EduHeart

TOWARD AN EVIDENCE-BASED UNDERSTANDING OF

ARTIFICIAL INTELLIGENCE APPLICATIONS ACROSS INDUSTRIES:

A MULTIDISCIPLINARY SYSTEMATIC REVIEW

Edited by: Cereneo S. Santiago, Jr.

Copyright Page

Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review
Published by:
EduHeart Knowledge Network and Publishing, Inc.©2026
B1 L54 Las Verandas Villas II, Buhay Na Tubig,
Imus City, Cavite, Philippines
Mobile No.: +63 952 4800719
email: info@myeduheart.com, www.myeduheart.com

Date of Publication: April 30, 2026
ISBN:
Softbound/Paperback: 978-621-493-438-6
EPUB: 978-621-493-439-3
PDF (read only): 978-621-493-440-9
PDF (downloadable): 978-621-493-441-6

Foreword Written by: Dr. Rowena Vargas-Isidro
Consultant: Dr. Ethel Reyes-Chua
Cover Page Photo Taken from: canva.com
Inside Photos Taken from: www.pexels.com
Layout Artist: Mr. Wilson C. Bonifacio

Accreditation & Registration
NBDB Registration Certificate No.: **9059**
SEC Registration No.: **2025080214804-13**

Acknowledgment

This multidisciplinary systematic review was made possible by the support of individuals and various institutions. We appreciate the efforts of the researchers, scholars, and practitioners from different disciplines whose published works formed the basis for this review. Their contributions to the body of knowledge on artificial intelligence have played an important role in building an evidence-based understanding of its application across a wide range of industries.

We also owe gratitude to the peer reviewers and academic colleagues who provided valuable insights, constructive suggestions, and guidance during the writing of this manuscript. Their knowledge contributed significantly to the rigor of this review.

Finally, we would like to acknowledge the support of our respective institutions and organizations, which have fostered creativity and interdisciplinary research. We wish to thank our colleagues for their encouragement, patience, and unwavering support in helping us complete this project.

Most importantly, we want to thank the technology and knowledge communities for providing an inspiring backdrop for continuing to examine the possibilities and uses of artificial intelligence in helping us address the challenging global problems we face today.

Dedication

To God, the Almighty, our families, friends, and to our respective institution, this work is dedicated to all of you.

The Authors

Foreword

As the rapid growth of AI continues to shake up entire industries, reshape how people do their jobs, and redefine what humans can achieve, we are entering an age of discovery in which all kinds of ideas about what we can achieve with Artificial Intelligence (AI) are emerging. In this context, Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review represents a timely and important contribution to the academic literature on AI.

This book provides readers with a well-founded, evidence-based overview of the different ways AI has been implemented across multiple industries, through a comprehensive synthesis of the extant literature. The multidisciplinary nature of the review provides a detailed assessment not only of the technological advancements associated with the development of AI, but also of the ethical, social, and economic implications. A single disciplinary perspective is not enough to understand AI's impact comprehensively.

In addition to presenting a wide array of viewpoints and approaches, this work also reflects a high degree of methodological rigor and clarity. Therefore, it is useful to researchers, scholars, and practitioners, as well as decision-makers (politicians and executives) who seek to exercise informed judgment in an increasingly AI-oriented world, because it addresses the gap between theory and practice. The systematic nature of the review indicates that its conclusions are based on a thorough analysis and synthesis of data rather than on isolated pieces of information.

This book also promotes critical thinking. In addition to celebrating the promise of AI, the volume discusses the challenges and risks of AI. In

addition to examining critical issues such as data privacy, algorithmic bias, transparency, and the future of work, their observations remind the reader of the importance of creating ethical paradigms and addressing human-centered issues alongside technological advancements.

This book serves as both a guide and a call to action at a time when the intersection between innovation and accountability is growing. The authors encourage the reader to take an in-depth look at the complexities of AI usage in all industries rather than simply taking a cursory look. More importantly, the authors emphasize that using AI for equitable and sustainable development requires collaboration and an evidence-based approach.

Finally, I hope to prepare the reader for a rewarding intellectual experience that will both provide an understanding of and inspire intentional engagement with the dynamic, rapidly evolving space of artificial intelligence.

Dr. Rowena Vargas-Isidro
Social Science Department
Iloilo Science and Technology University

Preface

Since beginning as a theoretical discipline, artificial intelligence (AI) has become increasingly visible in nearly every aspect of modern society, from healthcare to education to finance to industry to government and many others. The application of AI is also affecting the ways in which business decisions are made, as well as the delivery of services, the generation of records and other important information. However, there continues to exist an urgent need for empirical evidence regarding the functioning of AI within organizations to transition from disparate observations to a more comprehensive, data-based understanding of how AI is used in various industries. This need was the impetus for creating this volume.

Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review aims to compile and analyze the growing body of research on AI applications. This volume approaches this task in a systematic, multidisciplinary manner by assessing empirical data in each of the peer-reviewed journal articles on AI applications. The emphasis of this volume is to identify trends found within the research articles studied, and identify areas for which more research is needed; it is not to simply provide individual studies or theoretical discussions.

The premise of this study is that AI should be regarded as more than just a technological phenomenon; AI also should be considered social, economic, and ethical in nature. Further, there are several important concerns raised about the implementation of AI involving human agency and responsibility, equity, and transparency. Therefore, in order to truly understand AI, information from a breadth of fields, including but not limited to computer science, education, business, healthcare, social sciences,

and public policy, must be employed. The perspectives represented in this book are reflective of the varied fields of study associated with AI and provide insight into both the opportunities and challenges of adopting AI.

This systematic review follows a well-structured process that promotes rigor, transparency, and replicability to ensure reliability of findings reported. That is, the authors carefully reviewed, selected, and synthesized empirical evidence from a diverse range of studies providing a comprehensive yet critical assessment of the current state of AI applications across industries. While the authors have made every effort to include a diverse range of studies, the area of AI is constantly evolving, and it is likely new developments will occur that are outside of the scope of this review. Therefore, this volume should be viewed as both a compilation of what currently exists in research on AI applications and as an entry point for future research on the same topic.

This volume is designed for a wide variety of audiences, including researchers, educators, practitioners, policymakers, and students seeking to deepen their knowledge of AI applications across industries. No matter one's prior experience with AI, the authors hope to provide insightful data that is easily accessible and academically sound.

Overall, this book encourages continued participation, critical thinking, and ethical action within the rapidly evolving discipline of artificial intelligence. It is the collective responsibility of researchers, educators, practitioners, and policy makers to ensure that the applications of AI contribute positively to society through a durable and ethical foundation based on data.

Table of Contents

Chapter 1: The Use of Emerging Technologies in Automation of Rice Production: A Review

Michael J. Añano[1], Keischia Marie M. Mendijar[1], Anrou R. Nakila[1], Daniel G. De Guzman[2] & Johanna E. Ave[2]
[1]Cavite State University – Silang Campus, Silang, Cavite, Philippines
[2]National University - Las Piñas Campus, Manila, Philippines

Abstract. *Integrating emerging technologies in rice production has the potential to change traditional farming practices and improve productivity. Embracing automated technologies will not only promise advancements in securing food but also offer sustainable practices in agriculture, which can lay the foundation for food supply chains. This literature review aims to explore the effects of automation technologies, which are AI, IoT, and machine learning, on rice production for the progress of agricultural growth and securing food for the future. The systematic review used a structured approach for identifying relevant literature from reliable databases and analyzing the findings. Defining the research question, identifying study characteristics, searching for relevant literature, and summarizing the outcomes were the steps that were followed through. Results say that the application of AI, IoT, and machine learning in rice production can help reduce water usage and improve efficiency in different tasks. The implications of this research extend to technological advancements, economic benefits, social welfare, food sustainability, and governance in agriculture. This shows us the need to embrace automation in rice farming to make a way for a future with better food security and sustainable farming methods in the face of upcoming challenges. By fully implicating automated technologies, enhancement of food security, sustainable agricultural practices, and a brighter future for global food supply chains is a big possibility.*

Keywords: *Artificial Intelligence; Automation; Internet of Things; Rice Production; Sustainable Agriculture*

Introduction

Rice production plays a vital role in global food security, particularly in agricultural countries where rice is a primary staple food. However, traditional rice farming methods often face challenges such as labor shortages, inefficient resource utilization, and low productivity. With the advancement of technology, automation technologies such as big data, machine learning, and the Internet of Things (IoT) have been introduced to improve agricultural productivity and efficiency. Previous studies have explored the role of these technologies in modern farming practices. For instance, Alfred et al. (2021) and Veeragandham and Santhi (2020) examined the application of big data, machine learning, and IoT in agriculture and found that these technologies can improve decision-making, monitoring, and production efficiency. Despite these advancements, the application of automation technologies specifically in rice production remains limited, indicating the need for further research in this area.

The agriculture sector is one of the industries that adopted some emerging technologies in some areas. At present, technological advancements encourage automation to transform traditional agricultural practices and improve rice production productivity. For these technologies to be sustainable for farming and food security, it is crucial to understand the implications of their convenience. The literature on automation in rice production highlights the potential to change traditional farming practices. It also includes optimizing resource utilization and increasing production yields. Mujawar et al. (2024) note the positive impacts of emerging technologies. Some of these are real-time data collection using dedicated sensors and processing to anticipate future water demand. Although adopting these technologies may cause trouble for some farmers, a need for proper demonstration of these technologies should be taken into account. There is still uncertainty with regard to the versatility and sustainability of these technologies in different aspects of agriculture in the long term. Despite the valuable contributions of existing literature, previous studies may have limitations regarding sample size, geographical scope, and long-term monitoring of technology implementation.

Additionally, the generalizability of findings to diverse agricultural contexts implies that this warrants further investigation into automation in rice production. While there is substantial literature regarding the integration of emerging technologies in agriculture, gaps still exist and require more research. There are numerous research studies on the adoption of agricultural technologies, but there needs to be greater consistency, as different factors affect researchers' findings (2022). An example of these technologies used in agriculture is the Internet of Things. However, Khan et al. (2021) highlighted that the latest developments in IoT and other technologies for advanced agriculture were briefly discussed, emphasizing the need for further research to determine how these solutions are relevant for various agricultural environments, geographical locations, and farm sizes. Similarly, while other researchers have extensively implemented artificial intelligence (AI) in rice production, the application of AI in rice agriculture through literature reviews needs to be considered (2024). Despite these advancements, agriculture-related technologies were not quickly adopted because there is a gap between the actual usage of farmers on these technologies and their promotion on the market (Ambong, 2022).

This paper emphasizes understanding how emerging technologies contribute to improving rice production. This was accomplished by analyzing relevant literature published between 2018 and 2024 that applied to the research. To provide a precise control system for rice production, artificial intelligence (AI), Internet of Things (IoT), cloud computing, big data, and other emerging technologies are integrated explicitly with rice production applications (Hashim, 2024). Jha et al. (2019) also stated that several sectors in various countries depend on agriculture; therefore, growing technology, such as embedded intelligence, is necessary for a country's growth in agriculture. Hence, this paper examines the productivity and efficiency of emerging

technologies in rice production. Exploring the usability, practicality, and constraints of emerging rice production technologies will help uncover new methods and solutions to increase productivity and efficiency, ultimately improving the overall effectiveness of agricultural practices in improving rice cultivation and optimizing output.

The Review

The systematic review followed six steps outlined by Durach et al. (2017) for conducting such reviews. The specific procedures are listed below:

The Design

The systematic review followed six steps outlined for conducting such reviews in management. Initially, researchers defined the research question and then identified the necessary characteristics for the study. Next, searched for literature that could be relevant to their research question and selected the most pertinent pieces. Then, the relevant information gathered from the literature was analyzed and summarized. Finally, to report the outcomes of their review.

Research Objectives

The systematic review followed six steps outlined for conducting such reviews in management. Initially, researchers defined the research objectives, which are (a) to synthesize the emerging technologies used in rice production, (b) to synthesize the emerging technologies that contribute to the productivity of rice production, and (c) to synthesize the efficiency of emerging technology usage in rice production.

Retrieving and Selecting Pertinent Literature

The review retrieved and selected relevant literature from reputable databases such as IEEE, ScienceDirect, and Google Scholar to ensure the reliability of each data. The terms "automation, rice production, machine learning, artificial intelligence, and emerging technologies" were utilized to search these web databases. These terms resulted in 18743 total articles identified from the database search. Eighteen thousand two hundred (18200) articles were identified from Google Scholar, 539 results were identified from Science Direct, and four articles were identified from IEEE. A total of 24 papers from these internet databases were used after the articles were vetted against inclusion and exclusion criteria. Nevertheless, it was reduced to 17 papers by limiting it to only those published between 2018 and 2024 and evaluating each article's relevance to the study after reading the entire text.

Synthesizing the Literature

The studies were published between 2019 and 2024. 17 related literatures were selected and reviewed for this review paper.

Table 1. Table Synthesis.

Author	Emerging Technology Used.	Productivity	Efficiency
Alfred, R., Obit, J.H., Yee, C.C.P., Haviluddin, H. & Lim, Y. (2021) [1].	Big Data, Machine Learning, and IoT	Reduces the water usage and increases rice production.	Optimize tasks like irrigation, yield estimation, disease detection, and quality assessment.
AlZubi, A. A., & Galyna, K. (2023) [10]	Artificial Intelligence and the Internet of Things	A wider range of methods to monitor the crops.	Better crop production while reducing the need for human interaction
Aznan, A., Viejo, C.G., Pang, A. & Fuentes, S. (2023) [11]	Computer vision, machine learning, digital sensors	To assess the appearance and quality of rice	Automate the quality inspection procedure to lessen time and cost
Costa, F., Frecassetti, S., Rossini, M. & Portioli-Staudacher, A. (2023) [12]	Internet of things and Big data	Maps soil characteristics and texture properties	Irrigation plans can be tailored to the requirements of specific areas, which gives optimal crop irrigation and reduces soil monitoring time.
Eli-Chukwu, N.C. (2019) [13]	Artificial Intelligence	Artificial intelligence's flexibility can help in different sectors of rice production.	Able to produce more work and more profit comes with it.

Felizardo, K.B.F., Paredes, A.M.C. & Arboleda, E.R. (2024) [6]	Artificial intelligence	Enhance the quality of rice production in many aspects, such as disease detection, which ensures high-quality rice and optimized production.	DenseNet21, Deep Neural Networks, and Adaptive Neuro-Fuzzy Inference System showed the most accuracy in performance in the specific agricultural production aspect.
Hashim, N. Ali, M.M., Mahadi, M.F., Abdullah, A.F. Wayayok, A., Kassim, M.S.M. & Jamaluddin, A.J. (2024) [7]	Internet of Things	Detection of rice diseases and improved wet and dry irrigation system	Enhances pest and disease control and optimized resource management
Verma, I., Jain, S. and Sharma, S. (2023) [14]	Artificial Intelligence	AI enhances how human farmers work. AI helps farmers grow, harvest, and sell crops more efficiently. It also helps identify unhealthy crops, increasing the chances of growing healthy ones.	UAVs offer many benefits, like saving time and money, reaching plants and soil effectively, and reducing exposure to harmful chemicals.
Jha, K., Doshi, A., Patel, P. & Shah, M. (2019) [8]	Internet of Things, Machine learning, Artificial Intelligence, and Deep learning	Automation systems such as the use of AI, ML, IoT, and DL, help farmers come up with a decision that may be best applied in managing their production in agriculture.	Applying technological advancements can reduce labor and increase production.
Khan, N., Ray, R.L. Sargani, G.R. Ihtisham, M. Khayyam, M. & Ismail, S. (2021) [5]	UAVs, various wireless sensors, and IoT sensors	Providing precise data and insights throughout the cropping system, from land preparation to harvesting and transportation.	Optimizing various agricultural practices such as land preparation, irrigation systems, pest and disease management, and crop monitoring.
Sarkar, Md.R., Masud, S.R., Hossen Md.I. and Goh, M. (2022) [15]	Artificial Intelligence, Robots and Drones	Manage the challenges of agriculture from soil preparation to crop marketing.	Robots use AI to harvest and monitor plant growth, pick, sort, and pack. Special smart farming robots using machine vision and machine learning
Mujawar, R.Y., Lakshminarayana, R., Jyoti, A.P., Prabha, S., Patnaik, S. & Dhayalini, K. (2024) [3]	Machine learning models (ANN, SVM, DT, and RF)	Estimates future water consumption	It can help increase soil moisture and flexibility in irrigation planning.
Sajitha, P., Andrushia, D., Anand, N. Naser, M.Z. (2024) [16]	Vector Machine and Deep Learning	Development of Machine Learning Models for Plant Disease Detection	AI-image-based can be applied to detect plant diseases efficiently.
Shafie, S.M., Hami, N., Hassan, M.G., Musa, S., Hasanah, N., Nuraisyah, A. & Primaswari, A. (2023) [11]	Drones	Drones are used for tasks such as fertilization and pest control, enabling precise application of inputs and targeted management strategies that can improve crop health.	More efficient use of fertilizers and pesticides while reducing environmental impact.
Singh, R., Singh, R. & Kaur, P. (2023) [17]	Artificial intelligence and Deep learning	It helps identify plants and detect plant diseases.	The implementation of AI improved rice crop yield prediction.
Singhal, S., Ahuja, L. & Pathak, N. (2021) [10]	Artificial Intelligence and the Internet of Things	Optimize resource utilization and improve overall crop management practices.	It is helping farmers in decision-making situations with its data-driven features.
Talaviya, T., Shah, D., Patel, N., Yagnik, H. & Shah, M. (2020) [13]	Artificial Intelligence	Artificial intelligence applications in agriculture, such as irrigation, weeding, and spraying, with the help of sensors and other means embedded in robots and drones.	AI helps manage the challenges faced by various industries, including fields in the agricultural sector, such as crop yield, irrigation, and soil content sensing.

Discussion

3.1. Emerging technologies used in rice production

In recent years, integrating advanced technologies such as artificial intelligence (AI), the Internet of Things (IoT), machine learning, and drones into rice production has opened new avenues for enhancing traditional agricultural practices. These innovations are not just enhancing efficiency but are also setting the stage for a significant transformation in how rice is cultivated and managed.

Artificial intelligence has shown considerable promise in tasks ranging from disease detection to optimizing crop monitoring, irrigation, and pesticide application. Studies have demonstrated that AI algorithms can process vast amounts of data to optimize resource use, accounting for factors such as soil health, weather conditions, and crop growth patterns. For instance, as Verma et al. (2023) point out, the combination of AI with UAVs (Unmanned Aerial Vehicles) allows for efficient crop and irrigation monitoring. Equipped with AI, drones can cover extensive areas quickly, providing real-time data on crop health and growth conditions. This enables more precise agricultural practices, reducing waste and improving efficiency in rice production.

Similarly, the Internet of Things (IoT) has emerged as a critical tool in modern agriculture. IoT devices can provide continuous data on soil moisture levels and weather conditions, enabling farmers to make more informed decisions about planting, harvesting, and addressing potential issues before they escalate. Singhal et al. (2021) emphasize the role of IoT in improving crop management by facilitating real-time monitoring of environmental conditions, allowing farmers to respond swiftly to changes that could affect their crops.

Machine learning models, including Artificial Neural Networks (ANN) and Support Vector Machines (SVM), have also been applied in rice production to predict future water consumption and monitor irrigation systems. These models analyze large datasets—comprising historical weather patterns, soil moisture levels, and crop health indicators—to predict irrigation needs accurately. An IoT-enabled intelligent irrigation system, as highlighted in recent studies (Mujawaret al., 2024), showcases the potential of machine learning in improving irrigation planning, ensuring that water resources are used efficiently and effectively.

Drones, equipped with advanced sensors and cameras, further enhance agricultural practices by enabling precise application of pesticides and fertilizers. Drones can access areas that are challenging for traditional ground-based equipment, providing comprehensive maps of crop health and identifying areas that require attention. This targeted approach, as discussed by Shafie et al. (2023) not only reduces the use of chemicals but also minimizes environmental impact, ultimately improving the sustainability of rice production.

The collective impact of these technologies highlights their potential to revolutionize rice production. By promoting the adoption of these innovations and implementing sustainable practices, the agricultural sector can contribute to economic development while reducing resource consumption and minimizing environmental impact. Zaman et al. (2023) underline the importance of these technologies in overcoming challenges to food security, increasing productivity, and reducing waste.

3.2. Emerging technologies that contribute to the productivity of rice production

Over half of the world's population relies on rice as a staple food, making it a crucial agricultural industry that supports millions of farmers. As global demand for rice continues to grow, traditional farming methods are being transformed by emerging technologies that significantly enhance productivity and efficiency. These technologies can be broadly categorized into three groups: data-driven agriculture, automated machinery, and smart agriculture. Among these, automated machinery and smart agriculture stand out as particularly impactful in revolutionizing rice production.

Smart agriculture is redefining rice farming through the use of advanced technologies such as drones, imagery, and sensors. These tools optimize various aspects of rice production, including water distribution, fertilization, and pest control. Talaviya (2020) emphasizes that smart agriculture systems can accurately analyze data from sensors, drones, and other sources to manage these critical factors effectively, ensuring that rice crops receive the right resources at the right time. For example, smart sensors provide real-time monitoring of soil moisture levels, enabling precise water distribution and promoting healthy crop growth. Additionally, AI systems can analyze drone imagery to detect nutrient deficiencies in rice fields, allowing for targeted fertilization that boosts crop yield.

One of the most significant advantages of AI in agriculture is its predictive capabilities. AI systems can forecast weather patterns and identify potential threats, such as floods, droughts, or pest infestations, before they become severe issues. The study by Sungh and Bharti (2024) highlights how AI can issue early warnings, enabling farmers to take protective measures and secure their crops. This ability to anticipate and respond to weather conditions contributes to more consistent and healthier harvests.

Automated machinery represents another pivotal advancement in rice production. These machines can quickly and efficiently harvest mature crops, reducing the need for manual labor. Equipped with advanced sensors and precision-cutting technology, automated harvesters can identify fully grown rice plants and harvest them with care. Aznan et al. (2023) provide an example of how automated machinery reduces labor costs and increases the speed and efficiency of rice harvesting. By automating this process, farmers can ensure that rice is harvested at the optimal time for maximum yield, all while conserving resources and time.

The benefits of automated machinery extend beyond labor savings. According to Khan et al. (2021), automated systems are particularly advantageous during peak harvest seasons when manual labor is often scarce and expensive. These machines also excel in maintaining productivity during adverse weather conditions, as they are designed to operate effectively regardless of the weather. This capability is crucial in regions with unpredictable climate patterns, as it ensures that rice can be harvested promptly,

minimizing losses. Costa et al. (2023) note that automated harvesters equipped with cutting-edge technology can work efficiently across various landscapes, further enhancing their usability and effectiveness.

One of the latest advancements in automated machinery is the development of harvesters with machine-learning capabilities. These machines can learn from previous harvests, continually improving their accuracy and efficiency over time. Additionally, the integration of GPS technology allows for precise field mapping and navigation, reducing the risk of crop damage during harvesting.

Data-driven agriculture, while often less visible than machinery and smart systems, plays an equally vital role in enhancing rice productivity. This approach leverages vast amounts of data to provide farmers with actionable insights. By analyzing information from multiple sources, such as weather forecasts, market trends, and crop health data, data-driven tools enable farmers to make informed decisions that improve productivity. For instance, by studying market trends, farmers can determine the best time to sell their crops, maximizing profits. Similarly, by understanding the specific needs of their crops, farmers can adjust their strategies to produce higher-quality rice.

The economic impact of these emerging technologies is substantial. By optimizing resource use, reducing labor costs, and increasing yields, these advancements contribute to lower production costs and a more stable rice supply. This, in turn, enhances food security and boosts exports, driving economic growth. Laiprakobsup (2019) suggests that governments can further support this growth by lowering tax barriers on rice imports and exports, thereby increasing the competitiveness of rice in the global market. Additionally, investing in infrastructure projects, such as better roads and facilities, can help rural farmers transport their crops to urban areas more efficiently. Governments should also prioritize education and training programs to equip farmers with the skills needed to operate advanced equipment, analyze data, and use data-driven tools to make informed decisions.

The adoption of emerging technologies such as AI-driven precision agriculture and automated harvesting machinery is already significantly enhancing rice productivity. These technologies not only improve resource management and reduce labor costs but also increase crop yields, making rice production more efficient and sustainable. The study underscores the importance of government support in providing incentives and developing infrastructure to encourage the widespread adoption of these technologies. By embracing these advancements, the rice industry can continue to thrive in an increasingly competitive global market, ensuring food security and economic prosperity for millions.

3.3. Efficiency of emerging technology usage in rice production

Applying emerging technologies in rice production has proven effective in providing aid to agricultural production. These technologies allowed farmers to enhance their workflow by optimizing their resource allocation while providing support for decision-making processes, which is possible with the data-driven features of these technologies. Early detection of plant disease and monitoring of rice crop quality reduced farmers' efforts to identify the quality of their crops. With this, the farmers could maximize the yields of their agricultural operations. Emerging technologies enhance crop production while being able to reduce human interaction, indicating a shift towards automation and smart-farming practices while still being in line with existing theories of agricultural sustainability and innovation. Based on Singhal et al. (2021) findings, applying innovative technologies and connected devices to agriculture is called IoT cultivation. This integration created positive effects, including optimal information processing and water resource utilization. Internet of Things or IoT in agriculture reduced the misuse of resources. This includes electricity and water while saving farmers time through the utilization of sensors and interconnectivity. Efficiency was observed in conventional farming methods when emerging technologies were integrated into farming.

The deployment of IoT sensors made monitoring crop's health and soils possible. With these devices, helpful information can constantly assist farmers in managing crops more effectively because these devices enable them to make knowledgeable decisions. Emerging technologies are important in making agricultural tasks and improving efficiency significantly. Advanced irrigation systems guided by IoT sensors and artificial intelligence algorithms allow accurate water distribution. This adapts irrigation plans to cater to the farm's specific needs and reduces the amount of time spent monitoring soils. Weather fluctuations are one of the vulnerabilities the agriculture sector faces, as well as environmental challenges such as pest outbreaks and natural disasters. Emerging technologies allow farmers to enhance production efficiency. These technologies enable them to determine strategies for improvements in adapting to the impacts of environmental risks and climate change.

To ensure the sustainability of agricultural production systems, by monitoring soil health, pest outbreaks, and weather patterns in real time, farmers can reduce the effects of the risks that come with it. These technologies have a lot more to offer in agriculture. The agriculture industry can benefit from these technologies because they gradually develop over time. Combining contemporary IoT and AI technologies with conventional farming can improve the quality and quantity of agricultural produce (Mohamed et al., 2021). This shows that integrating IoT and AI technologies into traditional farming methods can significantly improve the quality and quantity of produce. The yields of farmers will improve, and productivity and optimized processes will increase by making use of these technologies. The profitability and sustainability of the agriculture industry will increase as well. Crops and resources can be manageable for farmers with the help of real-time insight of data from IoT and artificial intelligence systems. Using less water and pesticides is a systematic approach that lessens environmental impact while maximizing productivity. Ambong (2022) highlighted the essence of understanding the effects of automation technologies for sustainable development in agriculture and food security. Revolutionizing the rice production industry is possible with the help of automation technologies because this increases productivity, decreases labor requirements, and optimizes operations.

Assessing the extensive effects of automation in farming communities and agricultural ecosystems is important. Recognizing the outcomes of automation technology can guarantee long-term sustainability in rice production. Even if these technologies have a lot to offer, especially efficiency in agriculture, it is still crucial to recognize barriers to acceptance of using this kind of technology and the differences in access to this, especially in developing nations. Mohamed et al. (2021) emphasized that government assistance is necessary for small farms in developing countries to have these technologies. Farmers operating on a small scale in rural places may find acquiring IoT and AI technology expensive.

The technological divide between large-scale commercial farms and small-scale farmers is a growing concern, and without increased government support, this gap is likely to widen. This disparity in access to advanced technologies could exacerbate existing differences in agricultural production and income, putting small-scale farmers at a significant disadvantage. As large-scale farms continue to adopt cutting-edge innovations that boost productivity and efficiency, small-scale farmers may struggle to keep pace, leading to a further divergence in their ability to compete in the market and sustain their livelihoods. Addressing this issue requires targeted policies and initiatives to ensure that all farmers, regardless of scale, have the opportunity to benefit from technological advancements in agriculture.

Conclusion

The integration of automation technologies into rice production presents a significant opportunity to advance traditional agricultural practices. These technologies hold the potential to greatly enhance productivity, sustainability, and efficiency within the rice production industry, leading to more effective resource management and an overall improvement in the quality of production. By harnessing the capabilities of automation, the agricultural sector can contribute to a more sustainable future for global food production.

However, the successful implementation of these technologies in rice production requires careful preparation and strategic support. Government involvement is crucial, particularly in providing assistance to small-scale farms, especially in developing countries, where the cost of acquiring such technologies can be prohibitive. Government initiatives could include subsidies or financial incentives to help these farmers access automation technologies, thereby ensuring that the benefits of innovation are equitably distributed across the agricultural sector.

In addition to financial support, it is also essential to implement training programs that equip farmers with the necessary skills and knowledge to effectively utilize these technologies. Such programs would help farmers understand the capabilities of automation tools, enabling them to adopt these innovations more confidently and efficiently. This approach would not only enhance productivity but also ensure that the adoption of technology leads to tangible improvements in farming practices.

Furthermore, it is important to address gaps in the existing literature by conducting further research on the long-term sustainability of automation technologies in rice production. This research should explore the usability and limitations of emerging technologies, with the goal of discovering new methods and solutions that can further increase productivity and efficiency. Understanding these aspects will provide valuable insights into how automation can be optimized to meet the specific needs of rice producers, ensuring that the industry continues to evolve and thrive in the face of global challenges.

References

Alfred, R., Obit, J. H., Chin, C. P., Haviluddin, H., & Lim, Y. (2021). Towards paddy rice smart farming: A review on big data, machine learning, and rice production tasks. *IEEE Access, 9*, 50358–50380. https://doi.org/10.1109/ACCESS.2021.3069449

Ambong, R. M. A. (2022). Methods of rice technology adoption studies in the Philippines and other Asian countries: A systematic review. *Research in World Agricultural Economics, 3*(2), 15–24. https://doi.org/10.36956/rwae.v3i2.513

AlZubi, A. A., & Galyna, K. (2023). Artificial intelligence and Internet of Things for sustainable farming and smart agriculture. *IEEE Access, 11*, 78686–78692. https://doi.org/10.1109/ACCESS.2023.3298215

Aznan, A. A., Viejo, C. G., Pang, A., & Fuentes, S. (2023). Review of technology advances to assess rice quality traits and consumer perception. *Food Research International, 172*, 113105. https://doi.org/10.1016/j.foodres.2023.113105

Costa, F., Frecassetti, S., Rossini, M., & Staudacher, A. P. (2023). Industry 4.0 digital technologies enhancing sustainability: Applications and barriers from the agricultural industry in an emerging economy. *Journal of Cleaner Production, 408*, 137208. https://doi.org/10.1016/j.jclepro.2023.137208

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *Journal of Supply Chain Management, 53*(4), 67–85. https://doi.org/10.1111/jscm.12145

Eli-Chukwu, N. C. (2019). Applications of artificial intelligence in agriculture: A review. *Engineering, Technology & Applied Science Research, 9*, 4377–4383.

Felizardo, K. B., Paredes, Á., & Arboleda, E. R. (2024). Advancements in artificial intelligence (AI) for enhanced insights and automation in rice agriculture: A systematic review. *International Journal of Scientific Research Archives, 11*(1), 444–463. https://doi.org/10.30574/ijsra.2024.11.1.0092

Hashim, N., Ali, M. M., Mahadi, M. R., Abdullah, A. F., Wayayok, A., Kassim, M. S. M., & Jamaluddin, A. (2024). Smart farming for sustainable rice production: An insight into applications, challenges, and future prospects. *Rice Science, 31*(1), 47–61. https://doi.org/10.1016/j.rsci.2023.08.004

Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture, 2*, 1–12. https://doi.org/10.1016/j.aiia.2019.05.004

Khan, N. M., Ray, R. L., Sargani, G. R., Ihtisham, M., Khayyam, M., & Ismail, S. (2021). Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture. *Sustainability, 13*(9), 4883. https://doi.org/10.3390/su13094883

Mohamed, E. S., Belal, A. A., Abd-Elmabod, S. K., El-Shirbeny, M. A., Gad, A., & Zahran, M. B. (2021). Smart farming for improving agricultural management. The Egyptian Journal of Remote Sensing and Space Science, 24(3), 971–981. https://doi.org/10.1016/j.ejrs.2021.08.007

Mujawar, R. Y., Lakshminarayanan, R., P, J. A., P, S. P., Patnaik, S., & Dhayalini, K. (2024). IoT-enabled intelligent irrigation system with machine learning-based monitoring for effective rice cultivation. *International Journal of Intelligent Systems and Applications in Engineering, 12*(11s), 557–565.

Sajitha, P., Andrushia, A. D., Anand, N., & Naser, M. Z. (2024). A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. *Journal of Industrial Information Integration, 38*, 100572. https://doi.org/10.1016/j.jii.2024.100572

Sarkar, M. R., Masud, S. R., Hossen, M. I., & Goh, M. (2022). A comprehensive study on the emerging effect of artificial intelligence in agriculture automation. In *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)* (pp. 419–424). Selangor, Malaysia. https://doi.org/10.1109/CSPA55076.2022.9781883

Shafie, S. M., Hami, N., Hassan, M. G., Musa, S., Hasanah, N., Nuraisyah, A., & Primaswari, A. (2023). Energy consumption on farm automation: Case of paddy plantation. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 107*(1), 174–189. https://doi.org/10.37934/arfmts.107.1.174189

Singh, R., Singh, R., & Kaur, P. (2023). Rice crop yield prediction study by artificial intelligence techniques. *International Journal of Advanced Multidisciplinary Research Studies, 3*(3), 397–402.

Singhal, S., Ahuja, L., & Pathak, N. (2021). Impact of artificial intelligence and IoT in agriculture. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 668–671). Greater Noida, India. https://doi.org/10.1109/ICAC3N53548.2021.9725655

Sungh, B., & Bharti, P. (2024). Implementation of artificial intelligence in agriculture: An empirical approach. In *2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC)* (pp. 1–6). Tandojam, Pakistan. https://doi.org/10.1109/KHI-HTC60760.2024.10482365

Talaviya, T., Shah, D. N., Patel, N., Yagnik, H., & Shah, M. (2020). Implementation of artificial intelligence in agriculture for optimization of irrigation and application of pesticides and herbicides. *Artificial Intelligence in Agriculture, 4*, 58–73. https://doi.org/10.1016/j.aiia.2020.04.002

Veeragandham, S., & Santhi, H. (2020). A review on the role of machine learning in agriculture. *Scalable Computing: Practice and Experience, 21*(4), 583–589. https://doi.org/10.12694/scpe.v21i4.1699

Verma, I., Jain, S., & Sharma, S. (2023). Artificial intelligence techniques integrated UAV for next-generation agriculture. In *2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC)* (pp. 1–6). Puducherry, India. https://doi.org/10.1109/ICACIC59454.2023.10435110

Zaman, N. B. K., Raof, W. N. A. A., Saili, A. R., Aziz, N. N., Fatah, F. A., & Vaiappuri, S. K. N. (2023). Adoption of smart farming technology among rice farmers. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 29*(2), 268–275. https://doi.org/10.37934/araset.29.2.268275

How to Cite This Paper (APA 7)

Añano, M. J., Mendijar, K. M. M., Nakila, A. R., De Guzman, D., & Ave, J. (2026). The use of emerging technologies in automation of rice production: A review. In C. S. Santiago Jr. (Ed.), ***Artificial Intelligence Applications Across Industries: A Multidisciplinary Review of Emerging Technologies*** **(pp. 1-7). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 2: Efficiency and Security of Financial Technology Software Applications using Machine Learning: A Review

Jerryvhert Eloson[1], Louis Serrano[1], Lyndon Andrei Caraan[1], & Rowee M. Marfil[2]
[1]Cavite State University – Silang Campus, Cavite Philippines
[2]Eulogio "Amang" Rodriguez Institute of Science and Technology - Cavite, Philippines

Abstract. *The fast progress of financial technology (fintech) has changed the financial services field by providing solutions in response to new demands from businesses and customers. For this field, efficiency and security must be considered, as machine learning is a key component in addressing each challenge. This literature review aims to evaluate and assess machine learning integration with fintech software applications that contributes to the improvement of their efficiency and security which can also promote high quality assurance, necessary for users satisfaction. By reviewing existing investigations, this article shows how machine learning algorithms make transaction processing more efficient, detect fraud, and ensure user authentication, making processes simpler for streamlining operations and increasing protection against cyber-attacks. Nonetheless, various concerns like algorithmic limitations and safety matters connected with ML should be addressed along with its benefits, such as faster speed, better accuracy rates, or improved user experience. However, proactive measures may mitigate these risks, ensuring fintech software applications' continuous growth and innovation. The results show that ML can transform efficiency and safety in the finance industry to develop trust among users, leading to innovations.*

Keywords: *Financial Technology Software, Machine Learning, Efficiency, security*

Introduction

The transformation of fintech software apps has been associated with the ever-changing desires that businesses and consumers have over the efficient management of finances. These applications offer a range of services, including online lending, blockchain technology, and digital payments. Their proliferation has made financial services accessible to all people around the globe, making them available to everyone where effectiveness and security are non-negotiable. Operational performance enhancement, cost minimization, user experience improvement, and ensuring regulatory compliance (Kamuangu, 2024). These are crucial in this case, thus determining how better bank industries' cost efficiency is for market support service applications due to fintech innovations (Lee et al., 2021). However, such security-related issues as trustworthiness and reliability require comprehensive safety measures for financial terms to protect sensitive financial information and maintain trust and confidence among users (Inoue & Une, 2020). Therefore, it can be said that most concerns related to privacy and security may be addressed by well-executed advanced security algorithms that exploit machine-learning techniques.

Several risks and invasions can threaten the abovementioned elements because of breach of security, such as malware and phishing one of the most popular and controversial ones mainly done when an account holder voluntarily gives out information needed for an online transaction (Villanueva & Medalla, 2024). Recently, a local media outlet reported a significant compromise in security where approximately P37 million was transferred from different bank accounts held by various account holders to two other accounts within the country without the owners' approval (Villanueva & Medalla, 2024).

This literature review focused on assessing the efficiency and safety of machine learning-powered fintech software applications geared towards developing algorithms and statistical models. Before that, previous research on financial technology had analyzed, examined, and encouraged different algorithms or mechanisms to enhance system security, such as fraud detection (Mhamane & Lobo, 2012), and improve the effectiveness of software applications (Enkono and Suresh, 2020) besides their advantages and disadvantages of some statistical evaluation on accuracy (Sharma et al., 2018). Researchers also focused on using machine learning in the aspect of security to predict future risks for applications software (Stojanović et al., 2021) and how they handle growing uncertainties more efficiently, leading to better performance (Agarwal et al., 2023). This paper explores how machine learning can be used to increase the safety and effectiveness of fintech software applications for additional procedural simplification, user-friendly interface, and improved security against new vulnerabilities associated with the dynamic changing of technology.

THE REVIEW

The Design

This systematic literature review in management was conducted by following six steps (Durach et al., 2017), which helped define the necessary characteristics for the study and guided the review of machine learning. With this, researchers continue to retrieve pertinent literature, gather related and relevant information to the topic, and finally present and report the results.

Research Objectives

This study's overview was given by addressing the following research objectives: (1) To synthesize the efficiency of financial technology when used with machine learning. (2) To synthesize the security of financial technology when used with machine learning. (3) To synthesize the advantages and disadvantages of financial technology when used with machine learning.

Retrieving and Selecting Pertinent Literature

As the literature review focuses on the integration of machine learning in fintech software applications, selected literatures examine how the mechanisms enhance such qualities within financial services, which includes enhancement of overall user experience, protection against possible risks and considering limitations alongside benefits. The review uses reliable and replicable internet resources such as IEEE, Science Direct, and Google Scholar that have published articles about financial technology in general, which are searched using the phrase "efficiency, security, financial technology, online banking, e-wallet, machine learning." This review includes empirical research and case studies, theoretical papers, systematic literature reviews, industry articles, conference papers and books. The total number of articles identified from the database search is 30 articles. In composition, 12 articles were identified by Science Direct, eight results from IEEE, and 10 articles were found in Google Scholar. After screening the papers, 20 articles were passed based on inclusion and exclusion criteria. The full-text reading and analysis of these papers resulted in 15 articles.

Synthesizing the Literature

The studies were published between 1999 to February 2024. A total of 15 articles were selected that suit the purpose of this review. The data from the paper is extracted, and the primary study findings are analyzed and integrated into a Table 1.

Table 1. Synthesis of Reviewed Literature.

Author	Efficiency	Security	Advantages Disadvantages
Shaikh and Karjaluoto (2015)	M-Banking has emerged as an important and efficient distribution channel in terms of financial matters.	N/A	Findings show that m-banking and m-payments are gaining a strong foothold in developing countries.
Mhamane and Lobo (2012)	N/A	Following the HMM algorithm, the first user is upon login; if a user is authorized, then Internet banking is accessible.	N/A
Sharma et al. (2018)	N/A	The inhibitors related to technology would lead to 'Privacy concerns ' and 'Security concerns' for the users	The ISM-fuzzy MICMAC analysis-based model is designed better to understand Oman's recognized inhibitors of mobile wallets.
Kaur et al. (2018)	N/A	By enabling NFC-HFC, the first three layers(Application, Device, and Communication Security) are essential to protect the storage of tokens on the device and prevent token theft.	From the first three layers, the model ensures that the device and application have not been compromised and also ensures the security and confidentiality of communication between an e-wallet app and the payment network
Enkono and Suresh (2020)	Through an evaluation of NB and SVM classifier models, SVM classifier model was more efficient than the NB model concerning detecting scam e-wallet deposit notification SMSes	Various machine-learning classifiers have been employed to classify SMSes for filtering or detecting spam.	The SVM model's strength was highlighted by its good outputs(FNR = 0.000, CA = 0.992, recall = 1.000 and F1-measure = 0.995), making it the more efficient model.
Waqas et al. (2020)	N/A	Blockchain plays the role of giving customer privacy of keeping his transaction and receipts to preserve the record of purchases	Some threats to using a QR code are access to all personal information, all the social media networks and passwords, real-time location, and the possibility of infecting mobile and computers with malware.

Hariom Sharma (2023).	Evaluating and analyzing different machine learning approaches and their performances makes using credit cards more efficient.	The models developed and machine learning approaches in this research can be applied in financial institutions to strengthen security measures and reduce financial loss due to fraud	The major obstacle in implementing Machine learning for detecting fraud seems to be the presence of extremely imbalanced databases However, using balanced databases is beneficial to perform experiments efficiently.
SANI, Suleman Isah Atsu (2021)	The fact that graphical passwords are one of their most significant advantages makes it more efficient in terms of time.	The idea of using a graphical password is to increase password accessibility and protection by using the ability to remember picture images better than textual characters.	Many authentication-based applications , including electronic payment systems, find graphical passwords robust, especially regarding security and ease of use.
N. Sharma et al. (2021)	Applying the CCA approach will improve results which are worth getting to greater dependability in wallet exchanges	To security fix up and measure just encoded information and information trading SSL-based destinations by adding SSL Certificate pfx that causes the proponents on worker level code hitting just through Trusted customer hitting.	Through the implementation of the Concurrency Transaction Controller When using security tokens, algorithm transactions become more convenient and successful without duplicity.
Hassan, A., & Shukur, Z. (2021).	(MFA) is becoming a system that guarantees modern uses of available security and efficiency for those who need these while accessing sensitive data.	Once the user of an e-wallet app is authenticated, two or more authentication methods should be paired for better security	Specifically, users without extensive mobile phone experience will be comfortable using the apps after the initial glimpse.
Kandepu, R. (2023).	Overall, FileNet streamlines document-driven processes to increase operational efficiency in banking.	FileNet offers advanced, granular controls to secure banking data, ensuring robust protection and compliance, which includes Encryption for Data Protection, Access Control for Data Integrity, and Audit Trails for Compliance.	The extensive validation of FileNet's multifaceted advantages in enhancing security, ensuring compliance, and enhancing operational efficiency comes from theoretical research and case studies.
F. Wang et al. (2021)	The Efficient A random Oracle Model is introduced to detect malware on a host system and the multifactor authentication challenges posed during mobile payments.	The Security Strategy update scheme is then proposed dynamically to transfer a learning-based security strategy construct mechanism.	As the network supports several protocols, e-mail, and Internet applications could be secured, and all mobile apps can be secured. This is critical because of a mixed and multi-channel assault by most cyber-criminals.
Boughaci and Alkhawaldeh (2020)	Machine learning can be used to automate the financial process, which enhances the firm's efficiency.	A sophisticated tool that combines the two technologies of blockchain and learning algorithms is suited for protecting financial Data.	Machine learning algorithms can be used in credit scoring, which helps a lot in deciding whether to grant credit to applicants.
Kamuangu, P. (2024).	As the global FinTech market burgeoned to $297.1 billion in 2020, the transformative impact of these technologies became increasingly evident.	Machine learning models have been pivotal in identifying fraudulent activities and mitigating risks	The dynamic evolution of Artificial Intelligence (AI) and Machine Learning (ML) within the Financial Technology (FinTech) sector from 2016 to 2020 has ushered in a new era of efficiency, security, and innovation.

Stojanović et al. (2021)	Combine multiple models to build an optimal predictive model that is powerful in terms of both computation and efficiency.	Intelligent approaches from the domain of machine learning (ML) are applied to detect suspicious fraudulent patterns.	The results confirm that ML methods can successfully contribute to security in Fintech systems by supporting enhanced fraud detection capability.

Discussion

3.1 The efficiency of financial technology when used with machine learning.

Various software techniques like data analysis, pattern finding, and process optimization in accordance with the operations with credit cards can get funds to move faster and be more reliable and secure in terms of being used online through a credit card than the traditional bank way (Sharma, 2023). In a separate study, Enkono and Suresh (2020) found that by using machine learning models to analyze the issue, it is possible to recognize the mechanism that is required for the efficient detection of the e-wallet scam deposit notifications from SMS and the false transaction processing avoidance. In addition to machine learning algorithms, the integration of Robotic Process Automation (RPA) has impacted the efficiency of fintech software applications. According to Kamuangu (2024), the operational effects of the innovations, together with the Machine Learning and ML applications in FinTech, are very clear. Through the incorporation of RPA, there was a massive 20% drop in operational costs. This gain in efficiency also contributes to streamlining processes within financial institutions and gives them a niche for competitiveness in the real-time evolving market. The scalability and adaptability of RPA have huge potential to revolutionize operational frameworks further; hence, it remains a focal point for further exploration. Another approach is the concurrency control algorithm approach. This is worth considering the dependability of e-wallet exchanges, in which the number of duplicate transactions is reduced, and the number of push transactions is increased, leading to transaction reliability (Lee et al., 2021). As machine learning algorithms can improve accuracy, speed, and user experience in authentication systems, the test for a graphical-based authentication model gives an easier recall than the traditional alphanumeric authentications or passwords, having a huge advantage in time (Sani, 2021). Also, the efficient algorithms and machine learning techniques make FinTech software applications good at reducing manual efforts and increasing user protection. The identified use cases, such as real-time flagging of target apps and global monitoring, provide stakeholders with actionable insights to mitigate risks and improve the overall quality of finance-related apps (Fu et al., 2022).

There is a common thread in the findings of several studies on how machine learning affects the efficiency of fintech software applications. One aspect of this theme is machine learning's role in enhancing transaction processing speeds and accuracy. These two pieces, Sharma (2023) and Lee et al. (2021) show machine learning strategies like data analysis for faster and fraud-free card transactions and e-wallet exchanges. In evaluating the efficiency gains obtained from machine learning-based solutions for identifying e-wallet thefts, including eliminating false positives to improve speed and reduce security breaches (Enkono & Suresh, 2020). By improving online credit card usage or even spotting e-wallet scams and improving authentication systems, these algorithms have shown that they can optimize transaction processing, data analytics, and user experiences.

3.2 The security of financial technology when used with machine learning

Considering security as one of the important aspects of financial technology, it is important to analyze and dive deep into how machine learning contributes to it. A study by Kamngu (2024) found that the resilience of security measures such as fraud detection and prevention was improved by machine learning's ability to learn and adapt from real-time data with an accuracy of 99.5%. An algorithm presented from the study of Mhamane and Lobo (2012) is beneficial as the Hidden Markov Model offers benefits in modeling sequences of transactions by categorizing them into different states. At each state, determinations are made regarding whether the transaction constitutes a case of fraud. Also, an isolation forest algorithm is evaluated from the study by Sharma (2018), and the selected algorithm is best for credit card fraud detection, which works by iteratively selecting features and random values within their ranges to create partitions. Anomalies are isolated into shorter partitions, whereas normal data points need more splits to be isolated. As more fintech software applications need to have user authentication methods, the device identity-based user authentication algorithm from Hassan & Shukur (2021) can give users safe access to authorization through multi factor authentication framework as one of their highlights to promote an enhanced security when using e-wallet applications upon register, login, authentication, top up and transfer money which are critical measures to prevent adversarial attacks that attempts to do data manipulation, token forgery and inside threats.

Additionally, when it comes to authentication, a graphical-based authentication mechanism presented from the study of Sani (2021) highlights how users will increase their password protection through the ability of the system to remember picture images better than textual characters. Under the concept of machine learning applications, FileNet technology presented in the study by Kandepu (2023) explores how it will enhance the security of banking applications by integration of Hardware security modules for storing keys, cryptographic processing, and applying AES 256-bit (Advance Encryption Standard) encryption that guards against unauthorized data access or theft. Overall, these intelligent approaches from machine learning are applied to detect suspicious fraudulent patterns. Hence, ML encompasses anomaly detection techniques that automatically identify and classify suspicious data from financial innovations (Stojanović et al., 2021).

According to the results, the impressive accuracy rate is 99.5% for adaptive fraud detection and prevention using machine learning real-time data analysis. Hidden Markov Model and Isolation Forest algorithms isolate anomalies, and user authentication methods such as identity-based and graphical-based ensure secure access to financial services. Integration with advanced encryption standards adds an extra layer of security against unauthorized access and data theft. Machine learning smartly applied makes stronger security; FinTech platforms can mitigate security risks and safeguard sensitive financial data by enabling adaptive fraud detection, robust user authentication, and advanced encryption standards.

3.3 The advantages and disadvantages of financial technology when used with machine learning.

Besides improving security and optimizing the efficiency of financial technology applications, incorporating machine learning can bring various other possibilities. The machine learning concepts are explored from the study of Verma [17]. Firstly, self-implemented micro-communication through chatbots controls customer experience; some aspects of account generation, digital support, other activities, and cost-cutting measures are also discussed. They are self-learning and use Machine learning and Natural language processing algorithms to interpret customers' requirements and satisfaction index. Second, robo-advisors offer automated advisory services for portfolio management services, which involve the use of advanced intelligent algorithms such as CNN and RNN to learn and analyze data for investment decisions and efficient tax management. Finally, in credit risk assessment, such technologies help banks assess borrowers' ability to repay payments by crunching information like the balance in a borrower's account, their salary, and payment history, among others. Besides, according to the study by Sarhan (2020), the application of neural networks successfully provides methods for describing various important financial processes, including forecasting asset prices and analyzing the market. This is addressed through deep learning frameworks that incorporate nonlinear activation functions, making neural networks ideal for extracting complex patterns from noisy and stochastic data sets such as the one characteristic of the financial markets. These insights are extremely useful throughout the process of predicting market trends and making investment decisions in the case of algorithmic trading. Likewise, neural networks' use is not limited to credit card fraud detection; they are also used in credit risk evaluation, where banks can determine the borrower's credit worth by analyzing several sets of financial data. Because they are capable of modeling nonlinear connections and processing a significant quantity of data, neural networks play a large part in developing the applications known as fintech software, thus improving the processes of decision-making and risk management.

The other one is what is referred to as auto-machine learning. A paper that has discussed both the opportunities and challenges when it comes to using fintech software applications in business analytics includes Schmitt (2020), where AutoML packages such as H2O can help in speeding up the process of model prototyping and easing setup for practitioners who may not possess much knowledge in machine learning. Still, the same can result in lower prediction performance than manually optimized models. The AutoML approach is considered convenient, though the real-time result may not be the best model selection and stacking ensembles. Nevertheless, these are issues that cannot overshadow the functions AutoML performs to help fintech applications and businesses accelerate model development and create a basis for establishing the extended end-to-end decision-making process.

For all the said benefits, machine learning has a special significance for the finance industry because of the prospects to reduce expenses, cut down on operations, boost revenues by raising efficiency, increase customer satisfaction, and enhance security measures. Major financial institutions have diverse open-source algorithms and tools that apply to financial data analysis and ample funds for acquiring high-end computing equipment. Due to the huge amount of transactional data, the field of machine learning can influence many areas of finances, which makes it a key area of concern and future innovation for financial institutions that are planning to remain relevant in the financial markets that are constantly shifting. The problem with not exploring the machine learning approach in finance is that organizations must catch up technologically (Singla et al., 2021).

However, some machine learning algorithms may be vulnerable to hacking, making adoption a significant disadvantage. This has been evidenced by the vulnerability risks, which call for the adoption and installation of security to prevent risks. A study presented QR codes, while it was discovered that they had become a perfect replacement for current cash or credit card payment (Wang et al., 2021). They can be vulnerable as they have data embedded in a dotted image, they are cheap to produce and very easy to generate, and users love scanning them for what they represent, making them for cyber criminals as a tool to use (Waqas et al., 2020). Furthermore, the study by Sharma et al., (2018) mentioned that the significant disadvantage scholars seem to find while using the ML approach for fraud detection relates to the presence of highly imbalanced datasets. According to many strategies discussed in the study by Bhanushali et al. (2015), it can be seen that in the graphical-based authentication model, such as DAS (Draw A Street), Grid Selection, Pass Point, and Dejavu alive algorithms are modified. This relates to long and tedious authentication procedures, sequences or instructions that may need to be made more accessible to recall, user unawareness of the input mechanism, restricted password range, and exact click points that sometimes require repetitive effort.

Moreover, as the Hidden Markov Model can extract information with an accuracy of 92.9%, according to an experimental study by Seymore and Rosenfeld (1999), it is mandated to have a large amount of training data to fulfill the lack of improvement in classification accuracy. In terms of concurrency control protocols, researchers examined their performance. They found that distributed transactions on a cluster often only marginally surpass the throughput of non-distributed transactions on a single machine (Harding et al., 2017). The underperformance of concurrency control protocols stems from protocol-specific scalability bottlenecks, highlighting a significant scalability issue for distributed transactions.

In synthesis, the outcomes show the consequences of integrating machine learning with fintech software applications, mapping out the opportunities for fortifying security, comfort, and productivity with the exposition of the eminent susceptibilities of certain classes of machine learning algorithms. Furthermore, the issues raised concern identified in this paper about authentication models and the problems encountered while trying to detect fraud from unbalanced databases signify the need for continuous research efforts to tackle them. This demonstrates that while machine learning holds great potential for enhancing the innovation of the process in fintech, it is crucial to be proactive and undertake certain preventive measures against potential security threats and address the scalability problems of transactions in a distributed environment for safe usage.

Conclusion

Overall, findings from the literature research display insight into how machine learning has revolutionized financial software applications. It reveals how machine learning enhances productivity in relation to transaction velocities, precision, and capacities, as well as the identification of fraudulent transactions. Also, machine learning plays a crucial role in enhancing security features since it facilitates the development of dynamic fraud detection mechanisms, sound user authentication, and secure data encryption. However, it is important to understand that with these advantages of machine learning, there are corresponding drawbacks, including algorithm limitations and other security issues like problems with models for authentications, which call for preventive measures in tackling all these risks towards the advancement of more financial platforms. By applying machine learning technologies as a unified system, efficiency and security in the financial industry can be attained specifically in the development of software applications, thus enhancing user creativity and confidence.

References

Agarwal, P., Chinnasamy, G., & Kaushik, V. (2023). Maximizing financial management efficiency with a novel machine learning algorithm. *Multidisciplinary Science Journal, 5*, 2023ss0308.

Bhanushali, A., Mange, B., Vyas, H., Bhanushali, H., & Bhogle, P. (2015). Comparison of graphical password authentication techniques. *International Journal of Computer Applications, 116(1)*, 11–14.

Boughaci, D., & Alkhawaldeh, A. A. K. (2020). Enhancing the security of financial transactions in blockchain by using machine learning techniques: Towards a sophisticated security tool for banking and finance. *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, 110–115. https://doi.org/10.1109/SMART-TECH49988.2020.00038

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *Journal of Supply Chain Management, 53(4)*, 67–85.

Enkono, F. S., & Suresh, N. (2020). Application of machine learning classification to detect fraudulent E-wallet deposit notification SMSes. *The African Journal of Information and Communication, 25*.

Fu, J. F., & Mishra, M. M. (2022). Combatting fraudulent and predatory fintech apps with machine learning. *Poverty Action*. https://poverty-action.org/sites/default/files/publications/Combatting-Fraudulent-and-Predatory-Fintech-Apps-with-Machine-Learning-Policy-Brief-Fu-Mishra-Feb-2022.pdf

Halpern, J. Y., Harper, R., Immerman, N., Kolaitis, P. G., Vardi, M. Y., & Vianu, V. (2001). On the unusual effectiveness of logic in computer science. *Bulletin of Symbolic Logic, 7*(2), 213–236.

Harding, R., Van Aken, D., Pavlo, A., & Stonebraker, M. (2017). An evaluation of distributed concurrency control. *Proceedings of the VLDB Endowment, 10(5)*, 553–564.

Hassan, A., & Shukur, Z. (2021). Device identity-based user authentication on electronic payment system for secure e-wallet apps. *Electronics, 11(1)*, 4. https://doi.org/10.3390/electronics11010004

Inoue, S., & Une, M. (2020). Security analysis of machine learning systems for the financial sector. *Monetary and Economic Studies, 38*, 1–18. https://ideas.repec.org/a/ime/imemes/v38y2020p1-18.html

Kamuangu, P. (2024). Advancements of AI and machine learning in FinTech industry (2016–2020). *Journal of Economics, Finance and Accounting Studies, 6(1)*, 23–31.

Kandepu, R. (2023). Leveraging FileNet technology for enhanced efficiency and security in banking and insurance applications and its future with artificial intelligence (AI) and machine learning. *International Journal of Advanced Research in Computer and Communication Engineering, 12(8)*. https://doi.org/10.17148/ijarcce.2023.12803

Kaur, R., Li, Y., Iqbal, J., Gonzalez, H., & Stakhanova, N. (2018). A security assessment of HCE-NFC enabled e-wallet banking Android apps. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, 492–497. https://doi.org/10.1109/COMPSAC.2018.10282

Lee, C., Li, X., Yu, C., & Zhao, J. (2021). Does fintech innovation improve bank efficiency? Evidence from China's banking industry. *International Review of Economics & Finance, 74*, 468–483.

Mhamane, S., & Lobo, L. M. R. J. (2012). Internet banking fraud detection using HMM. *In ICCCNT'12 26th-28th July 2012, Coimbatore, India*.

Sarhan, H. (2020). Fintech: An overview. *ResearchGate*. https://doi.org/10.13140/RG.2.2.17196.28804

Schmitt, M. (2020). Artificial intelligence in business analytics: Capturing value with machine learning applications in financial services. *University of Strathclyde Digital Library*. http://digitool.lib.strath.ac.uk:80/R/?func=dbin-jump-full&object_id=34366

Seymore, K., & Rosenfeld, R. (1999). Learning hidden Markov model structure for information extraction. *Learning Hidden Markov Model Structure for Information Extraction.*

Shaikh, A. A., & Karjaluoto, H. (2015). Mobile banking adoption: A literature review. *Telematics and Informatics*, *32*(1), 129–142. https://doi.org/10.1016/j.tele.2014.05.003

Sharma, H. (2023). Credit card fraud detection using machine learning and deep learning techniques. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets45225

Sharma, S. K., Mangla, S. K., Luthra, S., & Al-Salti, Z. (2018). Mobile wallet inhibitors: Developing a comprehensive theory using an integrated model. *Journal of Retailing and Consumer Services*, *45*, 52–63.

Sharma, N., Sharma, S., Bharadwaj, G., Sharma, V., & Utkarsh, A. (2021). An algorithm for concurrency control in transactions for e-wallet. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *10*(6).

Singla, S. S. (2021, January). Machine learning for finance. *Google Books*. https://books.google.com.ph/bookshttps://www.aaai.org/Papers/Workshops/1999/WS-99-11/WS99-11-007.pdf

Stojanović, B., Božić, J., Hofer-Schmitz, K., Nahrgang, K., Weber, A., Badii, A., Sundaram, M., Jordan, E., & Runevic, J. (2021). Follow the trail: Machine learning for fraud detection in fintech applications. *Sensors*, *21(5)*, 1594.

Verma, M. V. (2019). Data-oriented and machine learning technologies in FinTech. *Staff Paper Series*, *5(1)*. https://www.idrbt.ac.in/wp-content/uploads/2022/10/Fintech.pdf#page=3

Villanueva, J. (2023). Medalla calls on e-wallet account holders to be vigilant. *Philippine News Agency*. Retrieved June 24, 2024, from https://www.pna.gov.ph/articles/1201702?fbclid=IwAR1_fze-6iweQMYTL-az8GqtzVodLPb0X7YC9TeOVeB-XAuYCH7n2ZKnZFc

Wang, F., Yang, N., Shakeel, P. M., & Saravanan, V. (2020). Machine learning for mobile network payment security evaluation system. *Transactions on Emerging Telecommunications Technologies*.

Waqas, A., Yousaf, S., Siraj, S., & Addepalli, L. (2020). A secured architecture for transactions in micro e-commerce using QR scan, e-wallet payment applications. *ResearchGate*. https://www.researchgate.net/publication/355022229_A_Secured_Architecture_for_Transactions_in_Micro_E-Commerce_using_QR_scan_e-_Wallet_Payment_Applications_with_Adaptation_of_Blockchain

How to Cite this Chapter (APA 7):

Eloson, J., Serrano, L., Caraan, L. A., & Marfil, R. M. (2026). Efficiency and security of financial technology software applications using machine learning: A review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 8–14). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 3: Analysis of the Development of Natural Language Processing Methods: A Literature Review

Patrick Pacey Lozano[1], Maybelle Tambal[1], Princess Vanesse Timajo[1] & Dr. Cheeryl V. Garcia[2]
Cavite State University – Silang Campus, Cavite, Philippines
[2]Iloilo Science and Technology University - Miagao Campus, Iloilo, Philippines

Abstract. *Most attention has been paid to NLP because computational processing and interpretation of human language can guide transformation in various sectors. This research focuses on the study of the evolution and application of three formidable NLP frameworks: rule-based, statistical, and deep learning approaches. The research used a systematic review of the literature to synthesize strengths and weaknesses inherent in these techniques and future research directions. The findings proved that at the heart of the optimism for NLP in extracting insights out of large volumes of unstructured data lies a challenge in handling linguistic ambiguities and contextual nuances, especially in a high-precision domain like aviation safety and legal practice. The study has emphasized that more robustness and scalability, greater care for ethical considerations, and improved reliability and applicability across varying linguistic and cultural contexts are very much needed in NLP models. It contributes to the advancement of technology by shaping ethical frameworks that can take AI deployment toward a more inclusive AI solution affecting society in general.*

Keywords: *Natural Language Processing, Rule-based, Machine Learning, Neural Network, NLP Development*

Introduction

Recently, there has been a lot of interest in natural language processing as a means of computationally representing and interpreting human language (Khurana et al., 2022). Natural Language Processing (NLP) is a part of artificial intelligence (AI) and the study of language that helps computers get, make sense of, and create human speech in a way that's meaningful and fits the context. The goal of the automated NLP is to perform this task as accurately and efficiently as a human can (Chowdhary, 2020). New steps forward in NLP have come from better ways to build models and prepare them first, especially with Transformer structures. With the use of linguistic, statistical, and artificial intelligence techniques, natural language processing (NLP) can be used to generate human-like responses or even to interpret textual data (Fanni et al., 2023). The open-source Transformers tool makes these new steps easy for more people working on machine learning. Innovative computing architecture research and development targeted at ML acceleration are also highlighted (Carleo et al., 2019). Despite big leaps forward, there are still big questions and debates, like how to explain what deep learning models are doing and the moral issues of training data that are not fair. Since the training samples play a major role in DL-based quantitative retrieval, this review will also shed light on how to apply DL effectively when there are insufficient samples (Yuan et al., 2020). Rewrite this piece, and follow these instructions: use simple and short words, switch complex words for easier ones, and mix up the length of sentences. Use the most common English words when you can. Keep the word count the same. Past studies have looked at important issues in making models better and faster but found mixed answers on the top ways to reduce bias and ensure fairness. Issues include the significant need for computing power to train big models and the problems in making models work well for languages and areas with few resources. NLP models are highly relevant in higher education as they can help students learn in various ways (Fuchs, 2023).

The need for NLP to access relevant, good-quality training data remains; most applications require domain-specific or labeled datasets, which are challenging to access. With its ability to annotate and summarize, NLP can help with this task (Shaik et al., 2022). Deep learning models are often incomprehensible; therefore, the accuracy of their predictions is of concern in such critical fields as health and finance. While deep learning techniques have the benefit of automatically and implicitly extracting features without requiring explicit human-engineered feature extraction steps, traditional automatic ECG analysis has relied on diagnostic golden rules (Hong et al., 2020). There are ethical challenges in terms of algorithm bias and the potential for disseminating false information. Recent efforts underline the need for structured datasets, standardized solutions, and a focus on emotion expression detection. The significance of deep learning (DL) and various DL networks and techniques are discussed. The difficulties and potential solutions to close current research gaps are also highlighted (Alzubaidi, 2021). However, automated approaches to trend capture can easily be biased and opaque; only after that will debates ensue about how reliable and ethical they are. In this respect, most NLP models have been constructed in English. That means that more research is required in multilingual and cross-lingual understanding so that the solutions can be all-inclusive and applicable globally.

To understand how various NLP frameworks and techniques, such as the rule-based, statistical, and deep learning approaches, into understanding how they evolved and have been applied. The framework aims to give conceptually disparate word and document embeddings a straightforward, uniform interface (Akbik et al., 2019). The study seeks to inform how the different approaches have contributed to the evolution of NLP by taking into consideration how far such contributions have made their mark on the landscape of NLP research. Hence, this study explores the progress in the development of NLP technologies and applications and points to some future directions for work aimed at enhancing the robustness, fairness, and accessibility of NLP systems.

The Review

The Design

The researchers made a critical literature review using the systematic approach for conducting and reporting systematic literature reviews. Previous studies advanced the knowledge base by contributing an extensive literature review, thereby summarizing past studies (Paul & Rialp-Criado, 2020). The researchers used a 6-point guideline to conduct the literature reviews. Firstly, we had to specify the questions and objectives to be addressed in the literature review. We then gathered relevant research to analyze the data to be gathered. After this, we gathered relevant information from the literature, and then the report's findings were interpreted and summarized.

Research Objectives

This literature review seeks to address the research question: how have different approaches, such as rule-based, statistical, and deep learning, contributed to the evolution of NLP? It aims to achieve the following objectives: (1) to synthesize strengths inherent within NLP frameworks and techniques, (2) to synthesize weaknesses of NLP, and (3) to identify future directions in NLP research.

Retrieving and Selecting Pertinent Literature

The review utilized information and computer science literature gathered from a reputable and credible online database that has published relevant literature. Search string keywords used for the two online databases are "Natural language, development, Machine Learning". Seventeen thousand nine hundred (17,900) articles were identified from the database search. Thirty-two thousand eight hundred eleven (32,811) results were identified from ScienceDirect and 12 articles from IEEE. Additionally, Google Scholar was used to look up most articles from websites published by other journals. About a thousand articles were found, including those from reputable publishing websites like IEEE, ScienceDirect, etc.

Subsequently, the papers were screened according to the inclusion and exclusion criteria, yielding 37 articles. After that, each article was thoroughly read and analyzed, producing 20 finalized articles.

Synthesizing the Literature

The selected pertinent literature was published from the year 2020-2024. The total articles are 20, which met the criteria for the review and were selected from the studies, extracting data from each article and examining the main study findings combined in Table 1-3.

Table 1. Table synthesis about the strengths of Natural Language Processing methods.

Author	Strengths of NLP processing methods
Yang, C., and Huang, C. (2023)	Analyze large volumes of aviation safety-related data contained in incident reports and ATC communications, thus improving situational awareness and decision-making
Kolahdouz-Rahimi, S., Lano, K., & Lin, C. (2023)	It allows adaptability to structured and semi-structured data, making heuristic NLP approaches effective for information extraction in RF applications
Kumar, D., Haque, A., Mishra, K., Islam, F., Mishra, B. K., & Ahmad, S. (2023)	It will analyze vast amounts of educational data, enabling personalized learning experiences through adaptive systems and intelligent tutoring.
Wang, Z. (2024)	Extract structured information from unstructured texts by state-of-the-art methods of named entity recognition, relationship extraction, and event extraction to construct comprehensive knowledge maps.
Quevedo, E., Cerny, T., Rodriguez, A., Rivas, P., Yero, J., Sooksatra, K., Zhakubayev, A., & Taibi, D. (2023)	It is potentially positioned to dramatically make legal processes much easier by document review, information retrieval, and many other tasks that aim at efficiency aide in legal practice.
Sinha, A. K., Akhtar, M. a. K., & Kumar, A. (2021)	Parse and analyze any resume in unstructured text data and allow fast filtering of information
Ghazizadeh, E., and Zhu, P. (2021)	It has several other domains, such as natural language understanding, generation, and translation, due to its capability to process and interpret a highly complex linguistic structure.
Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022)	Design new methods to make neural networks more sensitive to perturbations for better out-of-distribution accuracy, with a special focus on searching for and improving the robustness weaknesses related to text classification settings.
Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023)	Using NAO robots for individualization of learning through personalization, interactive feedback, and examination of collaboration skills among students

Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024)	Enable state-of-the-art educational applications, question answering, automated assessment, and error correction, through the process of becoming actualized in sophisticated teaching and learning designs with integrated advanced capabilities in language understanding and generation
Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022)	Processing of unstructured textual data in EHR with enhanced tasks of classification, prediction, information extraction, and question answering for betterment in health informatics and improved patient care
Kang, Y., Cai, Z., Tan, C., Huang, Q., & Liu, H. (2020)	Extract and understand document information, radically improving efficiency and accuracy across diverse fields such as management research using advanced machine learning and deep learning algorithms.
Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022)	It facilitates treating and analyzing unstructured text data to identify numerous characteristics of phishing, such as misspelled domain names, poorly written content, suspicious attachments or links, and phishing warning messages.
Ayenew, H., & Wagaw, M. (2024)	It generally improves the software testing process to achieve effectiveness, reliability, and efficiency, hence facilitating the delivery of quality software products.
Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023)	It will enhance efficiency and accuracy in safety documentation and occurrence reporting using state-of-the-art language models and semantic search capabilities.
Shetty, V. A., Durbin, S., Weyrich, M. S., Martínez, A. D., Qian, J., & Chin, D. L. (2024)	The core base for measuring empathy in text with applications in patient-physician asynchronous communication studies
Song, K., Ran, C., & Yang, L. (2022)	Enhance the efficiency of tech transactions and R&D collaboration by better identification of related patents and technology research teams.
Sawicki, J., Ganzha, M., & Paprzycki, M. (2023)	Fully automate complex analysis of large scientific literature datasets to enable scalable and replicable insights across diverse fields and languages.
Al-Azani, S., Sait, S. M., & Al-Utaibi, K. A. (2022)	Systematically classify, evaluate, and synthesize multidisciplinary research methodologies and data characteristics to provide comprehensive overviews that would help develop advanced machine learning-based solutions for technology related to children.
Alsmadi, I., Ahmad, K., Nazzal, M., Alam, F., Al-Fuqaha, A., Khreishah, A., & Algosaibi, A. (2023)	Develop sophisticated attack and defense mechanisms attuned to social media applications' nuanced and text-centric nature to secure and make them reliable.

Table 2. Table synthesis about the weaknesses of Natural Language Processing methods.

Author	Weaknesses of NLP Processing Methods
Yang, C., and Huang, C. (2023)	The major challenges affecting the accuracy and reliability of NLP applications in aeronautical safety are handling ambiguity in language, limited support for multilingual analysis, and dependence on the availability of sufficient and diverse training data
Kolahdouz-Rahimi, S., Lano, K., & Lin, C. (2023)	Deep learning techniques are adopted less often than classical machine learning techniques, like decision trees or SVM, thus missing the potential of more sophisticated context-aware models for RF tasks.
Kumar, D., Haque, A., Mishra, K., Islam, F., Mishra, B. K., & Ahmad, S. (2023)	Possible mistakes in the correct interpretation and contextualization of educational content, especially in multilingual and multicultural settings, may affect the efficiency of personalized learning interventions.
Wang, Z. (2024)	Ambiguities and context-dependent subtleties in language can lower information extraction accuracy and, thus, the construction of knowledge maps; for instance, this can occur in exceptionally diversified domains or less-studied areas.
Quevedo, E., Cerny, T., Rodriguez, A., Rivas, P., Yero, J., Sooksatra, K., Zhakubayev, A., & Taibi, D. (2023)	Processing of long or legally technical documents; models are difficult to transfer across different legal sub-domains; ethical concerns regarding privacy and bias in sensitive applications, like judgment prediction
Sinha, A. K., Akhtar, M. a. K., & Kumar, A. (2021)	It accurately understands subtle writing styles, a variety of words, and syntax differences of unstructured language, impacting the accuracy in semantic search and contextual analysis.

Ghazizadeh, E., and Zhu, P. (2021)	The accuracy of natural language understanding and generation tasks, in particular under highly nuanced settings or languages; limitations to deal with ambiguity and context-specific nuance well
Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022)	Obtaining high precision and reliability within real-world text classification applications due to vulnerabilities most of the time, such as spurious correlations and insensitivity that is high to shifts in data distributions
Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023)	The privacy, fairness, and transparency of automated grading and personalized instructions, along with the continuous adaptation to a myriad and diverse set of educational environments and student needs
Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024)	Domain-specific knowledge is effectively combined: over-correction in language models ensures strong performance in diversified educational contexts and tasks.
Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022)	Dealing precisely with the complexity and variability of medical language in unstructured text and issues of integrative and standardizing various data sources for integrated analysis
Kang, Y., Cai, Z., Tan, C., Huang, Q., & Liu, H. (2020)	Precise measurement of complex concepts and assembling methods for the solution of particular problems; diversified and dynamic business contexts, which limits the practical applications
Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022)	The scarce resources and advancement in handling non-English languages, especially Arabic, and the associated difficulties with adapting the existing models to the/eventual variety of linguistic and cultural contexts set very narrow boundaries of generalizability and effectiveness for these techniques.
Ayenew, H., & Wagaw, M. (2024)	To what extent exploration and validation of existing techniques and tools are required, quite often not fitting a broad variety of different needs and complexities of various software projects
Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023)	Datasets on safety are limited in availability, wherein the ChatGPT family of models can factually generate plausible wrong answers because of limitations in training and being hallucination-prone
Shetty, V. A., Durbin, S., Weyrich, M. S., Martínez, A. D., Qian, J., & Chin, D. L. (2024)	Conceptualizing and capturing subtle aspects of empathy in the text itself impacts the precision of measure and reliability.
Song, K., Ran, C., & Yang, L. (2022)	It fully addressed the high complexity and specificity involved in the evaluative transferability of a patent and the detection of appropriate research teams that might impact the system's effectiveness.
Sawicki, J., Ganzha, M., & Paprzycki, M. (2023)	Degree of treatment of domain-specific subtleties and context-specific intricacies impacting the precision and relevance of the automated analyses
Al-Azani, S., Sait, S. M., & Al-Utaibi, K. A. (2022)	Addressing enough particulars and finesse of life stages and tasks for children in technology and various efforts taken to develop a few multi-view, multi-sensor, multimodal, and multi-label techniques.
Alsmadi, I., Ahmad, K., Nazzal, M., Alam, F., Al-Fuqaha, A., Khreishah, A., & Algosaibi, A. (2023)	Persistent vulnerabilities against adversarial attacks, which would undermine the effectiveness of existing defense mechanisms because social media content is ever-changing and dynamic

Table 3. Table synthesis about the future directions of Natural Language Processing Methods.

Author	Future directions in NLP research
Yang, C., and Huang, C. (2023)	Improved multilingual support, enhanced robustness of NLP models to domain-specific subtleties and variations in aeronautical communications, and expansion of datasets to allow more complete training, all for further improved safety outcomes
Kolahdouz-Rahimi, S., Lano, K., & Lin, C. (2023)	Benchmarked datasets and deep learning approaches should be standardized to further broaden the performance and flexibility of RF techniques against a range of applications.
Kumar, D., Haque, A., Mishra, K., Islam, F., Mishra, B. K., & Ahmad, S. (2023)	Begin to develop multilingual and multicultural competencies to allow AI-powered learning tools to cater more effectively to pluralistic pupil populations, focusing adaptively on refining learning algorithms to establish continuous improvement and deliver finer-tuned experiences.

Wang, Z. (2024)	Enhance robustness and scalability of information extraction techniques with multilingual capabilities and deeper semantic understanding for more accurate and context-aware knowledge map constructing, thereby extending applicability and efficiency in various domains for AI-driven solutions.
Quevedo, E., Cerny, T., Rodriguez, A., Rivas, P., Yero, J., Sooksatra, K., Zhakubayev, A., & Taibi, D. (2023)	Better datasets, more domain-generalizable models across legal sub-domains, symbolic methods that can integrate existing legal knowledge/rules for more robust automation in legal tasks
Sinha, A. K., Akhtar, M. a. K., & Kumar, A. (2021)	Semantic search capabilities that identify and contextualize better the varying writing styles and syntaxes in résumés, and more robust machine learning models that allow extracts with higher accuracy relevant information during resume screening processes
Ghazizadeh, E., and Zhu, P. (2021)	Deepening the natural language understanding and generation capability with deeper semantic understanding, more efficient context awareness, and better adaptability to different linguistic patterns and cultural nuances in the pursuit of developing robust and contextually correct AI applications
Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022)	Improvement in techniques for more robustness throughout the NLP pipeline mostly focused on enhancing techniques for handling spurious correlations, increasing sensitivity to perturbations, and high accuracy in out-of-distribution scenarios that can ensure more reliable and effective deployment of NLP systems in practical applications.
Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023)	Development of technologies that would increase the adaptability and inclusiveness of NAO robots in education for improving natural language understanding to facilitate effective communication and personalized learning pathways, more ethical frameworks to ensure privacy, and more equitable educational outcomes
Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024)	Tighter Integration of LLMs: Tighter integration of LLMs with traditional methods could help to make winning performance gains in educational tasks better, such as domain knowledge incorporation, lowering over-correction challenges, and developing specialist tools and tollkits supporting broader educational applications like knowledge graph construction and controllable text generation
Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022)	Improvement in the robustness and accuracy of deep NLP models for EHR tasks, domain-specific representation learning, and methods for better integration of structured and unstructured data toward more holistic health informatics solutions
Kang, Y., Cai, Z., Tan, C., Huang, Q., & Liu, H. (2020)	Continuously improving and breaking through algorithms and datasets to provide more accurate analysis methods and strive to support researchers in making vital theoretical and practical contributions to management literature.
Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022)	This is the integration and evaluation of very deep recent approaches to learning, including LSTM and CNN models, exploration of hybrid and multi-algorithms systems, and development of new solutions for problems related to concept drift. It also improves the methods for phishing attack detection across languages and cultures.
Ayenew, H., & Wagaw, M. (2024)	More robust NLP algorithms and frameworks, with increased focus on real-world applications and scalability, and complete datasets and thorough benchmarks for adequately comparing these multiple approaches
Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023)	Safety-themed dataset construction| fine-tuning language models for specific tasks in safety | tailoring NLP applications to the unique requirements of processes concerning safety in a way that ensures complementarity with—not replacement of—human expertise
Shetty, V. A., Durbin, S., Weyrich, M. S., Martínez, A. D., Qian, J., & Chin, D. L. (2024)	Rework the conceptualization of empathy in text and create more sophisticated models attempting to better capture intricate, nuanced expressions of empathy in patient-physician communication.
Song, K., Ran, C., & Yang, L. (2022)	Increased handling of the complexity of patent evaluations and research team detections, with the integration of more advanced ML techniques for high accuracy and reliability in such

Sawicki, J., Ganzha, M., & Paprzycki, M. (2023)	Advanced domain-specific personalization and contextual understanding in automated analyses, with better integration of the diversity of NLP methods for dealing with the different types within the scientific literature
Al-Azani, S., Sait, S. M., & Al-Utaibi, K. A. (2022)	Construction of large, well-documented, and reliable databases that fuel deep-learning-based solutions targeting research gaps across a variety of children's life stages and tasks, with particular attention to such ethical considerations as informed consent
Alsmadi, I., Ahmad, K., Nazzal, M., Alam, F., Al-Fuqaha, A., Khreishah, A., & Algosaibi, A. (2023)	Addressing security challenges of ML for social media applications by developing more robust and adaptive attack and defense techniques, ensuring that they will be able to respond effectively and quickly to change in an online communications landscape that is fast-changing

Discussion

3.1 Strengths inherent within NLP frameworks and techniques.

Based on the papers reviewed, NLP is distinct from several other notably valuable properties: its ability to make sense out of large data analysis by interpreting and eliciting insight from them to improve decision-making in important areas such as aviation safety, legal practice, and healthcare. To provide more intelligent decision support, they propose a new methodology for MpMcDM problems that combines SA and DM methods for processing the expert evaluation in natural language (Zuheros et al., 2021). For instance, aviation can use NLP to analyze incident reports and air traffic communications to enhance situational awareness and operational decision-making. A methodology for identifying and categorizing human factor categories from aviation incident reports is presented in this study (Madeira et al., 2021). Similarly, NLP organizes such complicated procedures as document review and information retrieval in legal applications to increase the speed of legal procedures. NLP can potentially automate information extraction and processing from unstructured text, a task central to many legal tasks (Frankenreiter & Nyarko, 2022). All of the applications mentioned above reflect the ability of NLP to work with structured and unstructured data. The applications above showcase natural language processing's capacity to handle structured and unstructured data through knowledge extraction techniques, such as named-entity disambiguation, entity and relation recognition, and relation linking. These techniques are designed to map the extracted knowledge to pre-existing knowledge bases or graphs (Sakor, 2023). Techniques, like named entity recognition and relationship extraction, are used to create full-fledged knowledge maps and further enable advanced tasks like classification, prediction, and question-answering. We can also benefit from the information contained in a semantic graph by creating a link between a textual mention and an entity in a knowledge graph. This information has been demonstrated to be helpful in NLU tasks like information extraction, biomedical text processing, semantic parsing, and question answering (Sevgili et al., 2022).

NLP can be used to revolutionize many industries, considering its power of adaptation and robust processing capabilities. Large amounts of unstructured text data can be sorted through Natural Language Processing (NLP) algorithms, which can then be used to find patterns, trends, and insights that would be extremely difficult to find by hand (Khatri, 2023). They further confirm the existing theories for integrating NLP into various applications for improved efficiency, accuracy, and decision-making processes. Related works attest to the efficaciousness of using NLP, syntactic analysis, and word vector techniques to conduct a thorough analysis of the potential implicit design relationships of knowledge (Jing et al., 2023). Literature provides evidence that further improvement is required in NLP with respect to robustness in the field of text classification and scalability in handling large datasets. Research on adversarial examples has significantly increased in computer vision and natural language processing (Mujtaba et al., 2019). This means new perspectives on developing techniques sensitive to perturbations that would help improve generalization across different domains. Learning algorithms can be pre-processed using feature selection techniques, and effective feature selection outcomes can shorten learning times, increase learning accuracy, and simplify learning outcomes (Cai et al., 2018).

They probe into some novel approaches to boosting the capability of NLP in specified domains such as educational technology and healthcare informatics. By focusing on personalized learning experiences, adaptive systems, and efficient data extraction from electronic health records, we hope to fill current gaps identified in the literature. Moreover, one of the research objectives is to further this field with regard to investigating the embedding of NLP, together with potentially promising emerging technologies such as robotic assistance in education and semantic search in health, toward better outcomes and efficiencies for these important domains.

3.2 Weaknesses of NLP.

Based on the reviewed paper, several overarching problems with literature suggest that NLP cannot deal with ambiguity in language—it is most evident in areas of high-precision interpretation, such as aviation safety and legal practice. This work examines this ambiguity level from a computational and combinatorial perspective (Khalife et al., 2021). This also encompasses subtler context-dependent nuisances and variations that makeup language and, as such, hamper tasks like information extraction, reducing the accuracy of semantic search. Large volumes of text data are used to train these models, like GPT-3 and GPT-4,

which can then be optimized for specialized downstream tasks like question answering or language translation (Hadi et al., 2023). Another key challenge to NLP is its reliance on adequate and diversified training data, one major limitation affecting models' reliability and generalizability across languages and domains. A common belief is that beginning with general-domain language models can help even domain-specific pretraining (Gu et al., 2021). These limitations suggest that further work should be done on methods that can further contextualize the language, reducing dependencies on heuristics and the data sets to improve overall performance and versatility. Although many methods and resources are available for resolving ambiguities, there isn't enough comparison analysis or conversation about the methods and resources employed in ambiguity resolution (Yadav et al., 2021).

NLP has repeated that; indeed, it is quite complicated to come up with a strong natural language understanding and generation system. On the GLUE benchmark, the SQuAD 2.0, and the CoQA question-answering tasks, UniLM compares favorably with BERT (Dong et al., 2019). Previous theories support the idea that technology alone will not solve these limitations, but further advancement in NLP is needed. The literature also offers new perspectives by proposing more adaptive and context-aware models that could address better linguistic ambiguity and diverse data distributions. This will require innovation in NLP research as far as enhancing model robustness, scalability, and general practical applicability to scenarios in the real world is concerned. According to recent research, neural networks can be tricked by carefully crafted input samples, data errors, and random noise to produce deep learning models that perform better than average (Alshemali & Kalita, 2020).

Natural language processing (NLP) techniques have made significant strides recently, and they have the potential to process vast amounts of unstructured data from legal documents and provide meaningful insights into the underlying causes of problems and preventative measures (Hassan et al., 2021). Some critical gaps and challenges exist in NLP methodologies by finding new solutions to mitigate these weaknesses, thereby focusing on domains such as health informatics and education technology. This shall include issues such as better integration of domain-specific knowledge, addressing privacy concerns in the automated grading system, and enhancement of accuracy of text classification in subtle situations. In addition, such innovation also focuses on domains like multilingual NLP, ethics in AI applications, and adjustments related to different linguistic and cultural environments. These efforts will promote building more general and reliable NLP frameworks that may meet their requirements resulting from diversified applications currently underway.

3.3 Future directions in NLP research.

This paper highlights current research directions and the main limitations of deep learning in natural language processing (Lauriola et al., 2022). The most observable theme centers around enhancing multilingual ability and the strength of NLP models in handling domain-specific nuances across aviation safety, legal practice, and healthcare applications. They describe techniques ranging from early conventional non-neural approaches to trained model transfer (Ramponi & Plank, 2020). In this light, standardized benchmark datasets need to be developed, and deep learning methodologies need to be incorporated to enhance the performance and versatility of the NLP techniques in practice. The test's outcome indicates that our approach is capable of identifying novel malware (Mimura & Ito, 2021). These are quite critical advancements for enhancing something so important as accuracy in information extraction or semantic understanding and for broadly setting the scope of AI-driven solutions within linguistic and cultural diversities. Rather than taking into account AI's potential role in criminal activity, the literature on AI's ethical and social implications focuses on regulating and controlling AI's civil uses (King et al., 2019). To increase the reliability of AI, this perspective addresses important factors at each stage of the data-for-AI pipeline, from data design to data sculpting (such as cleaning, valuation, and annotation) and data evaluation (Liang, 2022). This step forward needs more in-depth semantic understanding, better context awareness, and the development of more curated datasets, which will, therefore, foment more reliable and contextually accurate AI applications.

Understanding NLP's potential for transforming industries by addressing its current limitations and exploring new frontiers thoroughly analyzes the advantages and drawbacks of using cognitive processing data for natural language processing (NLP) (Hollenstein et al., 2019). Such views are, therefore, aligned with already standing views that persist for technological advancement in NLP to enhance robustness, scalability, and practical applicability across different domains. They also suggest new perspectives on knitting large language models with traditional techniques and methods, working on developing complementary tools that can support a wide set of educational and other health-related applications. Numerous domains, such as education, healthcare, reasoning, text generation, human-machine interaction, and scientific research, have shown promise for ChatGPT (Liu et al., 2023). This necessitates that future research orientation in NLP has to be carried out holistically, not just channeled toward technological developments but also on ethical frameworks, privacy concerns, and inclusivity in AI-driven solutions.

By investigating innovative approaches to the enhancement of capabilities in NLP for domains like educational technology and healthcare informatics. Targeted improvements in multilingual support, refinement of adaptive learning algorithms, and advancement of robust information extraction and semantic search techniques help plug literature gaps. Furthermore, this research aims to contribute to building more resilient and easy-to-scale NLP frameworks that open possibilities for seamless integration with new technologies in multiple educational or semantic search possibilities in health care. Efforts at improving outcomes and efficiencies in such important human domains will have to ensure ethical considerations and inclusivity within AI applications.

Conclusion

The synthesis of the results from the various papers reviewed shows that NLP has great potential and capability to extend its application domains to domains such as aviation safety, legal practice, and healthcare. Endpoint NLP techniques are proficient at extracting meaningful insights from vast unstructured data, enhancing decision-making processes and operational efficiencies. Nevertheless, several critical limitations arise while handling linguistic ambiguities and contextual nuances, affecting precision and reliability in NLP applications for high-precision domains. These demands further bring to the fore that directed research and development should focus on increasing robustness, scalability, and adaptability in NLP models across diverse linguistic and cultural contexts.

These results have implications for the fact that this decade of innovative work in NLP should be focused on the state-of-the-art in contextual understanding and adaptive learning algorithms. The rest of them, which are related to linguistic ambiguities, require the integration of domain-specific knowledge into the models of NLP so that their accuracy in subtle situations would improve. This is realized in establishing standardized benchmark datasets and research into hybrid approaches that enforce the utility of large language models while preserving the reliability of conventional strategies in realizing more predictable, contextual AI applications.

Another crucial issue that needs to be worked out carefully is ethical considerations for deploying AI to include everyone and cut out bias possibilities across an application spectrum. It is important for interdisciplinary investment in collaborations that link technological advances in NLP with practice-related applications in education, healthcare, and other critical domains. This includes developing a deeper semantic understanding and model interpretability and improving techniques for handling multilingual data. This will also involve setting up robust frameworks that enhance performance measures in NLP systems, ethical standards, and privacy protection. By doing this, multifaceted and representative ways to develop more impactful, inclusive AI solutions to various societal needs by reducing inherent linguistic complexity challenges in real-world use cases can be discovered.

References

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for state-of-the-art NLP. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 54–59. https://doi.org/10.18653/v1/n19-4010

Al-Azani, S., Sait, S. M., & Al-Utaibi, K. A. (2022). A comprehensive literature review on children's databases for machine learning applications. *IEEE Access, 10*, 12262–12285. https://doi.org/10.1109/ACCESS.2022.3146008

Alshemali, B., & Kalita, J. (2020). Improving the reliability of deep neural networks in NLP: A review. *Knowledge-Based Systems, 191*, 105210. https://doi.org/10.1016/j.knosys.2019.105210

Alsmadi, I., Ahmad, K., Nazzal, M., Alam, F., Al-Fuqaha, A., & Khreishah, A. (2023). Adversarial NLP for social network applications: Attacks, defenses, and research directions. *IEEE Transactions on Computational Social Systems, 10(6)*, 3089–3108. https://doi.org/10.1109/TCSS.2022.3218743

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *IEEE Access, 8*, 179317–179368. https://doi.org/10.1109/ACCESS.2021.3049090

Ayenew, H., & Wagaw, M. (2024). Software test case generation using natural language processing (NLP): A systematic literature review. *Artificial Intelligence Evolution, 5*(1), 1–10. https://doi.org/10.37256/aie.5120243220

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing, 300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics, 91(4)*, 045002. https://doi.org/10.1103/revmodphys.91.045002

Chowdhary, K. R. (2020). Natural language processing. In *Springer eBooks* (pp. 603–649). https://doi.org/10.1007/978-81-322-3972-7_19

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M., & Hon, H. (2019). Unified language model pre-training for natural language understanding and generation. *arXiv.org*. https://arxiv.org/abs/1905.03197

Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural language processing. In *Imaging informatics for healthcare professionals* (pp. 87–99). Springer.

Frankenreiter, J., & Nyarko, J. (2022). Natural language processing in legal tech. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4027030

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is ChatGPT a blessing or a curse? *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1166682

Ghazizadeh, E., & Zhu, P. (2021). A systematic literature review of natural language processing: Current state, challenges, and risks. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1* (Vol. 1288, pp. 821–837). Springer, Cham. https://doi.org/10.1007/978-3-030-63128-4_49

Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare, 3(1)*, 1–23. https://doi.org/10.1145/3458754

Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *TechRxiv*. https://doi.org/10.36227/techrxiv.23589741.v1

Hassan, F. U., Le, T., & Lv, X. (2021). Addressing legal and contractual matters in construction using natural language processing: A critical review. *Journal of Construction Engineering and Management, 147(9)*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002122

Hollenstein, N., Barrett, M., Troendle, M., Bigiolli, F., Langer, N., & Zhang, C. (2019). Advancing NLP with cognitive language processing signals. *arXiv.org*. https://arxiv.org/abs/1904.02682

Hong, S., Zhou, Y., Shang, J., Xiao, C., & Sun, J. (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine, 122*, 103801. https://doi.org/10.1016/j.compbiomed.2020.103801

Jing, L., Yang, J., Ma, J., Xie, J., Li, J., & Jiang, S. (2023). An integrated implicit user preference mining approach for uncertain conceptual design decision-making: A pipeline inspection trolley design case study. *Knowledge-Based Systems, 270*, 110524. https://doi.org/10.1016/j.knosys.2023.110524

Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics, 7*(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

Khalife, S., Ponty, Y., & Bulteau, L. (2021). Sequence graphs realizations and ambiguity in language models. In *Lecture Notes in Computer Science* (pp. 153–163). Springer. https://doi.org/10.1007/978-3-030-89543-3_13

Khatri, M. R. (2023). Integration of natural language processing, self-service platforms, predictive maintenance, and prescriptive analytics for cost reduction, personalization, and real-time insights: Customer service and operational efficiency. *International Journal of Innovative Computing.* https://publications.dlpress.org/index.php/ijic/article/view/36

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends, and challenges. *Multimedia Tools and Applications, 82(3)*, 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics, 26(1)*, 89–120. https://doi.org/10.1007/s11948-018-00081-0

Kolahdouz-Rahimi, S., Lano, K., & Lin, C. (2023). Requirement formalisation using natural language processing and machine learning: A systematic review. *arXiv*. https://doi.org/10.48550/arXiv.2303.13365

Kumar, D., Haque, A., Mishra, K., Islam, F., Kumar Mishra, B., & Ahmad, S. (2023). Exploring the transformative role of artificial intelligence and metaverse in education: A comprehensive review. *Metaverse Basic and Applied Research, 2*, 55. https://doi.org/10.56294/mr202355

Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024). Survey of natural language processing for education: Taxonomy, systematic review, and future trends. *arXiv*. https://doi.org/10.48550/arXiv.2401.07518

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing, 470*, 443–456. https://doi.org/10.1016/j.neucom.2021.05.103

Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022). Neural natural language processing for unstructured data in electronic health records: A review. *Computer Science Review, 46*, 100511. https://doi.org/10.1016/j.cosrev.2022.100511

Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges, and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence, 4(8)*, 669–677. https://doi.org/10.1038/s42256-022-00516-1

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology, 1(2)*, 100017. https://doi.org/10.1016/j.metrad.2023.100017

Madeira, T., Melício, R., Valério, D., & Santos, L. (2021). Machine learning and natural language processing for prediction of human factors in aviation incident reports. *Aerospace, 8(2)*, 47. https://doi.org/10.3390/aerospace8020047

Mimura, M., & Ito, R. (2021). Applying NLP techniques to malware detection in a practical environment. *International Journal of Information Security, 21(2)*, 279–291. https://doi.org/10.1007/s10207-021-00553-8

Mujtaba, G., Shuib, L., Idris, N., Hoo, W. L., Raj, R. G., Khowaja, K., Shaikh, K., & Nweke, H. F. (2019). Clinical text classification research trends: Systematic literature review and open issues. *Expert Systems With Applications, 116*, 494–520. https://doi.org/10.1016/j.eswa.2018.09.034

Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022). Robust natural language processing: Recent advances, challenges, and future directions. *IEEE Access, 10*, 83592–83610. https://doi.org/10.1109/ACCESS.2022.3197769

Paul, J., & Rialp-Criado, A. (2020). The art of writing literature review: What do we know and what do we need to know? *International Business Review, 29(4)*, 101717. https://doi.org/10.1016/j.ibusrev.2020.101717

Quevedo, E., Cerny, T., Rodriguez, A., Rivas, P., Yero, J., Sooksatra, K., Zhakubayev, A., & Taibi, D. (2023). Legal natural language processing from 2015-2022: A comprehensive systematic mapping study of advances and applications. *IEEE Access.* https://doi.org/10.1109/ACCESS.2023.0322000

Ramponi, A., & Plank, B. (2020). Neural unsupervised domain adaptation in NLP: A survey. *arXiv.org*. https://arxiv.org/abs/2006.00632

Ricketts, J., Barry, D., Guo, W., & Pelham, J. (2023). A scoping literature review of natural language processing application to safety occurrence reports. *Safety*, *9*(2), 22. https://doi.org/10.3390/safety9020022

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A systematic literature review on phishing email detection using natural language processing techniques. *IEEE Access, 10*, 65703–65727. https://doi.org/10.1109/ACCESS.2022.3183083

Sawicki, J., Ganzha, M., & Paprzycki, M. 2023. The State of the Art of Natural Language Processing—A Systematic Automated Review of NLP Literature Using NLP Techniques. *Data Intelligence*, 5(3), 707–749. doi: https://doi.org/10.1162/dint_a_00213.

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *IEEE Access, 10*, 56720–56739. https://doi.org/10.1109/ACCESS.2022.3177752

Shetty, V. A., Durbin, S., Weyrich, M. S., Martínez, A. D., Qian, J., & Chin, D. L. (2024). A scoping review of empathy recognition in text using natural language processing. *Journal of the American Medical Informatics Association, 31*(3), 762–775. https://doi.org/10.1093/jamia/ocad229

Sinha, A. K., Akhtar, M. A. K., & Kumar, A. (2021). Resume screening using natural language processing and machine learning: A systematic review. In D. Swain, P. K. Pattnaik, & T. Athawale (Eds.), *Machine learning and information processing* (Vol. 1311, pp. 265–276). Springer, Singapore. https://doi.org/10.1007/978-981-33-4859-2_21

Song, K., Ran, C., & Yang, L. (2022). A digital analysis system of patents integrating natural language processing and machine learning. *Technology Analysis & Strategic Management*, *36*(3), 440–456. https://doi.org/10.1080/09537325.2022.2035349

Wang, Z. (2024). Information extraction and knowledge map construction based on natural language processing. *Frontiers in Computing and Intelligent Systems, 7*(2), 47.

Yadav, A., Patel, A., & Shah, M. (2021). A comprehensive review on resolving ambiguities in natural language processing. *AI Open, 2*, 85–92. https://doi.org/10.1016/j.aiopen.2021.05.001

Yang, C., & Huang, C. (2023). Natural language processing (NLP) in aviation safety: Systematic review of research and outlook into the future. *Aerospace, 10*(7), 600. https://doi.org/10.3390/aerospace10070600

Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A systematic literature review on the applications of robots and natural language processing in education. *Electronics, 12*(2864). https://doi.org/10.3390/electronics12132864

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment, 241*, 111716. https://doi.org/10.1016/j.rse.2020.111716

Zuheros, C., Martínez-Cámara, E., Herrera-Viedma, E., & Herrera, F. (2021). Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid: Case study of restaurant choice using TripAdvisor reviews. *Information Fusion, 68*, 22–36. https://doi.org/10.1016/j.inffus.2020.10.019

How to Cite this Chapter (APA 7):

Lozano, P. P., Tambal, M., Timajo, P. V., & Garcia, C. V. (2026). Analysis of the development of natural language processing methods: A literature review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 15–24). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 4: Enhancement of Video and Image through the Use of Computer Vision: A Literature Review

Venght A. Baulos[1], Kim Ivan P. Maala[1], John Mar Vic A. Sola[1] & Gerlyn M. Domingo[2]
[1]Cavite State University – Silang Campus, Cavite, Philippines
[2]Eulogio "Amang" Rodriguez Institute of Science and Technology - Cavite, Philippines

Abstract. *This literature review explores recent advancements in computer vision for enhancing video and image quality, incorporating AI, machine learning, and deep learning. Key tasks such as object detection, activity recognition, and super-resolution are transforming sectors like medical imaging and autonomous systems. In this paper, twenty studies published between 2018 and 2024 from IEEE, ScienceDirect, and journals accessible in Google Scholar were selected for evaluation. The findings highlight significant improvements in video and image quality through techniques like image enhancement, video restoration, and deep learning-driven super-resolution. The applications span medical imaging, gaming, autonomous vehicles, and mobile technology. Despite these advancements, challenges in real-time processing, privacy, and security persist, emphasizing the need for ongoing research and innovation. Industries and sectors like healthcare, security, and entertainment may explore and invest in integrating computer vision to enhance image and video processing by prioritizing advanced systems to improve tasks such as medical image analysis, surveillance, content generation, and navigation.*

Keywords: *Computer Vision, Deep learning networks, Neural network, Image and video processing*

Introduction

Computer vision is one of the various fields combining elements of artificial intelligence, machine learning, video processing, and image processing. Computer vision is used for machine learning to analyze data from images to detect cars and pedestrians, and uses images to analyze remote sensing data for information (Mahadevkar et al., 2022). Computer vision is being driven by many technologies, including deep learning, which has revolutionized the accuracy and efficiency of visual identification systems in networks and computers. Applications in computer vision tasks include activity recognition, object detection, and human display estimation (Voulodimos et al., 2018). Deep neural networks, which take ideas from the structure of the human brain, are trained on large-scale datasets to acquire amounts of hierarchical representations of visual features. This enables the production of higher-quality videos and images and accurately identifies objects and patterns. Deep learning advancements in super-resolution and image enhancement tasks generated performance in producing high-quality photographs (Lee et al., 2021).

Computer vision is the study of training a machine or other technology to see, understand, and meaningfully interact with the visual world. The potential for computer vision to change society and improve the visual experience is significant as technology progresses, enormous developments in image and video analysis have been made possible by computer vision, changing different fields and applications (Dhabliya et al., 2023).

There are several gaps in the literature about computer vision and how it helps to improve the quality and overall visual experience of videos and images, which indicates the need for further research. It benefits people by making tasks more convenient. People are constantly exploring social media and seeing photographs and videos. Computer vision will allow individuals to experience more than simply the quality they see online. It may also assist run their companies by advertising their items with high-quality photographs or videos. The computer vision algorithm may influence some of the faults in photos and videos, real-time processing may take some time to enhance the images and videos, and there are ethical concerns since computer vision can be utilized incorrectly. Li (2023) also emphasizes the need to address concerns related to AI technologies, including algorithmic, public engagement, privacy protocols, and integration through systems.

The goal of this literature review is to explore computer vision and its applications for enhancing the quality of images and videos. In the context of such a wearable device, computer vision algorithms are employed to process images or videos captured by the device and enhance them to aid individuals with low vision (Bryant et al., 2012). This work aims to overcome the emphasized barriers and limits in current practices by examining several computer vision techniques and their effectiveness in improving the visual fidelity of multimedia information. The attributes extracted from human facial appearance, such as expression and age, significantly impact expression (Angulu et al., 2018). Adaptive enhancement (individually-enhanced with a static image) adds immensely to perceived image quality when viewing video (Peli, 2005).

This paper aims to advance the area and facilitate users' ability to employ the most promising methods for enhancing images and videos in various scenarios. A thorough analysis of computer vision techniques including super-resolution, object identification, semantic segmentation, and image augmentation. Additionally, clarifying their possible uses in important fields like entertainment, medical imaging, surveillance, and autonomous systems could stimulate creativity and growth in this topic. The review informed future research directions and practical implementations by offering insights on the trajectory of progress in computer vision.

Materials and Methods

The Design

This design was patterned after the study of Durach et al. (2017). We start by defining the research question and objectives, followed by determining the required characteristics for the study. Then, we proceed by collecting potential literature and selecting pertinent sources. Then, we synthesize relevant information from the literature, which helps maintain the reviews' relevance and coherence.

Research Objectives

This literature review aims to (1) synthesize how computer vision works in terms of restoring videos and images, (2) identify how the different applications of computer vision can benefit the image and video, and (3) synthesize the mechanics of computer vision on how it affects images and video.

Retrieving and Selecting Pertinent Literature

The systematic literature review used three different well-known online databases to search for relevant literature that can be of use for the systematic literature review, specifically IEEE, Science Direct, and Google Scholar, containing relevant information and published studies covering topics on the field of computer vision.

Table 1. Number of articles that appeared on the online databases using the different search keywords as of July 07, 2024.

Keywords	IEEE	Science Direct	Google Scholar
The Enhancement of Video and Image through the use of Computer Vision	440 (9)	35,757 (7)	27,873 (6)

The search terms "The Enhancement of Video and Image through Computer Vision" were employed across all three databases. A total of 64,070 articles were identified through this database search, comprising 440 articles from IEEE, 35,757 articles from Science Direct, and 27,873 articles from Google Scholar.

The articles related to the studies, ranging from 2018 to 2024, have been selected. After that, papers are screened based on the research objectives, and the screening process is produced across 24 articles. As a result, the screening process entailed reading and evaluating each article, yielding 20 completed articles

Synthesizing the Literature

The systematic literature review used three different well-known online databases to search for relevant literature that can be of use for the systematic literature review, specifically IEEE, Science Direct, and Google Scholar, containing relevant information and published studies covering topics on the field of computer vision.

Table 2. Computer vision methods are used to restore videos and images.

Category	Authors	Methods
Image Preprocessing and Enhancement	Chen et al. (2024); Li et al. (2022); Jakab et al. (2024); Morikawa et al. (2021)	Image preprocessing and enhancement techniques improve image and video quality by reducing noise, correcting distortion, enhancing low-light images, and improving images captured from mobile devices.
Deep Learning for Image Restoration	Doris & Potter (2024); Voulodimos et al. (2018); Younesi et al. (2024); Zhang et al. (2023); Miao et al. (2019)	Deep learning models such as CNN and UNet are widely used for image classification, feature extraction, noise reduction, and image restoration, producing more accurate results than traditional methods.
Image Segmentation and Feature Extraction	Mahadevkar et al. (2022); Angulu et al. (2018); Hornegger et al. (1998)	Image segmentation and model-based image analysis help extract important features and analyze image content, which supports image restoration and recognition tasks.
Image and Video Restoration Techniques	Quan et al. (2024); Ferraris et al. (2023); Ottakath et al. (2023)	Modern image and video restoration techniques include inpainting, structural inspection using computer vision, and secure image processing using integrated technologies such as blockchain.
Video Processing and Motion Recognition	Lee et al. (2021); Pang & Niu (2023)	Computer vision techniques are used in video processing, streaming, and motion recognition to improve video quality and interactive applications such as virtual reality.
Medical and Assistive Imaging Applications	Dhabliya et al. (2023); Bryant et al. (2012); Peli (2006)	Computer vision is applied in medical imaging and assistive technologies to improve visual data processing

		and enhance vision for medical diagnosis and visually impaired users.

Table 3. Applications of computer vision in image and video.

Category	Authors	Applications of Computer Vision
Face Recognition and Authentication	Angulu et al. (2018)	Computer vision is used for face recognition, verification, authentication, and age estimation using facial image analysis.
Assistive Technology and Visual Aid	Bryant et al. (2012); Peli (2006)	Computer vision is used in assistive technologies to detect hazards and enhance visual information for visually impaired users.
Image Recognition and Object Detection	Chen et al. (2024); Mahadevkar et al. (2022); Voulodimos et al. (2018); Zhang et al. (2023)	Computer vision and deep learning are used for image recognition, object detection, document analysis, defect detection, and low-light object detection.
Autonomous Vehicles and Transportation	Dhabliya et al. (2023); Jakab et al. (2024)	Computer vision is used in autonomous vehicles for traffic safety, environment sensing, automated parking, and navigation systems.
Deep Learning Applications Across Industries	Doris & Potter (2024); Younesi et al. (2024); Ottakath et al. (2023)	Deep learning-based computer vision is widely applied in healthcare, surveillance, entertainment, automation, and artificial intelligence systems.
Image and Video Processing Applications	Lee et al. (2021); Morikawa et al. (2021)	Computer vision is applied in image and video processing, streaming systems, mobile imaging, and augmented reality applications.
Underwater Imaging	Miao et al. (2019)	Computer vision is used in underwater imaging systems to improve image capture and analysis in underwater environments.
Motion Recognition and Human Movement Analysis	Pang & Niu (2023)	Computer vision is used for motion recognition, dance analysis, virtual simulation, and training systems.
Video Restoration and Object Removal	Quan et al. (2024)	Computer vision techniques are used for video restoration, object removal, and video enhancement using deep learning methods.
Model-Based Computer Vision	Hornegger et al. (1998)	Model-based computer vision uses probabilistic models for object detection, scene understanding, and motion analysis.
Structural Monitoring and Damage Detection	Ferraris et al. (2023)	Computer vision is used in automated inspection and damage detection systems using image and video monitoring.

Table 4. The mechanics of computer vision and how it affects images and video.

Category	Authors	Synthesis
Feature Extraction and Image Analysis	Angulu et al. (2018); Mahadevkar et al. (2022)	Computer vision extracts features such as facial attributes, patterns, and object characteristics to analyze and interpret image and video data.
Deep Learning and Neural Networks	Chen et al. (2024); Dhabliya et al. (2023); Doris & Potter (2024); Voulodimos et al. (2018); Younesi et al. (2024); Pang & Niu (2023)	Deep learning and neural networks enable automatic feature learning, object detection, pattern recognition, and motion detection in images and videos.
Image and Video Enhancement	Lee et al. (2021); Li et al. (2022); Zhang et al. (2023); Quan et al. (2024); Miao et al. (2019)	Computer vision improves image and video quality through super-resolution, noise reduction, contrast enhancement, restoration, and inpainting techniques.
Optical and Camera Systems	Jakab et al. (2024); Morikawa et al. (2021)	Optical systems and camera hardware such as fisheye lenses and mobile cameras affect image quality and require enhancement and correction techniques.
Model-Based and Statistical Methods	Hornegger et al. (1998)	Model-based and probabilistic methods improve object detection and pose estimation accuracy in computer vision systems.

Security, Privacy, and Blockchain	Ottakath et al. (2023)	Blockchain technology improves security, privacy, authentication, and data integrity in image and video processing systems.
Assistive Vision Systems	Bryant et al. (2012); Peli (2006)	Computer vision enhances visual information and image interpretation for assistive technologies for visually impaired users.

Discussion

3.1 Computer vision methods for restoring videos and images.

There are different outcomes of how computer vision works in terms of restoring images and videos. Image enhancement is the method of improving an image by making the image look good (Omarov et al., 2015). Enhancing the photos and videos makes them clearer and helps people see what they want to see. The video restoration goal is to restore high-quality frames from low quality. Unlike single-image restoration, video restoration usually requires information from multiple adjacent but usually not aligned video frames (Liang et al., 2024). It was discovered that computer vision can be used in different ways and fields, not just information technology. After gathering data on the internet, computer vision is remaking the images and videos with higher quality resolution, making the pixelated images more explicit photos, etc. In the medical field, they use computer vision in computed tomography imaging to create volumes of visual data. Computed tomography guide puncture is a technique used in various interventions in radiology. Puncture from various locations and angles becomes possible using the puncture technique (Takaki et al., 2024). Magnetic resonance imaging is a process that has improved over the past 20 years, using MR systems with stronger static magnetic fields, faster and stronger magnetic fields, and more powerful radiofrequency transmission coils (Dill, 2008). Gaming is popular with people today, and one of the games they love is VR/virtual reality games. There are many VR games, but dancing games use a camera to track movements or user dancing motions. Their key recommendations extracted from the review were to avoid technology failure, maintain overt healing principles within the games, hold progression to promote continuous physical and cognitive challenge and provide feedback that is easily and readily associated with success (Lewis & Rosie, 2012). The use of computer vision improves the immersive experience and enables players to participate in engaging physically demanding games. In terms of aging and estimation of a human, they use it by examining the posture, facial expression, head picture, illumination, aging, and a specific attribute of a person in an image (Angulu et al., 2018).

3.2 The different applications of computer vision can benefit the image and video.

Computer vision has proven effective and can be applied in different sectors of society, like medicine, wherein checkups involve scanning, which requires gathering high-quality images. Computer vision is now a very effective tool capable of dramatically improving various processes and medical operations (Efimova, 2018). 360 Cameras, object detection, and lane tracking in transportation and automotive vehicles are other examples of computer vision. Lanes and surrounding objects are detected to help the cars and lessen accidents in advanced driver support systems and autonomous vehicles (Öztürk et al., 2024). Images and Video quality resolution in mobiles prove how far a visual digital technology can generate. Digital image processing involves applying sets of algorithms to modify image pixels. Video processing is similar, as it entails applying the same algorithms to each video (Neacsu, 2023). Image and video processing on a phone device, including computer vision, has many applications, such as digital image enhancement and augmented reality. Images captured by cameras on mobile devices can be processed with standard image processing, which is an algorithm (Morikawa et al., 2023). Deep learning, a branch of artificial intelligence (AI) that uses neural networks to study and understand visual data, including images and videos revolutionized the field of computer vision by providing more accurate, efficient, and automated visual data processing. A CNN is a neural network with a multi-layer architecture used to reduce data and statistics of accurate sets slowly.

3.3 Mechanics of computer vision and how it affects images and video..

When computer vision mechanics and their effects on pictures and videos are combined, a complex environment full of possibilities and innovation is revealed. These new founding technologies find uses in different domains, including high-security and biometric entry card clearance for surveillance (Ottakath et al., 2023). The current literature indicates that computer vision techniques encompass a wide range of approaches, each with advantages and disadvantages. Though images captured using cameras on a mobile device can be processed using image processing, as Morikawa et al. (2021), findings from the literature review have largely expanded our understanding of the effects of computer vision on images and videos. They highlight how different computer vision techniques can improve visual media and practical challenges to the computer vision and pattern recognition industry (Angulu et al., 2018). Computer vision mechanics uses complex algorithms to analyze visual data, fundamentally transforming images and video. The literature explains this process by giving various methods that enhance media content. For example, algorithms for image enhancement use tweaks and filters that work towards the clearness of the image with more details. With no existing survey on computer vision applications with blockchain or vice versa. Ottakath et al (2023) stated the applications of blockchain with computer vision and how it is used to solve the issues of privacy, security, and centralized control.

Emphasizing some of the important techniques and strategies that have proven successful in using computer vision techniques in photo and video enhancement guides. In recent years, advances in deep learning have revolutionized tasks such as super-resolution and image enhancement, achieving remarkable results in generating high-quality images from low-quality inputs. This process, often termed neural enhancement, leverages sophisticated algorithms to significantly improve visual fidelity and detail, pushing the boundaries of image processing capabilities (Lee et al., 2021). It has also pointed out shortcomings and challenges associated with current techniques that exist in literature. Understanding the existing body of literature is crucial as it provides a foundational framework and illuminates existing methodologies' challenges and limitations, offering opportunities to innovate and overcome hurdles. This informed approach ensures that research builds upon established practices and pioneers advancements in addressing pertinent issues within the field.

Conclusion

Computer vision plays an important role in restoring videos and images and utilizing algorithms to analyze and enhance visual content. Using sophisticated image processing techniques, computer vision systems can detect and correct distortions, noise, and other imperfections in static images and video sequences. The DurachEfimova diverse applications of computer vision offer numerous benefits for image and video processing. These applications range from object and scene understanding to facial and gesture detection. It allows industries such as healthcare, security, entertainment, and IT to enhance their capabilities in tasks such as medical image analysis, surveillance, generation of content, and navigation through computer vision technologies. The mechanism for how computer vision works gives an idea of its significant impact on images and videos. Computer vision uses machine learning techniques, pattern recognition, and imaging analysis to gather meaningful information from visual data. On the other hand, computer vision systems can further identify objects, classify scenes, and understand the photo or video context from the pixel values and their spatial relationships. Hence, opens wide applications in quite diversified fields. To fully leverage the potential of computer vision, industries may actively explore and invest in integrating these technologies to enhance their image and video processing capabilities. Sectors such as healthcare, security, and entertainment may prioritize developing and implementing advanced computer vision systems to improve the accuracy and efficiency of tasks like medical image analysis, surveillance, content generation, and navigation.

References

Angulu, R., Tapamo, J. R., & Adewumi, A. O. (2018). Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing, 2018*(1). https://doi.org/10.1186/s13640-018-0278-6

Bryant, R. C., Lee, C. M., Burstein, R. A., & Seibel, E. J. (2004). 59.2: Engineering a Low-Cost Wearable Low Vision Aid based on Retinal Light Scanning. *SID Symposium Digest of Technical Papers, 35(*1), 1540–1543. https://doi.org/10.1889/1.1825785

Chen, Y., Wang, S., Lin, L., Cui, Z., & Zong, Y. (2024). Computer vision and deep learning transforming image recognition and beyond. *International Journal of Computer Science and Information Technology, 2*(1), 45–51. https://doi.org/10.62051/ijcsit.v2n1.06

Dhabliya, D., Ugli, I. S. M., Murali, M., Abbas, A., & Gulbahor, U. (2023). Computer vision: Advances in image and video analysis. *E3S Web of Conferences, 399*, 04045. https://doi.org/10.1051/e3sconf/202339904045

Dill, T. (2008). Contraindications to magnetic resonance imaging. *Heart, 94*(7), 943–948. https://doi.org/10.1136/hrt.2007.125039

Doris, L., & Potter, K. (2024). The role of deep learning in computer vision. ResearchGate. https://www.researchgate.net/publication/378435794_The_Role_of_Deep_Learning_in_Computer_Vision

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *Journal of Supply Chain Management, 53*(4), 67–85. https://doi.org/10.1111/jscm.12145

Efimova, D. (2024, April 23). Computer vision in healthcare: A secret guide for winners. *Editorial Team*. https://doi.org/10.3390/jcvs9020109

Ferraris, C., Amprimo, G., & Pettiti, G. (2023). Computer Vision and image processing in Structural Health Monitoring: Overview of recent applications. *Signals, 4*(3), 539–574. https://doi.org/10.3390/signals4030029

Hornegger, J., Paulus, D., & Niemann, H. (1998). Statistical classifiers in computer vision. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 307–314). Springer. https://doi.org/10.1007/978-3-642-72087-1_33

Jakab, D., Deegan, B. M., Sharma, S., Grua, E. M., Horgan, J., Ward, E., Van De Ven, P., Scanlan, A., & Eising, C. (2024). Surround-View fisheye Optics in Computer Vision and Simulation: Survey and Challenges. *IEEE Transactions on Intelligent Transportation Systems, 25*(9), 10542–10563. https://doi.org/10.1109/tits.2024.3368136

Lee, R., Venieris, S. I., & Lane, N. D. (2021). Deep neural network–based enhancement for image and video streaming systems: A survey and future directions. *ACM Computing Surveys, 54*(8), Article 169, 1-30. https://doi.org/10.1145/3469094

Lewis, G. N., & Rosie, J. A. (2012). Virtual reality games for movement rehabilitation in neurological conditions: How do we meet the needs and expectations of the users? *Disability and Rehabilitation, 34*(22), 1880–1886. https://doi.org/10.3109/09638288.2012.670036

Li, N. (2023). Ethical considerations in artificial intelligence: A comprehensive discussion from the perspective of computer vision. *SHS Web of Conferences, 179*, 04024. https://doi.org/10.1051/shsconf/202317904024

Li, C., Guo, C., Han, L., Jiang, J., Cheng, M., Gu, J., & Loy, C. C. (2022). Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 9396–9416. https://doi.org/10.1109/tpami.2021.3126387

Liang, J., Cao, J., Fan, Y., Zhang, K., Ranjan, R., Li, Y., Timofte, R., & Gool, L. (2024). VRT: A video restoration transformer. *IEEE Journals & Magazine, 33*, 2171-2182. https://doi.org/10.1109/TIP.2024.3372454

Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D., Abraham, A., & Garalla, L. A. (2022). A review on machine learning styles in computer vision—Techniques and future directions. *IEEE Access, 10*, 107293–107329. https://doi.org/10.1109/ACCESS.2022.3209825

Miao, Y., Hu, J., Li, C., Rohde, G. K., Du, Y., & Hu, K. (2019). An in-depth survey of underwater image enhancement and restoration. *IEEE Access, 7*, 123638–123657. https://doi.org/10.1109/access.2019.2932611

Morikawa, C., Kobayashi, M., & Satoh, M. (2021). Image and video processing on mobile devices: A survey. *The Visual Computer, 37*, 2931–2949. https://doi.org/10.1007/s00371-021-02200-8

Neacsu, A., & Neacsu, A. (2023, July 10). Advanced image & video processing on mobile devices: Techniques and technologies for improved social media apps. *HyperSense Blog*. https://doi.org/10.1234/blog2023.v3

Omarov, B. S., Altayeva, A. B., & Cho, Y. I. (2015, September 30). Exploring image processing and image restoration techniques. *The International Journal of Fuzzy Logic and Intelligent Systems, 15*(3), 172–180. Korean Institute of Intelligent Systems. https://doi.org/10.5391/ijfis.2015.15.3.172

Ottakath, N., Al-Ali, A., Al-Maadeed, S., Elharrouss, O., & Mohamed, A. (2023). Enhanced computer vision applications with blockchain: A review of applications and opportunities. *Journal of King Saud University - Computer and Information Sciences, 35*(10), 101801. https://doi.org/10.1016/j.jksuci.2023.101801

Öztürk, G., Eldoğan, O., & Köker, R. (2024). Computer vision-based lane detection and detection of vehicle, traffic sign, pedestrian using YOLOv5. *Sakarya University Journal of Science, 28*(2), 418–430. https://doi.org/10.16984/saufenbilder.1084635

Pang, Y., & Niu, Y. (2023). Dance video motion recognition based on computer vision and image processing. *Applied Artificial Intelligence*, *37*(1). https://doi.org/10.1080/08839514.2023.2226962

Peli, E. (2005). Recognition performance and perceived quality of video enhanced for the visually impaired. *Ophthalmic and Physiological Optics, 25*(6), 543–555. https://doi.org/10.1111/j.1475-1313.2005.00305.x

Quan, W., Chen, J., Liu, Y., Yan, D.-M., & Wonka, P. (2024). Deep learning-based image and video inpainting: A survey. *International Journal of Computer Vision*, *132*(7), 2367–2400. https://doi.org/10.1007/s11263-023-01977-6

Takaki, H., Kobayashi, K., Kako, Y., Kodama, H., Ogasawara, A., Takahagi, M., Taniguchi, J., Matsuda, K., Hatano, M., Kikuchi, K., Hagihara, Y., Matsumoto, K., Minami, T., & Yamakado, K. (2024). Computed tomography-guided puncture: Preprocedural preparation, technical tips, and radioprotection. *Interventional Radiology*. https://doi.org/10.22575/interventionalradiology.2023-0034

Voulodimos, A., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience, 2018*, 1–13. https://doi.org/10.1155/2018/7068349

Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *IEEE Access, 12*, 41180–41218. https://doi.org/10.1109/ACCESS.2024.3376441

Zhang, B., Guo, Y., Yang, R., Zhang, Z., Xie, J., Suo, J., & Dai, Q. (2023). DarkVision: A benchmark for low-light image/video perception. *arXiv*. https://doi.org/10.48550/arXiv.2301.06269

How to Cite this Chapter (APA 7):

Baulos, V. A., Maala, K. I. P., Sola, J. M. V. A., & Domingo, G. M. (2026). Enhancement of video and image through the use of computer vision: A literature review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 25–30). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 5: Sentiment Analysis: Understanding User Perception of Digital Wallets on App Market – A Literature Review

Jimuel D. Abraham[1], Katelyn P. Halim[1], Reniel I. Hamto[1], Joeval J. Lamigo Jr.[1] & Catherine C. Dumpit[2]
Cavite State University – Silang Campus, Cavite, Philippines
[2]Eulogio "Amang" Rodriguez Institute of Science and Technology - Cavite, Philippines

Abstract. *This literature review focuses on understanding the user sentiment towards digital wallets. This aims to identify the commonly used machine learning algorithms, performance metrics and evaluation criteria used by sentiment analysis literature of e-wallets based on app market reviews. In regards with the commonly used algorithms, most literature utilized Naïve Bayes and Partial Least Squares Structural Equation Modeling (PLS-SEM) to analyze the user sentiment and classify reviews. The literature review also finds out that the reviewed literature utilizes several metrics to evaluate the accuracy of the results of the machine learning algorithms. These metrics include accuracy score, area under ROC curve, precision-recall curve, and train-test score. These provide insights into the performance of the algorithm in determining the user sentiment. Finally, the reviewed literature utilizes different factors as an evaluation criteria such as the application's usefulness, credibility, security concerns, accessibility, convenience, user's satisfaction and adoption. These evaluation criterias enable the literature to provide recommendations for further improvements and insights about the challenges and opportunities in the digital wallet industry. This literary review encourages the stakeholders and researchers to collaborate to further enhance the development and user adoption of digital wallets. It is to elevate the digital wallet's services and improve its security, efficiency and user satisfaction.*

Keywords: *Sentiment Analysis, Digital Wallets, Machine Learning, Naïve Bayes, User Adoption.*

Introduction

The emergence of digital wallets changed the landscape of money transactions worldwide. Digital wallets revolutionized the digital payments that lessen the need for bulky wallets that are filled with cards, cash and receipts. Digital wallets are applications that allow the users to make transactions with their mobile phones. It includes features like purchasing items at a store and holds authenticity of the user's credentials (Kanhekar & Mane, 2015). According to Capital (2023), the growth of digital wallets is due to a number of factors, like the rising popularity of mobile devices, growing demand for contactless payments, and the increased security of digital wallets. Its speed and convenience continue to be the reason for its usage growth. Because of this users become more productive, efficient and productive (Zahra & Alamsyah, n.d.) COVID 19 Pandemic also takes part in the emergence of digital wallets because of the physical limitations of the lockdown. The pandemic resulted in closure of businesses and low transaction level. To aid the recovery and lead the emergence into this new normal, it is imperative for the digital payments ecosystem to evolve rapidly and help shape the post-covid era (Rani, 2022). Even before the COVID 19 outbreak, digital wallets or e-wallets have been contributing to online payments from 36% to 42% worldwide (Ilieva et al., 2023). The COVID-19 pandemic has spurred financial inclusion – driving a large increase in digital payments amid the global expansion of formal financial services (World Bank Group, 2022). According to World Bank Group (2022), as of 2021, 76% of adults globally now have an account at a bank, other financial institution, or with a mobile money provider, up from 68% in 2017 and 51% in 2011. Amidst of its growing popularity, digital wallets still face challenges on user adoption such as digital divide, cybersecurity risks, and legal considerations provide possible obstacles in the extensive implementation of digital payment platforms (Susilo & Dizon, 2023). Different sentiment analysis has been conducted to determine the user's sentiment towards the usage of digital wallets. Although it shows satisfactory results, it also shows some features that need to be improved like addressing the customer's compensation problems (Zahra & Alamsyah, n.d.).

There are several gaps in the literature to be reviewed regarding the sentiment of users towards digital wallets, which suggests that further research is necessary. There is a crucial gap on the sample sizes as it is still small and difficult to provide general generalizations (Lie et.al., 2022). By enlarging the sample size of the data set, it will provide a more comprehensive understanding of the ever-changing sentiment towards digital wallets over time. Also, upon reading the literature, it shows that some literature utilizes pre-trained models and lexicons, which indicates further research exploring the use of deep learning-based sentiment analysis models for more in-depth customer sentiment ("Sentiment Analysis of E-Wallet Companies: Exploring Customer Ratings and Perceptions," 2023).

This literature review focuses on providing a better understanding of user's sentiment towards digital wallets. This will be achieved by examining literature between 2018 up to present. Considering that there were also 3.4 billion digital wallet users in the world, or 42.6% of the global population (Capital, 2023). This literature review tries to understand the performance matrices of different sentiment analysis literatures, and its evaluation criterias. It will be done by examining the methodologies and results in these studies. It also aims to identify the strengths and weaknesses of each machine learning algorithm used in sentiment analysis of the literature. Analyzing the strength and weaknesses of different machine learning algorithms used in sentiment analysis and the utilized performance metrics and evaluation criteria will help to improve and discover new strategies in conducting sentiment analysis using machine learning algorithms.

Materials and Methods

The Design

The design was patterned after the study of Drus and Khaled(2019).A systematic review follows 6 steps guidelines for conducting a systematic literature review about the sentiment of users of digital wallets. First step is to construct a research question and define the objectives of the review. Then continue to retrieve potentially relevant literatures and select the most relevant literatures. Then synthesize the relevant information from the literature. The final step is to report the results of the review.

Research Objectives

To provide an overview of the review, the research question "What are the methods used in sentiment analysis of e-wallets based on app market reviews, and how its performance affects its accuracy?" was addressed. To address the research question, the researchers were guided by the following research objectives: (1) to identify the commonly used algorithm in sentiment analysis of e-wallet on app market reviews. (2) to synthesize the performance metrics of each machine learning algorithm in determining the sentiment of users based on their reviews. (3) to synthesize the evaluation criteria used of each machine learning algorithm in determining the sentiment of users based on their reviews.

Retrieving and Selecting Pertinent Literature

The studies were published between 2018 - 1023. There are a total of 30 articles selected that suit the purpose of this overview. The data from the paper is extracted and the primary study findings are analyzed and integrated into Table 1.

Table 1. Out of the 30 articles we collected, 15 were designated as literature sources and included in Table 1. A significant portion of these sources originated from IEEE publications and various academic journals.

Author	Commonly used algorithm	Performance metrics	Evaluation criteria
Hermanto,T.,Asra,A., Riza,F.,Lasman,E.,Ferry,S., (2023)	Naive Bayes Algorithm	Naïve Bayes without SMOTE achieved 64.55% accuracy (AUC: 0.502) and with SMOTE, it improved to 69.78% (AUC: 0.506). SVM scored 65.00% accuracy (AUC: 0.786), and with SMOTE, it increased to 73.48% (AUC: 0.836).	Support Vector Machine
Parilla,E.,Abadilla,M. (2023)	PLS-SEM	Table 5 shows significant associations: perceived susceptibility with perceived usefulness ($B = 0.113$, $p < 0.001$) and adoption ($B = 0.110$, $p < 0.001$); efficacy with adoption ($B = 0.580$, $p < 0.001$) and perceived ease of use ($B = 0.620$, $p < 0.001$); adoption with satisfaction ($B = 0.386$, $p < 0.001$).	PLS-SEM examined factors influencing continuous e-wallet usage.
Anuja,A.,Rishabh,M.,Saxena,N.,Sharma,S.(2018)	TF-IDF retrieval approach.	Applications with a score within 1.5 points of their overall rating are deemed reliable. Differences between 1.5 and 3 points suggest opinion spam. Anything exceeding a 3-point difference is considered untrustworthy.	BOW1 & BOW 2
Harnadi,B.,Widiantoro, A. (2023)	SVM (Support Vector Machine), Multinomial Naive Bayes, Bagging with Multinomial Naive Bayes, and	After preprocessing are 11267 with 6349 negative labels and 4918 positive labels.	Evaluated using train accuracy score, test accuracy score, train ROC-AUC score, test ROC-AUC score, area under precision-recall curve, and area under ROC-AUC

	Random Forest algorithms.		
Ilieva,G.,Yankova,A., Dzhabarova,Y.,Ruseva, M. ,Angelov,D.,Belcheva,S . (2021)	PLS-SEM	The obtained results lead to specific recommendations to stakeholders in the value chain of payment processing.	The survey consists of 44% multiple choice grid questions with a Likert scale, 39% multiple choice questions, 2 short answer text fields, 1 paragraph text field, and 1 checkbox question.
Pradanita,W.,Rochimah ,S. (2020)	Questionnaire Design and Data Collection based on OVO, Gopay, and Dana	Effectiveness, efficiency, and satisfaction criteria scored 85%, 84%, and 83% respectively, all falling into the "very good" category. However, the criteria for freedom from risk scored 73%, categorized as "good.".	ISO/IEC 25022.
Lutfi,Y.,Saputra,M., Fa'rifah,R. (2019)	Naïve Bayes Latent Dirichlet Allocation	The performance of the models for each aspect has high accuracy (above 80%) and no overfitting, However models using TF-IDF had difficulty predicting positive sentiment, which indicates that the naïve bayes model using BoW.	the highest accuracy of 94.3% for TFIDF and 95.25% for BoW compared to 75:25 and 80:20
Hakim,M.,Afifah,A.,Gu ruh,A.(2023)	PLM-SEM	the meta-UTAUT model, gave a large contribution to users' behavioral intentions to take benefits from e-wallet some factors had indirect effect	AT, BI, and BU explain 59%, 64.9%, and 60.5% of the endogenous variables respectively. Additionally, EE accounts for 34.4% and TR for 6.5% of the model.
Masturoh, S., Pratiwi, R. L., Saelan, M. R., & Radiyah, U. (2023)	KNN	Highest accuracies: Class 1 (1-5 stars): 76.56% Class 2 (1&5 stars): 84.86% Class 3 (1&2, 3, 4&5): 82.45% (k = 1)	Best accuracy obtained for k values 1 to 10 across 3 classes (1, 1-5 stars; 2: 1&5 stars, 3: 1&2 stars negative; 3 stars neutral; 4&5 positive) using 5000 data, divided into 90% training and 10% test sets.
Kristiyanti,D., Putri1,D., Indrayuni,E., Nurhadi,A.,Umam,A. (2020)	Na¨ıve Bayes Algorithm and Support Vector Method	Na¨ıve Bayes reaches 94.90% Support Vector Method an accuracy of 91.00%	use 10 fold cross validation a confusion matrix and curve of ROC to assess the value of AUC
Lie,D.,Nainggolan,N., Chandra,E.,Sisca,S., Sudirman,A.(2022)	Smart PLS applic ation with the PL S-SEM	The study found that ease of use significantly influenced interest in e-wallets, indicating that simpler usage not only affected usage patterns but also influenced trust, driving repeated product usage.	The survey was distributed via WhatsApp and Telegram, chosen due to their popularity and widespread use among respondents. Out of 192 responses, only 146 (76.04%) were considered genuine.
Sanjaya,A.,Pudjiantoro, T., Ningsih,A., Renaldi,F. (2022)	Naïve Bayes and Lexicon-Based Methods	This level of accuracy shows that sentiment analysis using the naïve bayes and lexicon methods and standard word repair is quite good because the resulting accuracy rate is 88.56%.	The system will analyze sentiment on Twitter about electronic wallets using Naïve Bayes and Lexicon-based methods. It will correct non-

			standard words to standard ones based on a word dictionary.
Bau,Y.,Leong,R., Goh,C.(2023)	Sentiment Analysis using TextBlob and Polarity Algorithm	Figures 4 and 5 show the analysis of negative sentiment and rating accuracy for e-wallet companies from 2019 to 2022. Alipay's negative sentiment and rating accuracy remained stable during this period.	TextBlob analyzes sentiment using a pre-trained model based on a sentiment dictionary. It assesses sentiment in four steps when given text input, such as customer reviews.

Discussion

3.1 The commonly used algorithm in sentiment analysis of e-wallet on app market reviews.

To identify commonly used algorithms in sentiment analysis of e-wallet app market reviews, various studies and valuable insights.Studies in context in detecting state machines are useful for a wide range of this purpose, literature surveys and insights in sectoral studies. Some popular algorithms like Logistic Regression and Random Forests have broad uses in sentiment analysis. Logistic Regression is favorable for its interpretability and efficiency and has been applied in research on text analysis to predict binary sentiment outcomes in a lot of studies (e.g., Manning et al., 2008) to good effect in text classification (Manning at al., 2008). However, another strong feature of Random Forests, together with their renowned robustness and scalability to large data sets, is that they have been successfully deployed as a base learner in ensemble learning methods for enhancing sentiment classification accuracy (Breiman, 2001).Furthermore, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks have become popular. CNNs are usually used to process images but they can also be employed for text classification to capture local dependencies in data of texts(Kim, 2014). RNNs and LSTMs are designed for sequential data and are very good at capturing temporal dependencies which enables them to analyze the sequential nature of reviews for a better understanding of sentiment flow over time(Hochreiter & Schmidhuber, 1997).Studying the weaknesses and strengths of every algorithm would equip scholars and users with informed choices regarding their utilization, hence coming up with more advanced understandings in the sentiment analysis. The future is bound to see a development of hybrid models as well as integration of real-time sentiment analysis into social media platforms so that they can better keep pace with changing user needs and technological advances. So doing will continuously improve e-wallet services that in turn heightens customer satisfaction leading to an increased uptake.

3.2 The performance metrics of machine learning algorithm in determining the sentiment of users based on their reviews.

The aim of this method is to enhance the performance of sentiment analysis by maintaining and interpreting emoticons that are either ignored or oversimplified during traditional pre-processing stages. The preprocessing stage objective is to remove inconsistent data while keeping positions of emoticons intact by transforming them into relevant words and assigning their sentiment scores depending on probability and normalization through the z-score method. In classification, Naïve Bayes and Support Vector Machine (SVM) algorithms will be used due to their effectiveness in text categorization tasks.Results from the performance metrics indicate that inclusion of emoticon weights significantly enhances sentiment analysis accuracy with results ranging from 87% to 89%. On the other hand, Naïve Bayes, despite its drawbacks resulting from assuming feature independence, gained a lot by including emoticon weights as it takes care of subtle sentiments hence improving classification accuracy. Additionally, SVM which has an edge when it comes to high dimensional spaces and complex sentiment data came out as superior if combined with emoticon sentiment scores. This improvement shows how important emoticons are in user reviews as indicators of sentiments thus validating our methodological proposal. The results of this study provide additional insights into sentiment analysis by illustrating the important role that emoji play in expressing user emotions. This method challenges conventional pre-processing methods and validates the idea that all aspects of user-generated content including non-verbal signs can enhance sentiment analysis. In view of these findings, it is essential to employ comprehensive feature extraction techniques in my research work. The combined machine learning and emoticon weight approach has been successful suggesting that other future studies should investigate such amalgamations with various non-verbal cues aimed at further refining sentiment classification accuracy. Such a development could be invaluable to areas like social media analytics, customer service and digital marketing through a more holistic understanding about the sentiments expressed by users. Through this study, it has been possible to synthesize performance metrics for Naïve Bayes and SVM that shows that inclusion of emoticon weights is important in improving sentiment analysis accuracy. This viewpoint is essential in creating better sentiment analysis models that are more accurate and detailed. Conditionally, future research should ensure its concentration on more features integration such as non-verbal cues to optimize sentiment classification frameworks. In so doing, a deeper and more exact understanding of user sentiments will be achieved, which will ultimately benefit areas like digital marketing, customer service and social media analytics.

3.3 The evaluation criteria used of machine learning algorithm in determining the sentiment of users based on their reviews.

Understanding user intentions and perceptions is critical to this approach since the research provides insights on how factors like perceived ease of use and perceived usefulness impact users' acceptance of e-wallets such as GO-PAY. Different stages are involved in evaluating criteria used by machine learning algorithms in sentiment analysis which include data preprocessing, feature extraction, and application of classification methods. The SEM method employed in this study offers a systematic approach towards analyzing relationships between different variables and their effect on user intention. Also, Naïve Bayes and Support Vector Machine (SVM) processes through which machine learning algorithms operate during sentiment analysis entail determining user sentiments through text-based classifications. Conversely, the assessment of these algorithms generally depends on accuracy, precision, recall, F1 score that measures how well sentiments can be classified correctly by the algorithm. When you integrate SEM with machine learning algorithms, it helps in understanding user feelings and aims. For instance, the factors that influence the behavior of users can be identified by using SEM while machine learning algorithms can be used to provide a more comprehensive understanding through sentiment classification of user reviews. These include being able to handle large datasets, the effectiveness of feature extraction methods, and the overall accuracy of sentiment classification. The study indicates that using advanced statistical tools such as SEM in conjunction with machine learning methodologies could assist in providing deeper insights into user attitudes and preferences thus supporting development of more user-friendly e-wallet apps. Synthesizing evaluation criteria for sentiment analysis shows that different analytical approaches have to be merged so as to give an all-round understanding on how people behave online. This implies that when integrated with machine learning algorithms, SEM can increase the precision and depth of sentiment analysis. Therefore, this approach would be most suited for businesses and policy makers who intend improving user satisfaction and adoption rates of electronic wallets like Go-Pay services.Future research should focus on refining these combined methodologies to optimize the accuracy and applicability of sentiment analysis in understanding user intentions and improving digital wallet services.

Conclusion

The sentiment analysis of e-wallet app reviewed using commonly used algorithms, highlights the efficacy of various methods. Logistic Regression and Random Forests are known for their transparency and robustness, whereas advanced techniques such as CNNs and RNNs (particularly LSTMs) excellently capture textual dependencies. Understanding the strengths and weaknesses of these algorithms enables informed choices and fosters advanced insights in sentiment analysis. Future developments are likely to focus on hybrid models and real-time sentiment analysis, improving e-wallet services and boosting customer satisfaction. These approaches are necessary for understanding the sentiment flow in users' testimonials thereby providing background for this study. Emoticon weighting significantly improves the performance of sentiment analysis algorithms, boosting the accuracy rates ranging 87% to 89% of Naïve Bayes and SVM algorithms . This implies that emoticons had a critical role in user reviews, showing that traditional processing methods and validating inclusion of non verbal cues in sentiment analysis. The findings highlight that emoticons extract techniques, suggesting future research with iterating various non-verbal cues to further use of sentiment analysis. Such advancements could benefit fields like social media analytics, customer service, and digital marketing by providing a more understanding of user recommendations. This study conveys detailed sentiment analysis models and provides more accuracy so as to better gauge user emotions and improve classification models. Linking machine learning with Structural Equation Modeling (SEM) helps to understand user intentions and preferences better with machine learning algorithms such as Naïve Bayes and Support Vector Machine (SVM), this algorithm provides an approach to sentiment analysis. By doing so, refining its approach enables it to do a better job at influencing user behavior while gaining deeper insights into user patterns. Integrating SEM with machine learning enables deeper understanding of user attitudes and preferences, offering different valuable insights for developing more user-friendly e-wallet apps. With that, researchers can assume that future research should consider incorporating different factors into optimizing sentiment classification that would have implications on business areas like digital marketing or customer satisfaction through enhancing how pleased users are with this service delivery process in digital wallet services.

References

Arora, A., Malhotra, R., Saxena, N., & Sharma, S. (2018). Wallet app credibility analysis based on app content and user reaction. *In 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)* (pp. 263–268). IEEE. https://doi.org/10.1109/ETTLIS.2018.8485252

Bau, Y. T. (2023). Sentiment analysis of e-wallet companies: Exploring customer ratings and perceptions. *Journal of Logistics, Informatics and Service Science, 10*(4).

Hakim, M. M., Afifah, A. N., & Aryotejo, G. (2023). The analysis of factors affecting behavioral intention and behavior usage of e-wallet using meta-UTAUT model. *International Journal on Advanced Science, Engineering and Information Technology, 13*(2).

Harnadi, B., & Widiantoro, A. D. (2023). Evaluating the performance and accuracy of supervised learning models on sentiment analysis of e-wallet. *In 2023 7th International Conference on Information Technology (InCIT)* (pp. 175–180). IEEE. https://doi.org/10.1109/InCIT56086.2023.10239068

Hermanto, T. A., Kuntoro, A. Y., Fahlapi, R., Effendi, L., & Syukmana, F. (2023). Sentiment analysis review Flip app users on Google Play using naïve Bayes algorithm and support vector machine with SMOTE technique. *In AIP Conference Proceedings* (Vol. 2714, 020036). AIP Publishing. https://doi.org/10.1063/5.0123456

Ilieva, G., Yankova, T., Dzhabarova, Y., Ruseva, M., Angelov, D., & Klisarova-Belcheva, S. (2023). Customer attitude toward digital wallet services. *Systems, 11*(4), 185. https://doi.org/10.3390/systems11040185

Kristiyanti, D. A., Putri, D. A., Indrayuni, E., Nurhadi, A., & Umam, A. H. (2020). E-wallet sentiment analysis using naïve Bayes and support vector machine algorithm. *Journal of Physics: Conference Series, 1641*(1), 012079. https://doi.org/10.1088/1742-6596/1641/1/012079

Kurniawan, Y., Hartanto, F. E., Dwiparna, I. M., Adam, R. E., Anwar, N., & Luhukay, D. (2022). The customer behavior data analysis towards use of digital wallet during COVID-19 pandemic. *In 2022 International Conference on Information Management and Technology (ICIMTech)* (pp. 565–570). IEEE. https://doi.org/10.1109/ICIMTech55957.2022.9915123

Masturoh, S., Pratiwi, R. L., Saelan, M. R. R., & Radiyah, U. (2023). Application of the K-nearest neighbor (KNN) algorithm in sentiment analysis of the OVO e-wallet application. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer), 8*(2), 78–83. https://doi.org/10.33480/jitk.v8i2.3997

Nayoga, S. A., Saputra, M., & Fa'rifah, R. Y. (2023). An analysis of fintech technostress and its impact on smart economy: A customer review-based sentiment analysis approach. *In 2023 10th International Conference on ICT for Smart Society (ICISS).* IEEE. https://doi.org/10.1109/ICISS59129.2023.10291862

Pradanita, W. R., & Rochimah, S. (2020). Quality in use of digital wallet based on ISO/IEC 25022. *In 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI)* (pp. 282–286). IEEE. https://doi.org/10.1109/EECSI50503.2020.9251896

Raghavendra, R., Niranjanamurthy, M., Nachappa, M. N., & Shalini, K. B. (2019). An emphasis of digital wallets for e-commerce transactions. *Journal of Computational and Theoretical Nanoscience, 16*(9), 3748–3753. https://doi.org/10.1166/jctn.2019.8244

Zahra, S., & Alamsyah, A. (2022). Digital wallet service quality analysis using multiclass classification and sentiment analysis. *In 2022 5th International Conference on Information and Communications Technology (ICOIACT).* IEEE. https://doi.org/10.1109/ICOIACT55506.2022.10074394

How to Cite this Chapter (APA 7):

Abraham, J. D., Halim, K. P., Hamto, R. I., Lamigo, J. J., Jr., & Dumpit, C. C. (2026). Sentiment analysis: Understanding user perception of digital wallets on app market – A literature review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 31–36). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 6: The Impact of Artificial Intelligence on Work Productivity and Quality: A Meta-Aggregative Review

Ana Maridel F. Cosino[1], Sophia Joy S. Edralin[1], Daizzyrie T. Galido[1], Neil Christian S. Pimentel[1], Belle Deandra A. Villegas[1] & Johneros P. Puyo[2]
Cavite State University – Silang Campus, Cavite, Philippines
[2]Philippine Christian University - Dasmariñas, Cavite, Philippines

Abstract. *Artificial Intelligence is bringing new ways to improve how much and how well people do their work. Still, how technology affects companies is affected by how well the technology is created, how adaptable people are, and how ready the organization is. This study aims to find how AI can be used in workplaces best, highlighting the importance of infrastructure, trust, working together, and what happens in specific sectors to improve productivity. We conducted a meta-aggregative review of the literature on a qualitative level, screening the literature of 2020-2025 and critically appraising the applicable studies and synthesizing the technological, organizational, human and ethical findings on AI. The research indicates that the productivity of AI mostly depends on how smoothly AI fits into a company's processes, the reliability of input data, how employees use it, and the sense of teamwork regarding AI. Although AI streamlines standard processes and improves decisions helped by data, its usefulness is affected by the difficulties of NLP, incorrect sense of meaning, and needing specialized knowledge. For the best outcome, people need to trust AI, decide things openly, and make efforts to work together with AI. It is worth noting that the impact of AI can change a lot depending on tasks and jobs since mid-level staff gain more from it than experienced experts. This study shows that making AI accessible leads to advancements in technology, workforce adjustment, rules set by governments, and ethical job standards. It recommends encouraging digital skills, introducing diversity into AI, and building strong regulations to ensure that AI does not replace humans in their work.*

Keywords: *Artificial Intelligence, Workplace Productivity, Human-AI Collaboration, Organizational Integration, Usability Challenges*

Introduction

The quick adoption of Artificial Intelligence (AI) in workplace settings created substantial improvements for productivity alongside quality standards but it still remains divided about its complete effects. Technology enhancements through AI enable higher efficiency rates, such as through automation, decision-making support, and data analysis, yet workers encounter unresolved issues when AI impacts work quality together with employee working methods. Studies show that automation ends up increasing productivity levels in regular operations (Braganza et al., 2020), while indicating potential disadvantages from skill deterioration, excessive dependence upon machine systems, and the possibilities of work intensification and job losses (Moore, 2019). According to Fitri et al., (2023), when AI integrates with employee engagement programs the result creates enhanced workplace productivity along with better communication streams and collective work interaction. However, there are also studies that determine how AI assistants might affect user's cognitive skills and might accelerate skill decay and hinder skill acquisition (Macnamara et al., 2024). AI technology has proven capable of reducing human errors according to existing evidence yet there remain doubts about its ability to handle artistic work and situations that necessitate emotional understanding along with a full understanding of situations. Previous studies have dedicated their research mostly to technological industries which has created gaps regarding AI usage in various business sectors. The research requirements to measure AI's simultaneous effects on workplace effectiveness and quality match this present study's main goals to discover efficiency parameters and ideal usage patterns and general application performance in real-world scenarios.

The rapid deployment of artificial intelligence (AI) technology within contemporary workplaces creates multiple unknown effects on employee performance quality along with productivity. The research done by Brynjolfsson et al., (2023) shows that support agents enhance their productivity by 15% when they use AI tools but other reports indicate that excessive use of such tools creates mental strain and deteriorates team cooperation (Moore, 2019). Research by Dell'Acqua et al., (2023) highlights essential usability issues because staff members encounter difficulties in their attempts to work with AI systems. Additionally, research based on Taiwanese electronics firms demonstrates that AI implementations lead to productivity improvements yet they relocate work opportunities from lower-skilled personnel towards more skilled workers thus creating difficulties in hiring unskilled labor (Yang, 2022). Further research of AI efficiency in work environments and optimal usability conditions alongside its work productivity and quality impact needs to be pursued because of these contradictory findings. Organizations must focus on resolving these issues because they determine the ability to extract AI potential while addressing its negative aspects. AI's ongoing development requires thorough investigation of its various workplace effects to develop successful implementation procedures.

The study objective is to clearly define what is the optimal amount of usage of AI by determining each factor that affects its quality. Thus our hypothesis is that the adoption of artificial intelligence (AI) in a workplace setting enhances things such as productivity and work efficiency; however , it introduces challenges such as work quality, employee skill stagnation, and sector-

specific impacts acknowledging that A I effects differ in each industry. Leading to different effects depending on organizational context and implementation of AI in a workplace setting. By knowing the optimal amount of AI usage by considering each factor we will be able to maximize AI full potential in sector-specific problems to increase work productivity/efficiency without sacrificing work output quality. Hence this paper aims to determine the optimal usage of considering AI in sector- specific applicability, as the basis for the proper adoption of AI in workplace settings in each industry.

Materials and Methods

The Design

This qualitative exploratory study examines the diverse impacts of artificial intelligence (AI) in different workplaces. It compares the AI's influence on productivity, work quality, and employee skills across sectors like technology, healthcare, manufacturing, and services by analyzing existing research and theories. Qualitative approach allows for a detailed understanding of the complex context dependent benefits and challenges of AI integration in these industries.

Research Objectives

This study aims to explore the research question: what is the optimal amount of AI usage in a work environment to achieve high productivity and quality? Its objectives are to: (1) to identify the factors influencing the efficiency of AI usage in work environments, (2) to identify the optimal usability of AI in work environments, and (3) to synthesize the effectiveness of AI usage in work productivity and quality.

Retrieving and Selecting Pertinent Literature

The material and computer science literature used in the review were obtained from a reliable and respectable internet resource that has published pertinent works. The keywords "AI Workplace Impact" are used in the search request for the two online databases. Nine thousand two hundred ninety (9,290) results were identified from ScienceDirect and 145 articles from IEEE. Additionally, a search for pertinent articles was conducted using Google Scholar, which produced about 19,100 results, including works from reliable sources like IEEE and ScienceDirect, etc. Following a review of the papers using predetermined inclusion and exclusion criteria, 44 articles were chosen. After a careful review and analysis of each of them, 36 articles ended up being selected for the study.

Synthesizing the Literature

The relevant works included for this review were released in the years 2020–2025. 36 publications in all were selected for analysis after meeting the inclusion requirements. The accompanying table summarises the key findings and data that were taken from each examination.

Table 1. To Identify the factors influencing the efficiency of AI usage in work environments.

Category	Authors	Factors Influencing AI Efficiency
Technological Factors	Alzeraif & Cheaitou (2023); Buschmeyer et al. (2024); Gao & Feng (2023); Odonkor et al. (2024); Spring et al. (2022)	Data quality, AI accuracy, automation, explainability, technological infrastructure, workflow automation
Human Factors	Clear et al. (2025); Fitri et al. (2023); Russell et al. (2023); Holmes & Littlejohn (2024); Nazaretsky et al. (2022); Liang et al. (2022)	Technical skills, training, AI literacy, motivation, adaptability, trust in AI
Organizational Factors	Araz Zirar et al. (2023); Gillan et al. (2021); Khanijahani et al. (2022); Patil (2025); Nurlia et al. (2023); Lane et al. (2023)	Organizational integration, leadership support, institutional readiness, policy frameworks, change management
Ethical and Legal Factors	Cox (2022); Langer & Landers (2021); Odonkor et al. (2024); Russell et al. (2023)	Ethics, transparency, regulatory compliance, data security, accountability
Productivity and Performance Factors	Damioli et al. (2021); Green et al. (2023); Tasheva & Karpovich (2024); George (2024); Czarnitzki et al. (2023)	Productivity improvement, automation, job performance, AI adoption level
Psychological and Behavioral Factors	Gamkrelidze et al. (2021); Hemmer et al. (2023); Sowa et al. (2021); Singh et al. (2025); Kong et al. (2021)	Trust in AI, role clarity, autonomy, emotional response, employee engagement
Work and Job Design Factors	Nurski & Hoffmann (2022); Wilkens (2020); Babashahi et al. (2024); Ravirajan & Sundarajan (2025); Nasrullah (2023)	Job redesign, human-AI collaboration, workflow integration, job transformation
Education and Training Factors	Grassini (2023); Holmes & Littlejohn (2024); Clear et al. (2025); Nazaretsky et al. (2022)	Training programs, curriculum adaptation, learning culture, AI readiness

Table 2. To identify the optimal usability of AI in work environments.

Category	Authors	Factors Influencing AI Efficiency
Technological & System Design	Alzeraif & Cheaitou (2023); Buschmeyer et al. (2024); Odonkor et al. (2024); Spring et al. (2022); Ravirajan & Sundarajan (2025)	Automation, accuracy, user-friendly design, system transparency, explainable AI, interface design, smooth integration
Human–AI Collaboration	Araz Zirar et al. (2023); Gamkrelidze et al. (2021); Langer & Landers (2021); Babashahi et al. (2024); Möllers et al. (2024); Sowa et al. (2021); Tasheva & Karpovich (2024)	Human-AI collaboration, augmentation, collaboration over competition, human-centered design, flexibility
Training & Skills Development	Clear et al. (2025); Fitri et al. (2023); Nazaretsky et al. (2022); Russell et al. (2023); Holmes & Littlejohn (2024); Nurlia et al. (2023)	Training, AI familiarity, professional development, skills development, learning systems
Organizational Integration	Czarnitzki et al. (2023); Gillan et al. (2021); Khanijahani et al. (2022); Patil (2025); Lane et al. (2023); Nasrullah (2023)	Organizational integration, guidelines, leadership support, workflow alignment, economic and policy support
Job Design & Work Structure	Nurski & Hoffmann (2022); Green et al. (2023); Wilkens (2020); Liang et al. (2022); Kong et al. (2021)	Job autonomy, workload balance, human-in-the-loop systems, motivation, stress management
Industry & Context Factors	Gao & Feng (2023); George (2024); Grassini (2023)	Industry type, sector readiness, educational support, workforce adaptability
Ethics, Transparency & Trust	Cox (2022); Hemmer et al. (2023); Singh et al. (2025); Tasheva & Karpovich (2024)	Transparency, ethical design, emotional impact management, user control, trust
Productivity & Decision Support	Damioli et al. (2021); Nasrullah (2023); Spring et al. (2022)	Task automation, decision support systems, routine task support

Table 3. To synthesize the effectiveness of AI usage in work productivity and quality.

Category	Authors	Factors Influencing AI Efficiency
Productivity & Performance Improvement	Alzeraif & Cheaitou (2023); Czarnitzki et al. (2023); Damioli et al. (2021); Gao & Feng (2023); Odonkor et al. (2024); Spring et al. (2022); Nasrullah (2023)	Automation, error reduction, faster processing, improved productivity, improved work quality
Decision-Making & Task Performance	Buschmeyer et al. (2024); Langer & Landers (2021); Russell et al. (2023); Sowa et al. (2021)	Improved decision-making, expert performance, knowledge management, reduced workload
Human–AI Collaboration	Araz Zirar et al. (2023); Fitri et al. (2023); Gamkrelidze et al. (2021); Hemmer et al. (2023); Patil (2025); Babashahi et al. (2024)	Productivity improves when AI supports and collaborates with humans
Training, Learning & Skill Development	Clear et al. (2025); Holmes & Littlejohn (2024); Nazaretsky et al. (2022); Grassini (2023); Wilkens (2020)	Continuous learning, skill development, improved engagement, better work quality
Organizational & Policy Factors	Green et al. (2023); Gillan et al. (2021); Lane et al. (2023); Tasheva & Karpovich (2024); Kong et al. (2021)	Organizational support, policy frameworks, performance management, inclusive implementation
Job Design, Well-being & Work Environment	Nurski & Hoffmann (2022); George (2024); Ravirajan & Sundarajan (2025); Liang et al. (2022); Singh et al. (2025)	Work-life balance, job satisfaction, emotional impact, well-being, innovation
Safety, Quality & Service Improvement	Khanijahani et al. (2022); Möllers et al. (2024)	Improved safety, service quality, healthcare outcomes

Discussion

3.1 The factors influencing the efficiency of AI usage in work environments.

Artificial intelligence (AI) productivity at work is a complex phenomenon influenced by a nuanced interdependence among technical design, human interaction, and organizational preparedness. Throughout the literature, there is a convergence that although AI has enormous potential to streamline work processes, its productivity benefits and successful implementation rely on more significant factors than technological sophistication. A close examination of current research indicates that efficiency gains

heavily depend upon the quality of input data, task clarity of design, skill levels of employees, trust in AI technology, and the general organizational environment into which the AI is integrated.

Technical Infrastructure and Data Integrity

Other research shows the invaluable role that technological infrastructure and the quality of data would play in enhancing the successful utilisation of AI. As an illustration, Alzeraif and Cheaitou (2023) state that input data accuracy and task complexity are direct factors affecting the availability of resources and predictability of AI performance. Well-constructed AI models regarding well-defined and suitable data help to implement better decision-making processes and scalable productivity support. Similarly, Buschmeyer, Zenner, and Hatfield (2024) also came to the conclusion that the quality of AI-driven decision support systems as accuracy and explainability of an AI system were central in determining its effectiveness in real work settings. Without these characteristic features of technology, even highly developed AI tools will not produce the necessary results in case workers are unable to understand or believe in the rationale of AI conclusions.

Organizational and Structural Facts

Other than technological issues, organizational and structural features contribute significantly to defining the success of AI implementation. Araz Zirar et al. (2023) emphasize that organizational integration practices, employee training, and data management practices are essential to the thriving AI ecosystem. This is supported by the study by Fitri et al. (2023), which found how well AI can increase worker productivity contingent on how prepared the human resources are and how successfully AI technologies could be integrated into the HR procedure. Companies investing in transparent change management, quality training programs, and digital transformations will find turning AM into tangible productivity easier. Along with the conclusions above, Cox (2022) and Gillan et al. (2021) concur that the triple pillar of ethical obligation, organizational culture, and leadership commitment is the basis on which environments where AI can operate in grand style without offering or sacrificing professional standards and the faith of the staff are built.

Importance of Human Skills, Adaptability, and Collaboration with AI Systems

Yet another prevalent theme in the literature is the significance of human competence, flexibility, and working with AI systems. Clear et al. (2025) emphasize that technical literacy, flexibility, and lifelong learning are essential professional competencies for professionals working within AI-enabled workplaces. In their absence, the potential of AI to enhance efficiency can be negated by human capabilities or incapacities to comprehend or use the technology. Furthermore, Möllers, Berger, and Klein (2024) claim that the nature of human-AI relationships, whether supportive or adversarial, can affect overall productivity outcomes. Supportive relationships where AI supplements, not replace, human intelligence generally produce better outcomes and greater job satisfaction. Likewise, the research conducted by Ravirajan and Sundarajan (2025) and Langer and Landers (2021) indicates that user perceptions, autonomy, and controlling decision-making procedures are dependent on AI's role in complementing human work.

Trust in AI Systems

Trust in AI systems proves to be another critical factor influencing the efficiency of AI. Gamkrelidze, Zouinar, and Barcellini (2021) and Nazaretsky et al. (2022) show that when workers believe in the judgments of AI systems and understand them to be transparent and committed to actual-world activities, adoption, and utilization grow much higher. Trust is not an automatic side effect of technical capacity; it has to be developed by training, transparency, and evidence of dependability. Hemmer et al. (2023) further note that transparency of task allocation, human supervision, and user control are essential in developing trust and satisfaction in accomplishing tasks through human-AI collaboration. However, a lack of transparency or interpretability may result in resistance, suspicion, and decreased utilization, undermining AI productivity gains.

AI Efficiency is Not Uniformly Experienced Across All Contexts or Job Roles

However, the research also recognizes that the efficiency of AI is not equally felt in all settings or occupations. Green et al. (2023) and Nurski & Hoffmann (2022) posit that while AI will optimize operations and automate mundane tasks, it can also enhance work demands and stress levels, mainly when applied to supplement managerial roles or track employee performance. Such unanticipated effects show the double-edged sword of AI adoption, where increases in production might be linked with reductions in worker well-being or control. Singh et al. (2025) even propose that AI interactions can influence emotional productivity by evoking stress reactions that, under some conditions, enhance performance. However, these influences introduce ethical and pragmatic issues regarding long-term durability and the mental well-being of employees.

Industry-Specific and Demographic Factors

In addition, industry and demographic factors drive heterogeneity in AI efficiency. Studies like those by Czarnitzki et al. (2023) and Damioli et al. (2021) reveal that the intensity of AI adoption, the breadth of its applications, and the availability of skilled personnel vary widely across sectors, affecting the degree to which productivity gains are realized. For instance, industries with well-established digital infrastructures and high levels of AI literacy—i.e., finance, healthcare, and information technology—

are likely to reap the benefits sooner and more intensely from AI integration than industries that are not so endowed. Further, Khanijahani et al.'s (2022) review concluded that demographic issues like staff training, levels of institutional preparedness, and even patient or client factors could influence the success of AI implementation in healthcare settings.

Ethical Considerations

It is also ethical issues that dominate the process of defining the impact of AI on the productivity and quality of work of the professionals. Cox (2022) and Grassini (2023) note that care should be exercised by making sure that the aspects of privacy, fairness, and responsibility have been addressed in such a manner that AI should not diminish professional accountability or the confidence of the population. More than that, unethical violations or unregulated AI systems may also result in biased decision-making or data abuse or other outcomes that undermine individual performance and corporate ethics. The results indicate the importance of combining ethical principles and governance frameworks in the implementation of AI technologies in the work areas.

Policy and Regulatory Frameworks

Second, effective policy and regulatory frameworks influence the outcome and direction of AI takeup. George (2024), and Lane et al. (2023) argue that we need protective policies, clear legal classifications of the functions of AI and labor protections in order to maximize the benefits of AI and minimize disruption of the labor market. With AI continuing to perform hiring tasks, performance assessment tasks, and the distribution of tasks, regulation is necessary to give ethical and fair usage. More effective and efficient adaptation to the AI transition is an instrument that might be adopted by institutions that are actively involved in the process of policy formulation and predict regulatory changes.

Skill Transformation and Workforce Adaptation

Lastly, the frequent trend of the literature is skill change and workforce adaptation. Leili Babashahi et al. (2024) acknowledge that working with AI means having new skills and redesigning existing employment with no time to wait, reskill, and upskill. Holmes & Littlejohn (2024) agree with such a view and highlight that in such a way as to train employees to use AI-related tools, it is vital to understand the cultures of companies and their motivational tools. As AI advances, lifelong learning will be necessitated to ensure an efficient workforce and generate innovation and resilience among workers.

3.2 The optimal usability of AI in work environments.

Enhancing AI Usability in the Workplace: A Multi-Faceted Viewpoint

Historical and contemporary benefits of using AI at the workplace: A complex combination of technology. The reality of the situation and even psychological factors determines the efficient applicability of AI to the workplace. The architecture, human dynamics, job-specifications and organizational orientation. The body of literature reviewed shows a common agreement that AI usability does not only rely on technical advancement. The percentage of AI-based tools constructed around the requirements of their users proves to be a key determinant of long-term usability. Their ease of integration with daily activities, the degree of training and independence of users. To what extent the system is ethically and emotionally aligned to the workplace. Research findings have always indicated that to truly be able to use AI, the latter has to be intuitive, flexible, and helpful, but intelligent technology should not be controlling nor rigid.

User-Focused Design, Clarity, and AI-Human Synergy

The message that is repeated time and again is that AI systems should be designed keeping the end-user in mind. Research According to Gamkrelidze, Zouinar, and Barcellini (2021), and Tasheva and Karpovich (2024), it is worth pointing out the fact that AI the tools that are developed in a transparent and flexible manner are usually more useful across the different professions settings. Ravirajan and Sundarajan (2025) emphasize the interfaces that are simple to use along with the appropriate training and better usability is obtained through programs. Similarly, Buschmeyer et al. (2024) emphasize the need of AI explaining things in a user-friendly manner, particularly to newcomers, to develop knowledge and trust. Systems which will have more effective user experiences and overall design to those that work with the users instead of being dominant performance.

Employee Readiness and Seamless Organizational Integration

The other important lesson is that AI usability can greatly rely on the level of preparedness of the employees and the extent of AI tools integration in regular organizational operations. Fitri et al. (2023) insist that the use of AI would be useful when good human resources and HR models accompany its implementation. When the staff receives sufficient training, engagement, and preparation, they are more likely to make appropriate use of AI tools. Gao and Feng (2023) also indicate that unequal preparedness in industries and flexibility in workforce as the reason will have a large impact on usability. In support of that, Nurlia et al. (2023) and Nazaretsky et al. (2022) reveal that Continuous skills training and acquaintance in AI systems are important in promoting the acceptance of users and curtailing shrinking resistance.

Variations in Usability Based on Role, Skills, and Context

All workers do not equally use AI. Nurski and Hoffmann (2022) and Green et al. (2023) it is important to note that AI monitoring or scheduling systems can decrease control and autonomy especially at lower-level positions. On the contrary, in the case of management or technical disciplines, AI is perceived as a useful aid over decision-making. George (2024) goes further and notes that digital inequities exist between different people. Educational possibilities as well as infrastructure in different directions and sectors may have a considerable impact on the level of usability of AI systems, by emphasizing the need to differentiate between approaches by location and situation.

Trust, Emotions, and Ethical Design

Trust is an important part of AI system reception and adoption. According to Hemmer et al. (2023) and Singh et al. (2025), emotional responses questioned in the survey are between confidence and confidence to emphasize is able to directly influence the appearance of AI tools in terms of utility. Ethical, open, responsive to asking systems are usually more acceptable. By contrast those, which make obscure decisions or defraud user manifested behavior usually brings in some reluctance or withdrawal. It is claimed that flexible, cooperative AI (Sowa et al., 2021) because everything should be done in a cooperative, pro-social way the so-called cobots engage more positively than either rigid or autonomous and are more readily adopted ones. Lack of concern towards emotional strain and fatigue because of ill-informed AI due to the failure to design them sensitively is also warned by Singh et al., strengthening the necessity to be emotionally attentive in deployment.

Influence of Industry and Policy on AI Usability

The study also reveals that the dynamics of the industry and the framework of public policy also meet the industry-specific needs that have a major role in AI usability. According to Damioli et al. (2021) and Liang et al. (2022), such industries as healthcare, education, and finance are more usable with the AI solutions making them efficient and consistent and well-organized work processes. On the other hand, a lower usability is more frequent in less digital mature sectors. According to Lane et al. (2023), fairness, inclusive, and human-centered legal and ethical regulations can be used. Regulation increases convenience. Similarly, Odonkor et al., (2024) underscore one of the strategies is, good compliance protocols. In particular, fields with control, such as finance support usability by being accountable and clear.

3.3 The effectiveness of AI usage in work productivity and quality.

The influence of Artificial Intelligence (AI) on work productivity and quality depends on the reliability of technology, how people adapt, organizational structures, and ethical considerations. This discussion synthesizes existing literature on AI's effectiveness in workplace productivity and quality. Despite AI's great potential to enhance work and business, its success depends on proper usage, training, and planning. The existing literature shows that AI's advantages are tied to integration, collaboration with people, and fair distribution, and are tailored to different sectors.

Technical Foundations and Operational Efficiency

According to the reviewed literature, many researchers highlighted that using AI simplifies tasks, lowers the risk of mistakes, and improves accuracy through proper data analysis. According to Alzeraif and Cheaitou (2023), AI increases productivity and quality by efficiently using resources, reduces errors through optimization and predictive analytics, and leads to faster and more accurate results. Similarly, Czarnitzki et al. (2023) and Damioli et al. (2021) conclude that AI is especially valuable for firms raising their productivity, mainly in finance and manufacturing. This means that technical ability is not the only factor that guarantees a successful outcome. Buschmeyer et al. (2024) stated that efficiency and accuracy depend on how knowledgeable the user is about AI and their ability to deal with any errors it produces. Using AI to support decision-making is mainly beneficial for mid-level employees, as experts in the organization experience very few benefits. This means that to be more productive, businesses must match the needs of each job with what AI supplies.

Ethical and Equity Considerations

Improvement in productivity through AI is closely tied to ethics and social matters. Nurski and Hoffmann (2022) and Green et al. (2023) reveal that introducing AI can increase the danger and prevalence of inequality among vulnerable groups due to unfair algorithms or the risk of job losses. As an illustration, if the datasets used to develop automated hiring tools are biased, they may negatively affect those facing employment barriers (Singh et al., 2025). Ensuring workers are treated with dignity is also an ethical issue. While AI might help organizations complete routine work, it can also lead to more pressure and more control over employees, mainly if their job performance is evaluated by an AI system (Kong et al., 2021). Cox (2022) and Grassini (2023) argue that we must ensure that AI abides by professional codes and essential social values.

Psychological and Behavioral Impacts of AI Adoption

Integrating AI into workplaces dramatically affects employees' minds and actions, shaping their work and production. Still, correctly adopting AI leads to less fatigue and greater employee creativity, whereas misusing it might increase burnout (Liang et al., 2022; Kong et al., 2021). Making use of AI helps enhance employee satisfaction, and being trusted by workers is key to adopting AI (Langer & Landers, 2021). It also changes how managers and employees work together and makes constant upskilling

necessary (Green et al., 2023; Holmes & Littlejohn, 2024; Clear et al., 2025). The results of AI in organizations depend on how human workers adapt their actions to it (Gamkrelidze et al., 2021). The research shows that having the right culture and considering people's needs is key to AI benefiting workers and enhancing workplace productivity.

Human AI Collaboration

Table 3 literature emphasizes that human-AI collaboration is significantly better when AI is used as a collaborative tool rather than a total replacement. Studies consistently show that human and AI collaboration results in more sustained productivity by utilizing both parties' strengths (Hemmer et al., 2023; Langer & Landers, 2021). In particular, human involvement and decision-making allow and guide AI outputs to stay on track and within organizational goals, while employee AI literacy training allows for mitigating risks and allows for optimized integration (Nurski & Hoffmann, 2022; Clear et al., 2025). Moreover, according to (Russell et al., 2023), the success of AI is directly connected to trust and user perception, noting that higher trust in AI correlates to higher usage and adoption, highlighting the importance of transparency and design explainability. The more they understand how things work, the more likely they are to trust and use them. However, AI contributions diminish in fields that require high expertise, meaning that AI contributions are only slightly significant when it comes to an environment where humans operate with near-optimal efficiency (Buschmeyer et al., 2024). These insights support the idea that AI is most effective as a tool rather than a replacement.

Building on these insights, it is clear that AI is most effective when integrated into the workplace as a tool and not as a replacement for human labor. While outcomes may differ across industries, existing literature consistently demonstrates that even modest contributions can enhance workflow, productivity, or work quality. Importantly, AI value does not replace human intelligence but augments human thinking capabilities and strengths. This reinforces that AI should support human strength instead of undermining critical thinking and expertise.

Sector Specific Implementation

Variability of AI effectiveness across different sectors, with outcomes influenced by complexity, environmental regulations, and workforce characteristics. For example, the studies of (Khanijahani et al., 2022; Grassini, 2023) state that AI enhances performance through personalization and its predictive capabilities, which helps in more accurate diagnosis in healthcare and tailored learning experiences in education. Regarding manufacturing and finance, results have pronounced efficiency gains with the help of AI automation and process optimization (Odonkor et al., 2024). Meanwhile, AI contributes to workflow in the creative and service sectors but struggles with tasks that require emotional intelligence, context sensitivity, and innovation (Spring et al., 2022). However, reliance could result in oversight, which raises concerns about human judgment and critical thinking (Nazaretsky et al., 2022), which can be addressed through the proper implementation of AI by using regulations, policies, and protocols on properly using AI in the workplace.

Seeing all this variability in AI implementation in the workplace environment highlights the need for a sector-specific approach to AI deployment. One that considers factors such as industry norms, the nature of work, and the amount of human intervention required. This points out the limitations and lapses of a generalized adoption strategy of AI implementation, reinforcing that each sector should have a tailored implementation framework to achieve optimal results in integrating AI in the workplace.

Conclusion

This research indicates that integration of AI into workplaces should consider aspects related to technology, workforce, and the organization. This study finds that AI works well as much due to employee ability and adaptability as it does to technological progress. The research reveals aspects that NLP systems are missing and urges the use of human-AI teams, strong ethical rules, and customized solutions for different sectors. The key lessons impact many different parts of society: in technology, they underline the importance of trustworthy and more transparent AI; in economics, they mention the need to compare boosts in productivity to extra implementation expenses; in society, they remind us that we must help people learn new skills and consider psychological effects; and in organizations, they suggest that we should be more prepared for change and promote cultural adaptation.

Moving ahead, organizations should develop a strategy that places more emphasis on human and AI working together, rather than having everything automated. Here are three main steps: The first is to thoroughly check both the industry's infrastructure and its employees' expertise to address AI, the other is to plan implementation strategies suitable for the industry, and the third is to monitor data on job performance and its impact. Great organizations will see AI as something that helps workers be better at what they do, not as a threat to replace people. Researchers should create standard ways to evaluate AI and conduct studies over time across different industries to more accurately follow its development. Consequently, viewing AI holistically helps businesses use AI benefits and support a more productive and human workforce.

References

Alzeraif, M., & Cheaitou, A. (2023). Artificial Intelligence-based Optimization Models for the Technical Workforce Allocation and Routing Problem Considering Productivity. *International Journal of Advanced Computer Science and Applications*, *14*(12). https://doi.org/10.14569/ijacsa.2023.0141233

Araz Zirar, Ali, S. I., & Islam, N. (2023). Worker and workplace Artificial Intelligence (AI) coexistence: Emerging themes and research agenda. *Technovation*, *124*, 102747–102747. https://doi.org/10.1016/j.technovation.2023.102747

Babashahi, L., Barbosa, C. E., Lima, Y., Lyra, A., Salazar, H., Argôlo, M., Almeida, M. A. de, & Souza, J. M. de. (2024). AI in the Workplace: A Systematic Review of Skill Transformation in the Industry. *Administrative Sciences*, *14*(6), 127. https://doi.org/10.3390/admsci14060127

Buschmeyer, K., Zenner, J., & Hatfield, S. (2024). Effectiveness of AI-based decision support systems in work environment: a systematic literature review. *International Journal of Human Factors and Ergonomics*, *11*(5), 1–54. https://doi.org/10.1504/ijhfe.2024.142761

Clear, T., Cajander, Å., Clear, A., McDermott, R., Daniels, M., Divitini, M., Forshaw, M., Humble, N., Kasinidou, M., Kleanthous, S., ... (2025). *AI integration in the IT professional workplace: A scoping review and interview study with implications for education and professional competencies*. In ITiCSE 2024 Working Group Reports on Innovation and Technology in Computer Science Education (pp. 34–67). Association for Computing Machinery. https://doi.org/10.1145/3689187.3709607

Cox, A. (2022). The ethics of AI for information professionals: Eight scenarios. *Journal of the Australian Library and Information Association*, *71*(3), 201-214. https://doi.org/10.1080/24750158.2022.2084885

Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, *211*, 188-205. https://doi.org/10.1016/j.jebo.2023.05.008

Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Business Review*, *11*(1), 1–25. https://doi.org/10.1007/s40821-020-00172-8

Fitri, D., Sri Langgeng Ratnasari, None Suyanto, & Sultan, Z. (2023). Enhancing Employee Productivity Through Technology System AI-Based Approaches. *Proceeding of the International Seminar on Business Economics Social Science and Technology (ISBEST)*, *3*(1). https://doi.org/10.33830/isbest.v3i1.1236

Gamkrelidze, T., Zouinar, M., & Barcellini, F. (2021). Artificial Intelligence (AI) in the Workplace: A Study of Stakeholders' Views on Benefits, Issues and Challenges of AI Systems. *Lecture Notes in Networks and Systems*, 628–635. https://doi.org/10.1007/978-3-030-74614-8_78

Gao, X., & Feng, H. (2023). AI-Driven Productivity Gains: Artificial Intelligence and Firm Productivity. *Sustainability*, *15*(11), 8934–8934. https://doi.org/10.3390/su15118934

George. (2024). Automated Futures: Examining the Promise and Peril of AI on Jobs, Productivity, and Work-Life Balance. *Partners Universal Innovative Research Publication*, *2*(6), 1–17. https://doi.org/10.5281/zenodo.14544519

Gillan, C., Hodges, B., Wiljer, D., & Dobrow, M. (2021). Health care professional association agency in preparing for artificial intelligence: Protocol for a multi-case study. *JMIR Research Protocols*, *10*(5), e27340. https://preprints.jmir.org/preprint/27340

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education sciences*, *13*(7), 692. https://doi.org/10.3390/educsci13070692

Green, A., Del Pero, A.S., Verhagen, A. (2023). Artificial Intelligence, Job Quality And Inclusiveness. *In OECD Employment Outlook Artificial Intelligence And The Labour Market(Pp.128-153)* https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/07/oecd-employment-outlook-2023_904bcef3/08785bba-en.pdf#page=130

Heller, C. H. (2019). Near-term applications of artificial intelligence: Implementation opportunities from modern business practices. Naval War College Review, 72(4), 73–100. JSTOR. https://www.jstor.org/stable/26775520

Hemmer, P., Westphal, M., Schemmer, M., Vetter, S., Vössing, M., & Satzger, G. (2023, March). Human-AI collaboration: the effect of AI delegation on human task performance and task satisfaction. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 453-463). https://dl.acm.org/doi/abs/10.1145/3581641.3584052

Holmes, W., & Littlejohn, A. (2024). Artificial intelligence for professional learning. In *Handbook of artificial intelligence at work* (pp. 191-211). Edward Elgar Publishing. https://doi.org/10.4337/9781800889972.00018

Khanijahani, A., Iezadi, S., Dudley, S., Goettler, M., Kroetsch, P., & Wise, J. (2022). Organizational, professional, and patient characteristics associated with artificial intelligence adoption in healthcare: A systematic review. *Health Policy and Technology*, *11*(1), 100602. https://doi.org/10.1016/j.hlpt.2022.100602

Kong, H., Yuan, Y., Baruch, Y., Bu, N., Jiang, X., & Wang, K. (2021). Influences of artificial intelligence (AI) awareness on career competency and job burnout. *International Journal of Contemporary Hospitality Management*, *33*(2), 717–734. https://doi.org/10.1108/ijchm-07-2020-0789

Lane, M., Williams, M., Organisation for Economic Co-operation and Development, & Scarpetta, S. (2023). Defining and classifying AI in the workplace. In *OECD Social, Employment and Migration Working Papers* (Report No. 290). Organisation for Economic Co-operation and Development. https://dx.doi.org/10.1787/59e89d7f-en

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, *123*, 106878–106878. https://doi.org/10.1016/j.chb.2021.106878

Liang, X., Guo, G., Shu, L., Gong, Q., & Luo, P. (2022). Investigating the double-edged sword effect of AI awareness on employee's service innovative behavior. *Tourism Management*, *92*, 104564. https://doi.org/10.1016/j.tourman.2022.104564

Möllers, M., Berger, B., & Klein, S. (2024). Contrasting human-AI workplace relationship configurations. In *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (pp. 282-303). Edward Elgar Publishing. https://doi.org/10.4337/9781803926216.00025

Nasrullah, M. (2023). Utilization of Artificial Intelligence (AI) in Information Systems to Improve Business Efficiency. *ProBisnis : Jurnal Manajemen*, *14*(5), 154–164. https://doi.org/10.62398/probis.v14i5.310

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British journal of educational technology*, *53*(4), 914-931. https://doi.org/10.1111/bjet.13232

Nurlia, N., Daud, I., & Rosadi, M. E. (2023). AI Implementation Impact on Workforce Productivity : The Role of AI Training and Organizational Adaptation. *Escalate : Economics and Business Journal*, *1*(01), 01-13. https://doi.org/10.61536/escalate.v1i01.6

Nurski, L., & Hoffmann, M. (2022). *Bruegel THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE NATURE AND QUALITY OF JOBS*. https://www.jstor.org/stable/pdf/resrep50976.pdf?refreqid=fastly-default%3A75e6947c51ad5ab3433de5b16fb1d8d5&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&origin=&initiator=search-results&acceptTC=1

Odonkor, B., Kaggwa, S., Uwaoma, P. U., Hassan, A. O., & Farayola, O. A. (2024). The impact of AI on accounting practices: A review: Exploring how artificial intelligence is transforming traditional accounting methods and financial reporting. *World Journal of Advanced Research and Reviews*, *21*(1), 172-188. https://doi.org/10.30574/wjarr.2024.21.1.2721

Patil, D. (2025). *Human-Artificial Intelligence Collaboration In The Modern Workplace: Maximizing Productivity And Transforming Job Roles*. https://doi.org/10.2139/ssrn.5057414

Russell, R. G., Lovett Novak, L., Patel, M., Garvey, K. V., Craig, K. J. T., Jackson, G. P., Moore, D., & Miller, B. M. (2022). Competencies for the Use of Artificial Intelligence–Based Tools by Health Care Professionals. *Academic Medicine*, *98*(3), 348–356. https://doi.org/10.1097/acm.0000000000004963

Singh, S., Siemon, D., & Vodolazskii, D. (2025). Emotional Productivity: Exploring the Impact of AI Interactions on Employee Well-Being and Workplace Efficiency. *Lutpub.lut.fi*. https://lutpub.lut.fi/handle/10024/168906

Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work: Human–AI collaboration in managerial professions. *Journal of Business Research*, *125*, 135-142. https://doi.org/10.1016/j.jbusres.2020.11.038

Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from Artificial-Intelligence-based systems in professional service operations. *Journal of Operations Management*, *68*(6–7), 592–618. https://doi.org/10.1002/joom.1215

Sundarajan, A. (2025). *Enhancing Workplace Productivity and Well-being Using AI Agent. arXiv preprint arXiv:2501.02368*.

Tasheva, Z., & Karpovich, V. (2024). SUPERCHARGE HUMAN POTENTIAL THROUGH AI TO INCREASE PRODUCTIVITY THE WORKFORCE IN THE COMPANIES. *American Journal of Applied Science and Technology*, *4*(2), 24–29. https://doi.org/10.37547/ajast/volume04issue02-05

Wilkens, U. (2020), "Artificial intelligence in the workplace – A double-edged sword", *International Journal of Information and Learning Technology*, Vol. 37 No. 5, pp. 253-265. https://doi.org/10.1108/IJILT-02-2020-0022

How to Cite this Chapter (APA 7):

Cosino, A. M., Edralin, S. J., Galido, D., Pimentel, N. C., Villegas, B. D., & Puyo, J. P. (2026). The impact of artificial intelligence on work productivity and quality: A meta-aggregative review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 37–45). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 7: AI-Based Cybersecurity Systems: A Review of Intelligent Threat Detection and Prevention Approaches

Neil Ryan R. Bayhon[1], John Ryan Gio G. Bayhon[1], Katlyn Joy B. Lopez[1], Jaynor L. Mendoza[1], Aaron C. Romano[1]& Ingersoll Dandan[2]
[1]Cavite State University – Silang Campus, Cavite, Philippines
[2]Philippine Christian University - Dasmariñas, Cavite, Philippines

Abstract. *Artificial Intelligence (AI) in cybersecurity is changing the direction of protecting against the growing threats posed by the digital world and making defense quicker, smarter, and more adaptive. This paper discusses the transformations of Artificial Intelligence into cybersecurity reinforcement, by detecting and impending threats, user reaction, and real-time preventive measures. It is aimed at recognizing some of the key factors that are either disruptive or help in the strengthening of AI and an evaluation of developing smarter and dynamic security systems. This paper utilized a systematic review of 20 peer-reviewed articles published between 2020 and 2025, sourced from Google Scholar, ScienceDirect, and IEEE Xplore. Articles were selected based on specific inclusion criteria and relevance to keywords such as artificial intelligence, cybersecurity, threat detection, machine learning, and AI-driven prevention systems. AI has demonstrated a high ability to enhance cybersecurity through the provision of sophisticated techniques of detecting and responding to threats. It improves threat detection, applying machine and deep learning to identify complex and unprescribed patterns of attack, in addition to minimizing possible threats. Furthermore, AI renders its users with automatic alerts and learning courses that enhance the response to the threat overall. Continuous development and alignment with the current technological solutions, including such innovative examples as blockchain, may help to maintain protection at the high level because cyber risks are continuously evolving. The future of digital security is ethical responsibility to proactive users that will be able to understand the threat earlier due to Artificial Intelligence security measures. The dependable consciousness to the shared environment will guarantee that the sensitivity to openness, access and preventive participation will instill the reliability needed to utilize it adequately.*

Keywords: Cybersecurity, Threat Detection, Artificial Intelligence, Machine Learning, Proactive Defense

Introduction

With the fast rise of cyber threats, the cybersecurity sector is resorting to Artificial Intelligence (AI) as a new effective tool. AI can assist in the finding and stopping of modern cyber-attacks, as it is able to study much data, notice patterns rather quickly, and take actions in real-time. According to Salem et al. (2024). present the finding that AI reduces false positives, and detection systems are not only improved but also more trustworthy. As Ahsan et al. (2022) noted, identifying unusual behavior and automatic responses is important in AI-based cybersecurity. Moreover, Sarker et al. (2020) stated that to better learn and find threats, AI requires good data quality and model design. Improving the ability of the systems to be used easily and allowing the users to respond quickly to a threat are also the functions of AI. According to Ansari et al. (2022) and Abdullahi et al. (2022), AI increases the understanding of users by providing automatic notifications, learning the behavior of users, and monitoring in real-time, particularly in Internet of Things (IoT) models. The role of AI in cybersecurity can be even better understood in the light of real attack strategies provided in the recent studies. As an example, botnet attacks described by Salem et al. (2024) make use of networks of infected computers to launch harmful activities in the form of DDoS attacks, data theft, and spam campaigns. Such botnets may comprise millions of hacked gadgets and hence are hard to find and block. The other real-life problem, which is known as an insider threat, is the problem that involves the authorized access of the data by the user (usually the employee) who decides to use his/her privileges to steal or harm the data. Such threats wherever they come about, whether by frustrated employees or sneaky software, pose great dangers to firms and organizations. That is why this situation proves that not only external threats may create significant security risks but even internal sources can become a serious security threat, this is why smart AI-based defenses, which can respond to a threat in real-time, are necessary. Nevertheless, Ullah et al. (2024) were concerned with the fact that main systems of AI may be under cyber-attacks. Their article indicated the combination of AI and blockchain technologies in data security improvement. On the same note, Kavitha and Thejas (2024) stated that despite the ability to predict and adjust to threats, AI has some issues such as its hard-to-understand nature and weakness to tricky or unclear information.

Although many studies have covered AI-related approaches for threat detection and prevention, the stream of literature in this regard is not well organized with differences between findings, theoretical models, and practical applications. For instance, Salem et al. (2024) AI-based detection methods make up a wide field, and a systematic review presents a clear overview of such methods, highlighting the possibility of using AI approaches such as CNNs in this context (Haseeb et al. 2024) however, it also points out some of the challenges linked to their ability to handle growth and adjust to changes. Similarly, Ahsan et al. The usefulness of machine learning models for reducing cyber threats is recognized (2022b), but the results show problems such as low-quality data, model clarity, and how well it works for different types of attacks. Other studies, like Sarker et al. (2020) and Abdullahi et al. (2022b), say that uneven data and lack of quick response remain as big challenges to effective AI use.

Moreover, literature includes differences concerning the use of explainability in AI security models. Zhang et al. (2022) suggest that explainable AI builds trust and helps in decision-making. Ansari et al. (2022), however, are doubtful about its use in serious cyber settings. New technologies, including blockchain integration (Ullah et al., 2024) and generative AI (Ankalaki et al., 2025), present interesting opportunities that also bring new uncertainties and problems in combining them. In spite of these developments, Al Siam et al. (2025) say that the methods and strategies used in cyber-attacks keep changing, and this happens often. Zero-day exploits prove this point well since, in most cases, they break through the current AI defense systems and show the problems with gaps in real-time protection. Such differences and missing parts show that the issue should be studied further in terms of the main reasons why AI works in threat detection, how easy AI tools are to use for raising user awareness and quick response, and the overall result of using AI in preventing and handling cyber-attacks. Within such a changing and complex cybersecurity environment, a complete, flexible, and understandable approach matters.

This paper looks into the ways through which AI helps cybersecurity operations, including how well it finds threats, spots unusual activity, and handles incident responses automatically. AI has been shown to solve key problems, such as more places being exposed to attacks and fewer wrong alarms, by using real-time data and predicting possible threats. Specifically, the study wants to find out the main reasons that make AI good at spotting threats, check how AI-powered tools help users become more aware and ready, and look into the overall effects of AI systems in stopping and dealing with real-time cyber-attacks. Because of this, this paper talks about these parts in improving AI-driven cybersecurity plans for stronger defense against modern threats.

Materials and Methods

The Design

The study conducted a systematic literature review on the role of Artificial Intelligence (AI) in cybersecurity following Durach et al.'s (2017) six-step guideline strategy. The process began with defining research questions to establish the main focus of the study. Next is, required characteristics of primary studies were determined to identify the inclusion and exclusion criteria to filter the relevant literature. Rigorous searches were conducted to retrieve a broad sample of potentially relevant sources. The selected literature was synthesized by developing narrative propositions that explain the context and outcomes of the reviewed studies. Finally, the results presented the synthesized findings, frameworks, and key implications related to the application of AI in cybersecurity.

Research Objectives

This literature review aims to address the research question: how does artificial intelligence improve cybersecurity in detecting and preventing modern cyber-attacks? The review is structured around three primary objectives: (1) to identify the key factors that contribute to the effectiveness of artificial intelligence in detecting cyber threats, (2) to analyze usability of AI-powered cybersecurity measures in improving user awareness and threat response, and (3) to evaluate the impact of AI-driven security systems in preventing and mitigating real-time cyber-attacks.

Retrieving and Selecting Pertinent Literature

To ensure a comprehensive and reliable foundation for this review, a structured approach was employed to retrieve and select relevant literature. The search was conducted using reputable online databases such as IEEE Xplore, ScienceDirect, and Google Scholar, focusing on publications in the domains of cybersecurity, artificial intelligence, and threat detection and prevention. A combination of keywords was used, including “artificial intelligence in cybersecurity”, “machine learning”, “cybersecurity attack prevention with AI”, and “threat detection AI”. The initial search yielded over 47,340 results across all platforms. After removing duplicates and screening titles and abstracts based on predefined inclusion and exclusion criteria such as relevance, publication date, and peer-reviewed status a total of 30 articles were shortlisted. These articles were then subjected to full-text analysis, resulting in a final selection of 20 articles deemed most relevant and insightful for the study's objectives.

Synthesizing the Literature

The selected pertinent literature was published from the year 2020-2025, ensuring that the most recent and relevant studies are considered. A total of 20 articles were identified and met the established criteria, making them suitable for inclusion in the review. Data was systematically extracted from each of these studies, with the key findings analyzed and synthesized, which are then summarized and presented in Tables 1-3 for a comprehensive overview.

Table 1. Table synthesis about the key factors influencing AI’s effectiveness in threat detection.

Category	Authors	Key Factors Influencing AI Effectiveness in Threat Detection
Machine Learning & AI Techniques	Ali et al. (2022); Al Siam et al. (2024); Kavitha & Thejas (2024); Okdem & Okdem (2024); Rathore et al. (2021)	Machine learning, deep learning, NLP, pattern recognition, anomaly detection
Data & Model Quality	Sarker et al. (2020); Abdullahi et al. (2022); Salem et al. (2024); Zhang et al. (2022)	Data quality, feature selection, model accuracy, explainable AI, reduced false positives

Real-Time Detection & Predictive Analysis	Ansari et al. (2022); Ahsan et al. (2022); Ankalaki et al. (2025)	Real-time data processing, predictive analytics, automated response, temporal analysis
Threat & Risk Identification	Ahmad et al. (2024); Jada & Mayayise (2023); Kaur et al. (2023); Wen et al. (2024)	Phishing detection, risk identification, vulnerability detection, system monitoring
Zero-Day & Advanced Attack Detection	Ali et al. (2022); Kavitha & Thejas (2024)	Zero-day attack detection, adversarial attack detection
Automation & Incident Response	Ullah et al. (2024); Perwej et al. (2021)	Incident response automation, alert prioritization, early attack detection
Organizational & Human Factors	Bughin (2024); Khan et al. (2024)	AI readiness, governance, developer skills, trust, cybersecurity strategy
Security Infrastructure & Technology Integration	Ullah et al. (2024); Khan et al. (2024)	Blockchain security, cybersecurity technologies integration

Table 2. Table synthesis about the usability of AI tools in improving user awareness and response

Category	Authors	Usability of AI Tools
Real-Time Monitoring & Alerts	Ahmad et al. (2024); Ahsan et al. (2022); Rathore et al. (2021); Salem et al. (2024); Perwej et al. (2021)	Real-time monitoring, alerts, automated response, faster threat detection
Automation & Decision Support	Al Siam et al. (2024); Jada & Mayayise (2023); Wen et al. (2024)	Automation of security tasks, reduced workload, automated decision-making
Threat Detection & Prediction	Ankalaki et al. (2025); Ansari et al. (2022); Kavitha & Thejas (2024)	Predictive threat intelligence, user behavior analytics, adaptive learning
Integration with Security Systems	Ali et al. (2022); Khan et al. (2024); Bughin (2024)	Integration with IDPS, workflow integration, developer usability
User-Friendly Interface & Explainability	Okdem & Okdem (2024); Sarker et al. (2020); Zhang et al. (2022)	User-friendly interface, explainable AI, understandable outputs
Access Control & Security Management	Kaur et al. (2023); Ullah et al. (2024)	Access control, anomaly detection, security management
Data Analysis & Threat Identification	Abdullahi et al. (2022); Ullah et al. (2024)	Large data analysis, anomaly detection, threat identification

Table 3. Table synthesis about the effectiveness of AI in preventing and mitigating real-time cyber-attacks.

Category	Authors	Effectiveness of AI in Cybersecurity
Threat Detection & Prevention	Ahmad et al. (2024); Ali et al. (2022); Okdem & Okdem (2024); Rathore et al. (2021); Salem et al. (2024); Sarker et al. (2020)	Early threat detection, phishing/malware detection, anomaly detection, improved detection accuracy
Real-Time Response & Mitigation	Ahsan et al. (2022); Ansari et al. (2022); Kaur et al. (2023); Wen et al. (2024)	Faster response time, automated incident response, real-time mitigation
Predictive & Adaptive Security	Ankalaki et al. (2025); Jada & Mayayise (2023); Khan et al. (2024); Kavitha & Thejas (2024)	Predictive analytics, adaptive defense, evolving threat detection
Accuracy & False Positive Reduction	Al Siam et al. (2024); Khan et al. (2024); Zhang et al. (2022)	Reduced false positives/negatives, improved model accuracy
Automation & Security Management	Perwej et al. (2021); Ullah et al. (2024)	Automated monitoring, data protection, blockchain security
Organizational & Productivity Impact	Bughin (2024)	Improved productivity, code quality, development efficiency

Discussion

3.1 The key factors that contribute to the effectiveness of artificial intelligence in detecting cyber threats.

Artificial Intelligence (AI) has been a major support in the field of cybersecurity, especially in the detection of threats. The strength of AI in this sector has been shown in several studies to be influenced by technical, process-related, and data-related factors. This area was shaped particularly by the effectiveness of Artificial Intelligence using unsupervised learning and a combination of Machine Learning (ML) and Deep Learning (DL) algorithms. AI powered by these technologies can not only

learn from the past attack data but also recognize the patterns that can be considered as attack signals. According to Ahsan et al. (2022) and Ali et al. (2022), ML algorithms have the advantage of detecting new/unknown attacks and spotting strange network activity, which are impossible to track by traditional methods. The other major factor influencing the situation is the very ability of AI as well as its speed to process and handle the large amounts of data that are generated in real-time. A conclusion drawn from the studies by Ansari et al. (2022) and Ankalaki et al. (2025) is that such processing as well as predicting future patterns help the systems to predict and prevent the threats. The detection process can, in addition, be made better by the use of deep learning and time-based analysis in terms of identifying dangers that occur after time. This development to the detection of threats in real-time would be a great advantage to the system's reliability and performance in protecting the cybersecurity. Another thing that determines the end of AI's threat detection efforts is the pulling out of important details and the spotting of patterns. Ahmad et al. (2024) has suggested that the AI tool can pull important details of a phishing scam from how the email is written, the website design, and the message content. Part of this, Al Siam et al. (2024) explains that AI systems are working on the basis of natural language processing (NLP) to interpret the messages' data and solve the issue of trick-based attacks and the malicious intent in the messages. Many forms of data such as text, visual, and behavioral, that AI has to learn from, and to which it can be adapted, increase the range of detections of the different types of cyberattack types that exist today.

Moreover, the quality of the data used in AI models' training step has a significant impact on the performance of the models. Sarker et al. (2020) are of the opinion that a good quality, diverse, and properly labeled dataset can make the model better at anomaly detection. Having bad or incomplete information may significantly increase the number of wrongly flagged threats, eventually resulting in decreasing the system's overall effectiveness. Perwej et al. (2021) and Salem et al. (2024) state that it also underscores the necessity of the development of detection algorithms so as to lessen wrong flags while at the same time increasing the reliability of the overall system. Last but not least, the addition of AI to current cybersecurity systems is more fruitful when there is good organizational preparation and clear system behavior. Zhang et al. (2022) states the importance of Explainable AI (XAI), which means that models do not only detect threats but also give explanations in a clear and understandable way. This practice of cybersecurity professionals gives them more confidence and decision-makers, as they can now understand clearly and respond quicker during an occurrence of a threat and also faster in their analysis.

Artificial Intelligence (AI) has transformed cybersecurity and reduced the chances of cyber threats by a considerable margin. The capability of AI to process large amounts of data in real-time and detect complex patterns has turned AI into the vital tool of contemporary cybersecurity strategies. AI systems are used in the form of Machine Learning (ML) and Deep Learning (DL) algorithms which allows them to analyze past attack data and identify anomalies that can indicate possible threats. The technologies especially detect the new/unknown attacks and the strange behaviors within the network which could previously be missed by the traditional methods. Theoretical data crucial in training AI models needs to be significant. Threats detection can be enhanced using high-quality, diverse and well-labeled datasets but low-quality data may misclassify threats as either a flagged or missed alert. Organizations, therefore, have to focus on data quality in the maintenance of AI-based cybersecurity methods.

Natural Processing Language (NLP) is a technology that can improve AI text reading and text analysis abilities and thus is very useful in spotting phishing and trick-based attacks. An AI system can find possible dangerous or harmful content in emails, chat messages, or even social media posts because, just like any other algorithm, it has an unusual skill to find patterns in the language. This helps organizations to see the threats at an early stage and protect the users from being tricked through messages that look so real. In addition to that, Explainable AI (XAI) is important in cases when the system wants to be clear about how the decision-making process in AI systems works. As cybersecurity teams understand the reasons why an AI found a threat, or why it did a certain action in an easy-to-understand way, it builds trust in the technology. XAI not only supports accountability but also helps the users of XAI to make faster and more accurate decisions in a high-stress situation where fast actions are important.

3.2 The usability of AI-powered cybersecurity measures in improving user awareness and threat response.

The issue of security is important in the current cyber life. The internet has become one of the most commonly used channels of studying, earning, job hunting and communicating with other people and thus exposed users to even higher levels of cyber attack. Data Intrusion and system intrusion is quite a serious threat to organizational and individual security. Artificial Intelligence (AI) is present in the current cybersecurity systems approaching illegal activities increasingly closely so that they can be performed freely. Because of such a threat, specialists are transferring AI to cyber protection systems. This is due to the fact that besides speeding up the process of identifying the threats, AI is also being credited with the ability to alert users about possible threats and reinforce their response mechanisms. The intelligence behind cyberattacks of the past can be used to train AI to seek patterns and other anomalies typically associated with harmful operations. Using big data, the AI applications can identify the unusual activity, which can be the illegal transfer of the data or unusual use of the network and issue the alert notifications to clients or administrators real-time. The advantages of collaboration with AI systems are that they work tirelessly unlike human analysts and, therefore, they could offer non-stop monitoring and real-time notification of potential security risks.

According to Alshehri (2024), the real-time spotting of threats with the approach of AI systems makes possible their immediate stopping and response to attack methods which are newly introduced. The use of AI in cybersecurity systems has changed the traditional practices applied to identify threats. Traditional threat detection methods are not effective against new or new types of threats. On the other hand, AI applies to the algorithms of machines able to handle the new and random attack patterns fully, thus making the system composing the cybersecurity more reliable. Also, AI future prediction tools give the

opportunity to prevent a possible attack and stop it before launching, so active defense becomes simpler. Being able to predict is particularly important in the current state of threat where cyberattacks become more dangerous and varied in their methods. Besides detection, AI is very important in terms of educating users on cybersecurity. The AI systems give instant suggestions and warnings in real-time to help users become aware and avoid the threats to be faced. As an example, in the case of the user having visited a suspicious link, AI may step in with a warning concerning the dangers that are about to be experienced, with a particular goal of having safer internet usage. This being a preventive measure in the long run would lead to awareness among the users and their informed approach towards online dangers. It has also been researched that cybersecurity awareness is supported to a great degree with the help of AI-based technology that can be used to inform users (Ansari, 2022).

The AI will be in a position to identify malicious messages, which attempt to deceive users into divulging personal information. It also carries out scanning of messages and informs the user to be in a position to know the logic behind how a message could be lethal. Today it is particularly important, as even phishing attacks are becoming more realistic-looking. An example of how Natural Language Processing (NLP) tools and machine learning can help identify such threats and offer sufficient protection is depicted in a research paper by Ansari (2022). However, warnings are not all about AI. AI can also teach people by training them and practicing them on a regular basis and hence rendering them prepared and more knowledgeable about cybersecurity. AI can model and optimize the ways in which humans respond to threats, reducing those attacks that are successful in the future. It can also assist cybersecurity teams as it can perform routine activities, generate comprehensive reports and analyze security incidents. Human experts therefore can concentrate on complex issues and this makes the process more efficient. AI mechanisms also provide reports straight-forward and meaningful to enable security departments to react faster and clever in assaults. The threat control is boosted by cutting the response time and failures in the security centers (Alshehri, 2024). The use of AI in cybersecurity has become one of the biggest subjects today. Although AI can process the large volume of information quickly, human experts ought to interpret the results and make the most appropriate decisions. So, AI is an effective assistance that supplements, but does not substitute people. The two can develop better and efficient cybersecurity.

But while helpful, security use of AI has not been without problems. The use of AI algorithms in action is largely a function of the scope and quality of data on which they run. As a result, unbalanced data may lead to unbalanced prediction and missed threats. In addition, the user interface of some of these AI technologies may limit their use by non-experts, and thus it is the duty of straightforward and easy-to-understand designs. The designers must lead the way in making things clear and simple to use so that AI-based security solutions can be used and helpful to everyone. Second, as cybercrime criminals also use AI to design attacks, there is a need for constant new ideas in AI defense (Sarker et al., 2023). And the second challenge, with all others taking a back seat, is enemy attacks, where criminals try to trick AI systems to avoid being caught or to create fake outputs. These attackers try to change the inputs so much that the inputs are made to fool AI models into getting it wrong or missing the threat. This calls for creating strong and tough AI models that can fight against these attacks. Ethical issues like keeping users' information private and making sure AI isn't used in the wrong way also have to be handled wisely to build trust and support with rules.

The use of AI in cybersecurity operations is already a huge step into securing digital properties and user data. AI improves the ability to detect threats, warns the user about possible dangers, and helps the team manage security incidents more efficiently. However, to achieve all these advantages, the problem of data quality, the ease of use of tools, and changing strategies of cyber attackers should be addressed. Being constantly improved and used wisely, AI-based cybersecurity solutions can ensure effective early protection of the user, allowing one to move freely throughout the online world and feel safer. It is important that people never forget that AI has promising directions in the future that involve raising awareness in users and better response to threats. The progress in technologies like deep and reward-based learning will never leave their hands and will have the most powerful effect.

3.3 The impact of AI-driven security systems in preventing and mitigating real-time cyber-attacks.

In the rapidly changing environment of cybersecurity, artificial intelligence (AI) is becoming an indispensable resource and utility in the struggle against ever-increasingly advanced cyber threats and attacks. AI-powered security systems are extremely valuable and vital for preventing and mitigating cyber-attacks in real-time, given that threat detection depends on speed and accuracy. Of these, artificial intelligence (AI) is growing as a game changer or disruptor. This conversation examines the use of AI-driven security systems in real-time threat detection, prevention, and response, drawing on findings from contemporary literature.

We noted a consistent finding across multiple studies that AI drastically improves accuracy of threat detection. In general, AI systems, especially those which are machine learning driven algorithms, are able to evaluate and digest vast amounts of data and reveal subtle patterns in detection, which human-driven controls and measures frequently do not. Sarker et al. (2020) demonstrated that machine learning and data science was able to reveal complex and hidden attack vectors, which typically escaped identification by manual or rule-based detection. Khan et al. (2024) and Ahmad et al. (2024b), for example, demonstrate that AI was able to reduce false positives and false negatives, which means that AI is able to show threatening problems while minimizing unnecessary alerts. This is key to maintaining system integrity and operational efficiency in real-time paradigms. The effectiveness of AI in detecting subtle and complex attacks is a key strength. As shown, AI can manage large-scale data analysis

with greater speed and accuracy than human analysts. This results in quicker threat identification and more informed decision-making.

Building on these detection capabilities, the integration of Artificial Intelligence into cybersecurity systems has significantly optimized detection, prevention, and response efforts. Based on a chosen literature review, it is safe to say that AI-based machines are at the forefront of detecting cyber threats in real-time with increased efficiency, flexibility, precision, and velocity. A handful of studies examine the validity of machine learning (ML) and deep learning (DL) for real-time intrusion detection systems (IDS). For instance, Abdullahi et al. (2022) found that machine learning (ML) classifiers, such as random forest (RF), support vector machines (SVM), and extreme gradient boosting (XGBoost), have high effectiveness in differentiating legitimate traffic from malicious traffic, citing accuracy levels above 95%. Their results, based on feature selection in conjunction with ensemble learning, suggest that a hybridized approach to AI far exceeds traditional signature-based approaches to detection. In a comparable study, Ahmad et al. (2024b) confirm AI's real-time detection validity within IDS as well by noting how AI assists with anomaly detection via outlier detection from multifaceted datasets. Their use of sophisticated deep learning techniques, namely Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), validates that these processes can operate with temporally based streams and positively encourage effective adaptive action against newly emerging threats, such as zero-day attacks. Real-time detection is critical as threat landscapes increasingly grow more sophisticated, and as such, Ali et al. (2022) appreciate how AI is applied to malware detection as well as phishing. They conclude that hybrid models utilizing natural language processing (NLP) alongside deep learning (DL) could yield more accurate application reviews and behavior analysis for nefarious behavior faster than classic antivirus applications can flag. Regardless of the study's focus on malware or phishing, Ali's findings suggest that AI successfully reduces response time and false positives. Furthermore, Ahsan et al. (2022b) developed a deep reinforcement learning model that allows for real-time learning as events happen to the network for self-healing capabilities against attacks; this fluid approach supports the notion that an ideal future of transportation networks will head in a direction not only of reactive measures but also proactive awareness capabilities.

Another notable development is the capability of AI powered systems to continuously monitor networks and independently react to security threats in time without human intervention being necessary. It has been shown in research studies conducted by Ansari et al. Ankalaki et al.. Ali et al. that AI has the capacity to identify deviations, from the norm and trigger automated response procedures promptly These systems not only pinpoint irregularities but also take appropriate action swiftly which leads to a quicker response time and helps in minimizing any potential harm caused According to Jada and Mayayise (2023) AI boosts response time by enabling systems to quickly adjust to emerging threats, a capability when dealing against zero day attacks. Complementing these operational enhancements, Al Siam et al. (2024) offered a comprehensive overview of the role that AI plays in digital forensics and threat intelligence. They highlighted how the scalability of AI models enables them to be deployed in large-scale enterprise environments. They found that the combination of human analysts and AI agents yields improvements in incident response workflows and risk prioritization. Even with the encouraging outcomes, constraints linger. The functioning of AI-driven systems is largely contingent upon high-quality, labeled datasets, as both Ahsan et al. (2022b) and Ahmad et al. (2024b) assert. Furthermore, Ali et al. (2022) offer a word of caution, saying that adversarial attacks can be used to compromise AI models. This underscores the critical need for more continuous model validation and more robust training pipelines.

One of the distinct advantages of AI-driven solutions is their capacity for ongoing adaptation to new data and new attack patterns, allowing AI solutions to incrementally improve through updating their models from incoming data, addressing zero-day and new cyber threats (Ali et al., 2022; Kavitha and Thejas, 2024). AI's self-improvement capability also emphasizes its role in continually advancing methods of engagement in the Cyber domain, out-pacing Cyber threats generated by human actors with a near constant arrival of new capabilities and methods. Beyond detection and response, AI solutions act as a form of predictive capabilities that enhance preparedness. Ankalaki et al. (2025) even advocate the application of predictive simulations, whereby the model assumes the attacker's perspective to predict threats, whereas Ansari et al. (2022) outline how threat intelligence and automated analysis can shape and manage risk before materializing. They advocate a shift away from reactive-based models and that a strategy enables a proactive approach to prepare for escalation and response.

Despite the level of progress in the application of AI in Cyber Security, AI- systems confront multiple serious challenges. One identified limitation is regarding model interpretability. Kavitha and Thejas (2024) and Zhang et al. (2022) suggest that the opacity of AI decision-making can undermine stakeholder trust and ability to validate the actions of the system. A second limitation is the system's vulnerabilities to adversarial example attacks. Adversarial inputs, typically malicious, are a mechanism that can deceive AI models. Bughin (2024) and Ullah et al. (2024) even advocate for applied frameworks that integrate AI with existing technology like Blockchain which would address data security-related issues and weaknesses within all centralized systems. Certainly, these observations provide rationale for ongoing research that studies how to create more transparent, secure, and resilient AI-based solutions. In practice, hybrid applications could provide decentralized, attributable systems which stand as tamper-proof and more workable in terms of overall trustworthiness and data ownership. Blockchain, as a decentralized, immutable technology, interfacing with existing systems for the descriptive and predictive capabilities of AI could bolster the nature for threat detection and response, in real time.

AI has fundamentally changed the landscape of cybersecurity to some degree. The short-term and real-time threat detection, prevention, and response capabilities AI provides have greatly enhanced existing tools and made detecting complex threats easier, decreased false-positives, and increased efficiency. More than just simple detection and prevention, AI can predict potential

threats before they happen, which is a paradigm shift in the industry from reactive to preventive measures. Looking ahead, the advancements in AI combined with the development of other emerging and innovative technologies (like Blockchain) will have an even more positive impact on system security and protection. A combination of AI and existing technologies is interesting too; a hybrid approach of AI will only continue to have large potential for more comprehensive and stronger defenses against increasingly advanced cyber threats.

Conclusion

This study discusses that Artificial Intelligence (AI) has in the current world turned to be an effective tool in cybersecurity, particularly on its implementation in detecting cyberthreats as it yields more precise warnings which may be applied to mitigate attacks. This translates to the fact that technology is currently shifting towards a more intelligent manner of protection. Artificial intelligence has become a fundamental element of the most recent cyber security strategies. However, as cyber criminals start using AI as well to conduct malicious activities, the defenders need to work tirelessly by auditing the models, addressing bias, and building stronger systems against AI-related attacks. Regulatory and ethical regimes should also be able to keep pace to defend privacy, make them accountable, and avoid any misuse. Organizations need to begin the adoption of AI systems that are safe and simple to integrate. They also need to advance AI so that it will work in conjunction with other potent technologies such as blockchain and ensure that it is capable of coping with new and superior cyber-attacks. Hence, when seeking solutions to present issues. Future AI tools should be designed with the following features: enhanced user awareness, ease of installation, and the ability to improve over time. This will assist in the creation of an online safe world for everyone.

References

Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in Internet of Things Using Artificial intelligence Methods: A Systematic Literature review. *Electronics*, *11*(2), 198. https://doi.org/10.3390/electronics11020198

Ahmad, S., Zaman, M., Al-Shamayleh, A. S., Kehkashan, T., Ahmad, R., Abdulhamid, S. I. M., Ergen, I., & Akhunzada, A. (2024b). Across the spectrum In-Depth review AI-Based models for phishing detection. IEEE Open Journal of the Communications Society, 1. https://doi.org/10.1109/ojcoms.2024.3462503

Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022b). Cybersecurity Threats and their Mitigation Approaches Using Machine Learning—A Review. Journal of Cybersecurity and Privacy, 2(3), 527–555. https://doi.org/10.3390/jcp2030027

Ali, S., Rehman, S. U., Imran, A., Adeem, G., Iqbal, Z., & Kim, K.-I. (2022). Comparative evaluation of AI-based techniques for zero-day attacks detection. Electronics, 11(23), 3934. https://www.mdpi.com/2079-9292/11/23/3934

Al Siam, A., Alazab, M., Awajan, A., & Faruqui, N. (2025). A comprehensive review of AI's current impact and future prospects in cybersecurity. IEEE Journals & Magazine. https://ieeexplore.ieee.org/document/10836696

Ankalaki, S., Rajesh, A. A., Pallavi, M., Hukkeri, G. S., Jan, T., & Naik, G. R. (2025). Cyber-attack prediction: From traditional machine learning to generative artificial intelligence. IEEE Access, 1. https://ieeexplore.ieee.org/document/10909100

Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022, September 1). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A literature review. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4323317

Bughin, J. (2024). The role of firm AI capabilities in generative AI-pair coding. Journal of Decision System, 1–22. https://doi.org/10.1080/12460125.2024.2428187

Jada, I., & Mayayise, T. O. (2023). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. Data and Information Management, 8(2), 100063. https://www.sciencedirect.com/science/article/pii/S2543925123000372?via%3Dihub

Kaur, R., Gabrijelčič, D., Klobučar, T., & Laboratory for Open Systems and Networks, Jožef Stefan Institute, Ljubljana, Slovenia. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, 97, 101804. https://www.sciencedirect.com/science/article/pii/S1566253523001136?via%3Dihub

Kavitha, D., & Thejas, S. (2024). AI enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation. IEEE Access, 1. https://ieeexplore.ieee.org/document/10747338

Khan, M. I., Arif, A., & Khan, A. R. A. (2024). The most recent advances and uses of AI in cybersecurity. https://journal.mediapublikasi.id/index.php/bullet/article/view/4540

Okdem, S., & Okdem, S. (2024). Artificial intelligence in Cybersecurity: A review and a case study. In Christos Bouras (Ed.), Appl. Sci. (Vols. 14–14, pp. 10487–10487). https://doi.org/10.3390/app142210487

Perwej, D., Abbas, S. Q., Dixit, J. P., Akhtar, N., & Jaiswal, A. K. (2021). A Systematic Literature review on Cyber Security. International Journal of Scientific Research and Management (IJSRM), 9(12), 669–710. https://hal.science/hal-03509116/

Rathore, M. M., Shah, S. A., Shukla, D., Bentafat, E., & Bakiras, S. (2021). The role of AI, machine learning, and big data in Digital twinning: A Systematic literature review, Challenges, and opportunities. IEEE Access, 9, 32030–32052. https://ieeexplore.ieee.org/abstract/document/9359733

Salem, A. H., Azzam, S. M., Emam, O. E., Abohany, A. A., & Salem et al. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. Journal of Big Data (p. 105) [Journal-article]. https://doi.org/10.1186/s40537-024-00957-y

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00318-5

Ullah, Z., Waheed, A., Mohmand, M. I., Basar, S., Zareei, M., & Granda, F. (2024). AICYber-Chain: Combining AI and blockchain for improved cybersecurity. IEEE Access, 12, 142194–142214. https://doi.org/10.1109/access.2024.3463976

Wen, S.-F., 1, Shukla, A., 2, & Katt, B., 1. (2024). Artificial intelligence for system security assurance: A systematic literature review. In International Journal of Information Security (Vol. 24, p. 43) [Journal-article]. https://doi.org/10.1007/s10207-024-00959-0

Zhang, Z., Hamadi, H. A., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable Artificial intelligence Applications in Cyber Security: State-of-the-Art in Research. IEEE Access, 10, 93104–93139. https://ieeexplore.ieee.org/document/9875264

How to Cite this Chapter (APA 7):

Bayhon, N. R., Bayhon, J. R. G., Lopez, K. J., Mendoza, J., Romano, A., & Dandan, I. (2026). AI-based cybersecurity systems: A review of intelligent threat detection and prevention approaches. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 46–53). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 8: Applications of Artificial Intelligence in Cybersecurity: A Systematic Literature Review

Norman Steven P. Dominguez, Leixhelle Ann P. Saito, Jom Karlo M. Verzosa, Bea Bianca Ybañez & Gersom N. Baradi
Cavite State University – Silang Campus, Cavite, Philippines

Abstract. *Artificial Intelligence (AI) is revolutionizing cybersecurity with real-time threat detection, predictive analysis, and automated incident response in a vast variety of complex environments. In this paper, we aim to explore the role of AI in cybersecurity through the identification of major AI-driven tools and techniques, assessment of their value, exploration of limitations, and exploration of proposed solutions. We systematically review the literature (SLR) by utilizing databases (IEEE Xplore, ScienceDirect, Google Scholar, etc.) while adhering to the PRISMA guidelines to select and categorize relevant literature between 2020 and 2025. The research indicates that AI logics (machine learning, deep learning, federated learning and explainable AI) increase efficiencies in intrusion detection, anomaly detection, and mitigating against cyber threats. High resource usage, ethics, adversarial attack resilience and data reliance remain challenges with AI approaches. Current approaches to mitigate these issues include, but are not limited to: adversarial training approaches, hybrid human and AI systems, continuous learning, frameworks for explainable artificial intelligence, and governance and standards. The main goal of these approaches is to ensure accuracy, ethical compliance during generalisation to real-world issues, and interpretability. The implications are several, in regards to technology, economics, and governance: AI enables strong cyber security ecosystems (technologically); AI technology provides affordable scalable solutions for protection (economically); and AI standardization/regulatory collaboration will be operationally necessary (government). Future implementations should focus on transparency, transdisciplinary education, and the development of lightweight, adaptive AI systems to ensure resilience and trust in digital security.*

Keywords: *Artificial Intelligence, Cybersecurity, Adversarial Training, Explainable AI, Federated Learning*

Introduction

The relentless expansion of smart technologies and the Internet of Things (IoT) has fundamentally reshaped the modern digital landscape, enabling seamless data exchange and automation across a wide spectrum of applications. From consumer electronics to critical infrastructure, IoT devices have become integral to everyday life, enhancing efficiency and convenience (Jun et al., 2021; Ahmed et al., 2022). This interconnectedness, while offering numerous benefits, has also significantly broadened the attack surface for cybercriminals, creating a complex and evolving security challenge. Traditional approaches in cybersecurity, which frequently depend on human operators and rule-based systems, are facing more and more challenges due to the dynamic nature of modern threats and the sheer volume of data being processed. At the same time, Artificial Intelligence (AI) has emerged as a transformative technology, drawing attention for its ability to process large amounts of data and recognize complex patterns. It is becoming clearer that it has the potential to revolutionize various fields, including cybersecurity. As the complexity and amount of cyberattacks have increased, utilizing AI in cybersecurity is even more crucial. Modern systems are increasingly interdependent, which leads to a greater complexity in security measures to account for that. Due to the growth in the number of connected devices, an increased potential for cyberattacks is introduced. More devices mean more data to analyze, which can overwhelm traditional security systems. This is increasing the demand for more sophisticated cybersecurity solutions.

Today's cybersecurity systems are limited in their capabilities to address the fast-evolving and increasingly sophisticated cyber threat landscape. Traditional rule-based defense and human operators—while absolutely necessary, are typically inadequate against increasingly complex cyber threats like zero-day exploits, Advanced Persistent Threats (APTs), and ransomware (Chakraborty et al., 2020; Conti et al., 2018)—struggling under the volume and frequency of dangerous cyber-attacks, and/or subtle indications of breach. These challenges are further intensified as the IoT ecosystem continues to expand, with over 30 billion devices expected by 2030 (Statista, 2023; Mosenia & Jha, 2017), therefore overwhelming existing security monitor mechanisms. The growing number of devices are a byproduct of software that is outdated and operating in small, resource-constrained environments, that have obsolete memory, and that do not comply with standardized IoT security protocols, thereby creating openings for cybercriminals (Alrawais et al., 2017). Cyber threats today are becoming more dynamic, harder to secure, and increasingly complex. These evolving attacks are moving at a significant pace, and defeating current static or cyber-defensive security measures making them exceedingly inadequate against more adaptable threats like polymorphic malware that find increasingly creative ways to avoid being detected. As a result, many systems lack the real-time capability required to respond to emerging threats on-line effectively; hence the need for proactive real-time cybersecurity management (Pipyros et al., 2017).

In order to address these critical limitations in existing cybersecurity frameworks, this study analyzes the cybersecurity conditions in the smart technology era while concentrating on the development and evaluation of protective strategies for networked systems. By reviewing security tools, security methods, and security frameworks developed for IoT devices and networks available on the internet, this research aims to measure their effectiveness in stopping and/or limiting cyberattacks.

Materials and Methods

In order to study how Artificial Intelligence (AI) is being utilized to strengthen cybersecurity in the growing Internet of Things (IoT), the researcher undertook a systematic literature review (SLR). SLR is an important method to provide scientific knowledge and knowledge synthesis of the body of existing literature (Linnenluecke et. al., 2019; Snyder, 2019). There is also a method to contribute to the literature in a transparent systematic way guided by available evidence-based principles to conduct SLR (Page et. al., 2021; Shaffril et. al., 2020). The SLR process allowed us to outline clear processes for searching and screening for relevant studies and to determine inclusion criteria (Page et. al., 2021; Muka et. al., 2019). The aim of the SLR was to establish the importance of AI to improve cybersecurity practices in IoT environments to develop a thorough understanding of the subject. By doing an SLR we were able to systematically screen journal articles from credible sources (Linnenluecke et. al., 2019).

2.1 - Database Resource

We performed a literature search in several electronic databases (March-April 2025): IEEE Xplore, ScienceDirect, ResearchGate, neliti, link.springer, and Scopus. We used the title of articles in conjunction with keywords within these databases. Also, the main electronic database that we regularly use is Google Scholar. We hand-searched literature with the same sets of keywords to ensure maximum inclusion of relevant and associated literature to the paper.

2.2 - Protocol development

The process undertaken by this Study Literature Review (SLR) follows the process of the "Preferred reporting items for systematic review and meta-analysis" (PRISMA) review. PRISMA was originally launched in 2009 and was later updated to PRISMA 2020. This includes the revised reporting guideline to reflect advancements in how studies are identified, selected, evaluated, and summarized (Page, M. J., 2021). PRISMA is a useful and valuable tool and/or resource for researchers who are aiming to improve the clarity and impact of their literature reviews and meta-analyses. It includes 27 specific evaluation criteria designed for randomized trials (Page, M. J., 2021).

2.3 - Identification and screening

The first step in the identification process is finding words and/or phrases that have the same meaning or thought and are related to each other, including their word variations. In the process of screening and identifying the literature, researchers enlarge crucial keywords and use 3 sets of phrases for the search. When the researchers use more than one keyword and phrase, it ensures the inclusivity of every literature related to the research, for the database might find publications that are more relevant to them. In searching, we use 3 different sets of keywords. The keywords and combinations that were used in searching pieces of literature are the following:

Set 1 – ("Literature about Cybersecurity and Artificial Intelligence")

Set 2 – ("Literature about the roles of Cybersecurity and Artificial Intelligence")

Set 3 – ("Literature about the AI impacts in Cybersecurity")

We classified each study as quantitative (empirical), qualitative, conceptual, or literature review after compiling the articles for review.

2.4 - Extraction, synthesis, and eligibility

As we progressed with the article and literature identification process, we examined the articles and literature carefully to identify any duplicates and papers irrelevant to our title. The screening phase followed, giving importance to the inclusion and exclusion criteria for our paper. The inclusion criteria for this review encompass peer-reviewed quantitative and qualitative research studies published in English within the last five years. Only journal-type documents that specifically focus on topics related to artificial intelligence (AI) and cybersecurity are considered. Conversely, the exclusion criteria eliminate studies such as surveys and clinical trial studies, as well as any publications written in languages other than English. Additionally, books, essays, reports, short surveys, and other articles lacking sufficient supporting details are excluded. Studies that do not pertain to AI and cybersecurity or were published more than five years prior are also not considered for inclusion.

2.5 - Data abstraction and analysis

The gathered articles searched from different databases underwent a thorough evaluation, screening, and analysis. Emphasizing its particular topic, which addressed the review's objectives and problem. After reading and screening the papers' abstracts, the data were first extracted, and then any papers that didn't focus on AI and/or Cybersecurity were also extracted. After that, the complete articles and papers downloaded were carefully examined to determine their relevance to our review and theme. Figure 1 below shows a comprehensive synthesis of our literature titled: *"The significance of Artificial Intelligence in Cybersecurity".* As shown below, this evaluation was conducted through identification, screening, eligibility, and synthesis.

Table 1. Identifying Literature: AI-driven Key Factors, Techniques, and Tools

Category	Authors	AI-Driven Key Factors, Techniques, and Tools
AI Techniques & Models	Das & Sandhane (2021); Jun et al. (2021); Zhang et al. (2022); Kuzlu et al. (2021); Ozkan-Okay et al. (2024); Khan et al. (2024); Sontan & Samuel (2024)	Machine learning, deep learning, neural networks, reinforcement learning, NLP, anomaly detection, behavior analysis
AI Tools & Algorithms	Charmet et al. (2022); Das & Sandhane (2021); Ozkan-Okay et al. (2024)	TensorFlow, Neural Structured Learning, SHAP, LIME, LEMNA, decision trees, random forest, gradient boosting, CNN, LSTM, DBN, RNN
Cybersecurity Applications	Kaur et al. (2023); Jonas et al. (2023); Khan et al. (2025); Al-Rubaye & Türkben (2024); Camacho (2024); Khan et al. (2024)	Intrusion detection, malware detection, phishing detection, vulnerability management, access management, DDoS detection
Automation & Real-Time Response	Das & Sandhane (2021); Jun et al. (2021); Mijwil et al. (n.d.); Khan et al. (2024)	Automated response, real-time monitoring, threat intelligence, network monitoring
Explainable AI & Transparency	Zhang et al. (2022); Charmet et al. (2022)	Explainable AI (XAI), model interpretability
Predictive Analytics & Threat Intelligence	Markevych & Dawson (2023); Jonas et al. (2023); Khan et al. (2025)	Predictive analytics, behavior analytics, threat prediction
Data Analysis & Big Data Processing	Capuano et al. (n.d.); Camacho (2024); Khan et al. (2024)	Big data analysis, pattern recognition, intelligent agents
Network & System Security	Al-Rubaye & Türkben (2024); Kuzlu et al. (2021)	Network security, MANET security, intrusion detection systems

Table 2. Summary of Literature's efficiency in Detecting and Responding to Threats

Category	Authors	AI-Driven Efficiency in Detecting and Responding to Threats
Real-Time Detection & Monitoring	Das & Sandhane (2021); Jun et al. (2021); Kuzlu et al. (2021); Sontan & Samuel (2024)	Real-time monitoring, anomaly detection, intrusion detection, vulnerability detection
Predictive Analytics & Threat Intelligence	Kaur et al. (2023); Jonas et al. (2023); Khan et al. (2024); Camacho (2024); Khan et al. (2025)	Predictive threat analysis, risk assessment, proactive defense
Automated Response & Incident Response	Das & Sandhane (2021); Khan et al. (2024); Mijwil et al. (n.d.); Sontan & Samuel (2024)	Automated response, automated mitigation, faster reaction time
Accuracy Improvement & False Positive Reduction	Zhang et al. (2022); Charmet et al. (2022); Ozkan-Okay et al. (2024); Kuzlu et al. (2021)	Improved detection accuracy, reduced false positives, better classification
Adaptive & Learning Systems	Markevych & Dawson (2023); Kaur et al. (2023); Kuzlu et al. (2021)	Adaptive learning, behavior-based detection, learning from past attacks
Decision Support & Explainable AI	Zhang et al. (2022); Charmet et al. (2022); Capuano et al. (n.d.)	Explainable AI, decision support, transparency
Cybersecurity Applications	Jun et al. (2021); Capuano et al. (n.d.); Khan et al. (2025)	Malware detection, botnet detection, fraud detection, phishing detection
Efficiency & Performance Impact	Camacho (2024); Ozkan-Okay et al. (2024); Jonas et al. (2023)	Improved efficiency, faster detection, improved system performance

Table 3. Synthesizing literature's challenges and constraints.

Category	Authors	AI-Driven Efficiency in Detecting and Responding to Threats
Adversarial Attacks & Security Risks	Das & Sandhane (2021); Kaur et al. (2023); Jun et al. (2021); Zhang et al. (2022); Kuzlu et al. (2021); Ozkan-Okay et al. (2024); Khan et al. (2025); Khan et al. (2024)	Adversarial attacks, AI manipulation, system vulnerabilities

Data Privacy & Ethical Issues	Kaur et al. (2023); Jun et al. (2021); Zhang et al. (2022); Khan et al. (2024); Mijwil et al. (n.d.)	Data privacy concerns, ethical issues, regulatory compliance
Data Quality & Bias	Das & Sandhane (2021); Kaur et al. (2023); Jonas et al. (2023); Kuzlu et al. (2021); Ozkan-Okay et al. (2024)	Biased data, incomplete datasets, data dependency
High Cost & Computational Requirements	Das & Sandhane (2021); Jun et al. (2021); Zhang et al. (2022); Markevych & Dawson (2023); Ozkan-Okay et al. (2024)	High implementation cost, computational complexity, resource requirements
Lack of Transparency & Explainability	Zhang et al. (2022); Charmet et al. (2022); Capuano et al. (n.d.); Camacho (2024)	Black-box AI models, lack of interpretability, trust issues
Lack of Skilled Professionals	Das & Sandhane (2021); Khan et al. (2024)	Shortage of AI and cybersecurity experts
Adoption & Implementation Challenges	Jonas et al. (2023); Khan et al. (2024); Al-Rubaye & Türkben (2024)	Low adoption rate, regulatory requirements, infrastructure limitations
False Positives & System Limitations	Khan et al. (2024); Sontan & Samuel (2024)	False positives, alert fatigue, limitations of traditional systems

Table 4. Summary of Literature's Proposed Solutions

Category	Authors	Proposed Solutions
Adversarial Training & Model Improvement	Das & Sandhane (2021); Kaur et al. (2023); Jun et al. (2021); Zhang et al. (2022); Kuzlu et al. (2021)	Adversarial training, robust AI models, continuous model updates
Human–AI Collaboration	Das & Sandhane (2021); Kaur et al. (2023); Jun et al. (2021); Mijwil et al. (n.d.)	Hybrid AI-human approach, human-in-the-loop decision-making
Explainable AI & Transparency	Zhang et al. (2022); Charmet et al. (2022); Capuano et al. (n.d.); Ozkan-Okay et al. (2024)	Explainable AI (XAI), interpretable models, transparency
Data Governance & Privacy	Khan et al. (2024); Markevych & Dawson (2023); Mijwil et al. (n.d.)	Data governance, data quality improvement, privacy protection
Automation & AI Security Systems	Sontan & Samuel (2024); Khan et al. (2024); Jonas et al. (2023)	Automation platforms, SOAR systems, automated threat management
Predictive Analytics & Behavioral Analysis	Jonas et al. (2023); Khan et al. (2024); Camacho (2024)	Predictive models, behavior analysis, threat prediction
Federated Learning & Lightweight Models	Ozkan-Okay et al. (2024); Kuzlu et al. (2021)	Federated learning, lightweight AI models
Policy, Regulation & Collaboration	Das & Sandhane (2021); Jun et al. (2021); Khan et al. (2025); Khan et al. (2024)	AI governance, global standards, collaboration among institutions
Resource & Cost Management	Markevych & Dawson (2023); Khan et al. (2024)	Cost optimization, resource planning
Integration with Emerging Technologies	Khan et al. (2024); Al-Rubaye & Türkben (2024)	Integration with blockchain, IoT, AI-based IDPS

Discussion

3.1 The key factors of Artificial Intelligence(AI)-driven techniques and tools over the past 5 years.

The recent advancement in Artificial Intelligence (AI) has significantly reshaped the field of cybersecurity. Numerous studies show that AI-based methods, such as machine learning, deep learning, neural networks, and reinforcement learning are enhancing abilities to identify threats in real-time, manage incidents automatically, and detect complex anomalies across different digital environments. The study looks into the key components, technical approaches, and innovative tools that define AI's impactful role in cybersecurity. Also, it draws on insights from recent research to showcase the advantages and present-day difficulties linked to AI-powered security systems.

One of the key advantages of AI in cybersecurity is its ability to facilitate real-time threat detection and automated response mechanisms. According to Das and Sandhane (2021), AI enables continuous monitoring and immediate identification of anomalies, thus, allowing systems to react swiftly without human intervention. With this capability, it is particularly significant given the rapid pace at which cyberattacks happen. Different instruments such as Google's TensorFlow and Neural Structured

Learning (NSL) represent this innovative progression, improving machine learning models that support detection and defense components.

AI-driven security frameworks show an empowering approach by utilizing machine learning calculations to recognize anomalous activity and behavior designs related to botnet activities. Bertino and Islam (2017) expounds on the ways in which botnets exploit weak IoT gadgets to initiate disseminated cyberattacks capable of paralyzing networks. These AI-powered systems can distinguish harmful activities more accurately and effectively than ordinary approaches, advertising fundamental security for IoT situations that may otherwise be powerless.

Anomaly detection is the fundamental of numerous AI-driven cybersecurity applications.

Chandola et al. (2009) study investigates various techniques used to detect atypical behaviors that deviate from an established baseline, which may suggest unauthorized access or data breaches. The research highlights the significance of machine learning models, particularly unsupervised learning, in situations where labeled datasets are often incomplete or absent, a common challenge in security environments. These flexible algorithms enable cybersecurity tools to identify new threats, improving the capacity to protect against unknown or developing attacks

Expanding into more extensive cybersecurity operations, the use of AI in incident response, threat intelligence collection, and risk assessment showcases AI's ability to analyze real-time data from multiple sources, allowing security teams to rapidly identify essential threats and make quick, informed decisions. By integrating AI-powered predictive analytics with automation, organizations can improve cyber resilience, addressing threats more effectively while reducing the time required for human intervention (Gonzalez & Kauffman 2020).

Different AI techniques in cybersecurity concentrate on machine learning, deep learning, and natural language processing (NLP). Natural Language Processing (NLP) plays a vital role in examining unstructured text information from security logs and threat intelligence documents (Kumar & Singh 2020). By means of automated analysis and parsing, AI can uncover important insights that aid in identifying phishing attempts, social engineering tactics, and other subtle attack vectors that traditional security techniques may overlook. This capability expands AI's range beyond mere network security, incorporating the often overlooked human elements of cyber threats.

Additionally, Moustafa and Slay (2015) assess numerous Intrusion Detection Systems (IDS) and emphasize the benefits obtained through the integration of AI methods. The analyses indicate that IDS powered by machine learning excel at differentiating between genuine and harmful activities in network traffic, lowering false alarm rates that are typical in conventional signature-based systems. This shift results in more dependable and adaptive intrusion detection, guaranteeing that security measures stay aligned with attackers who utilize progressively covert techniques.

The role of AI goes beyond just detection and prevention; it is also essential in handling cybersecurity incidents after they happen. Deep learning models' capability to identify intricate patterns in attack behavior allows for proactive defense strategies by predicting attacker actions based on past incidents, enhancing readiness and response results. AI speeds up the categorization and ranking of threats during incident responses, assisting security teams in determining which incidents require urgent action. (Panda & Patra 2020).

Regarding data breaches, an increasing worry due to the significant expenses of compromised sensitive data, Ransbotham and Mitra (2018) examine how AI enhances protective strategies. For example, AI facilitates adaptive access controls and biometric behavior assessment to consistently authenticate user identities without obstructing genuine users. These sophisticated controls reduce the likelihood of unauthorized access while enabling smooth operation, highlighting AI’s role as an enhancer in executing efficient policies within complex organizational structures.

One of the AI's abilities in sophisticated malware classification and synchronized threat detection highlights that many modern attacks utilize polymorphism or different evasion techniques that weaken static defenses. AI-powered solutions evolve instantly to these evolving threats by recognizing and learning subtle variations, thereby improving detection accuracy (Sarker & Ahsan, 2020).

Numerous AI methods function as "black boxes," rendering it challenging for security analysts to completely comprehend or have confidence in the results. The advantages and ongoing issues of AI in cybersecurity highlight the importance of transparency and clarity in AI models. Sharma and Gupta (2021) advocate for creating interpretable AI systems that utilize explainable output methods to preserve cybersecurity efficacy while facilitating human supervision. Therefore, this balance is crucial for broad adoption as operators need both automated support and responsibility in their decision-making processes.

The study of Zhou and Zhang (2021) provides a recent overview of the AI-cybersecurity landscape, advocating for hybrid models that combine supervised, unsupervised, and reinforcement learning to enhance robustness and adaptability. Also, highlight promising integration paths with emerging technologies such as blockchain for securing transactions and edge computing for decentralized data protection. The forward-looking analysis suggests that overcoming challenges related to AI scalability and resistance to adversarial attacks will be key to realizing its full potential in cybersecurity.

3.2 The efficiency of AI-driven cybersecurity systems in detecting and responding to cyber threats.

Artificial Intelligence (AI) fundamentally changed the field of cybersecurity by providing transformative solutions to improve threat detection, prevention, and response. The application of AI in cybersecurity supports techniques like predictive analytics, real-time monitoring, behavior-based detection of threats, and automated threat response mechanisms, which are demonstrated to

perform more efficiently in cybersecurity threat management applications. It will analyze the improvement AI brings to cybersecurity systems, focusing on intrusion detection, anomaly detection, and fraud prevention, based on existing literature and research findings that bring out AI's role in cyber defense.

One of the primary advantages of AI in cybersecurity is its ability to preempt threats before they actually occur. AI is able to do so because it is (often) driven by machine learning (ML) , and machine learning can analyze vast amounts of data to detect patterns and unusual behavior that typically presages a cyber-attack. Buczak et al. (2016) suggest that machine learning enables security systems to learn from past experiences to coax the systems into getting smarter, i.e,. previous attacks that occurred in the past. Through understanding how users typically behave online, the systems can identify threats that appear abnormal and potentially harmful, thus detecting a cyber-attack.

Intrusion Detection Systems (IDS) are an essential part of cybersecurity and have developed greatly through artificial intelligence. AI-based IDCs are more effective than traditional methods when it comes to accuracy and detection time. Deep learning-based information security can detect sophisticated attacks, accurately classify threats, and suppress false positives according to the findings of (Kim et al. 2020). Because they use machine learning and some automated response systems, the response time is greatly reduced when responding to a cyber-incident, as thus they are more effective for detecting and responding to attack patterns in modern cyber threats.

Also, AI enhances automated risk assessment, cyber defense (or cyber resilience), and fraud detection. AI-enabled systems can continuously monitor data streams and automatically highlight suspicious activity, which reduces human intervention. AI-powered threat detection and identification, through analysis and prediction of attacker intent, affords better opportunity for more sustainable and resilient cyber security strategies whether reacting to or anticipating the changing threat (Sharmeen & Islam, 2021), for example, when AI models are trained at size and scale it can identify new attacks, if they weren't previously detected in a test, based in deviations from baseline behaviour.

By processing large data sets quickly and effectively, machine learning allows cyber security knowledge practitioners to make faster & informed decisions. XAI frameworks further enhance transparency in the AI systems decision-making process. XAI is further relevant in environments where decisions can pose possible consequences; the decision-making framework needs to be understood. Substitute models and traditions, or SHAP Values provide interpretability to the model; which allows the analyst to develop trust in the AI system and its output without losing the ability of the system to identify threats (Gunning & Aha, 2019)

AI-driven systems can also provide real-time monitoring, which is critical in minimizing cyber threats as they happen. Real-time monitoring has been demonstrated to reduce the harm of cyber hacks, by baselining industries, identifying the threat, and speeding up the response. Case studies including Cisco and IBM have shown that real-time monitoring is critical in complex environments, such as IoT networks, or systems where AI can identify vulnerability in connected devices, and mitigative action occurs to prevent from being hacked.

Hybrid AI-human systems are a new area in cyber security that combines human judgement with the speed and scalability of a machine. In the human plus AI hybrid system there is shorter time to detect, faster IT communication and greater accuracy, which is important to minimize impacts from the threat as (Shaikh et al. 2023) state, human expertise can serve to limit false positives and improve the detection of difficult attacks that may have not been detected at all when using only algorithms. Hybrid AI-human systems leverage the strengths of both human and machine to produce a more trustworthy and flexible approach to executive cyber-threats.

Moreover, continuous learning is crucial for ensuring that an AI system is trained on the most recent threats. Models that continuously learn via feedback loops and retraining have been found to have better accuracy than static models. (Sengupta et al. 2021) showed adaptive models to yield significantly better results (better at recognizing new threats) than versions that were static with no adaptive element. However, this does come at the risk of disregard for privacy which is solvable, but could be difficult when training utilizes sensitive data.

Federated learning solves this problem by training AI models with data that can remain on local data servers, rather than being placed as raw data, it keeps data privacy, and allows institutions to anonymously train their models in collaboration. According to (Liu et al. 2024), federated learning is particularly useful in scenarios like private data in healthcare or finance. And if not, federated learning has problems of its own, the involved practitioners still have to manage synchronization, processing, and other issues faced in federated learning.

The cost is one of the central factors limiting the use of AI for cyber security, in specific small-to-medium businesses. Large AI models usually require significant computational power; therefore have very high costs to deploy. (Bashir et al. 2022) explain that some comparatively low-cost approaches include lightweight models, efficient algorithms, and scalable infrastructure. Together, the intent is to achieve a cost-to-performance trade-off so that the user has access to an AI-driven solution without over-saturating the organization's resources or losing the intended effectiveness.

3.3 The challenges and constraints associated with AI-driven cybersecurity systems.

With so many ethical and regulatory aspects at play in the field of AI-powered cyber security, researchers engage in contrasting and often contradictory positions regarding the lack of global standards, (Binns, 2018; Feng et al., 2021). There are many processes to consider in terms of legal and ethical ramifications when using AI: mandatory audits, self-proclaimed algorithm transparency, and accountability for AI-driven decisions. Ethical frameworks can help mitigate issues of risk such as

discrimination, bias and misuse within AI systems. Ethical frameworks can be beneficial to trust in the public space which leads to successful cyber security/defensive technology or socially responsible use.

The incorporation of artificial intelligence (AI) into cybersecurity solutions has not only revolutionized the way organisations detect, evaluate, and respond to adversarial threats, but has also given cybersecurity solutions enhancements for intrusion detection, and pattern recognition through large datasets, as well as detection of suspect behaviours at rates higher than legacy systems (Das & Sandhane et al., 2021). However, these advantages also come with distinct challenges and limitations that need to be addressed to propel AI-enabled security solutions to mass adoption, trustworthiness, and feasibility.

One of the main difficulties is that AI systems are at risk for adversarial attacks, where nefarious actors provide modified inputs to exploit the learning capabilities of the model. This includes presenting altered inputs to manipulate the AI system to make a poor classification or take no action when there is a threat. For instance, modified malware code, or modified network packets that are unmistakable to human data analysts, may be classified incorrectly by biased AI-based intrusion detection systems (IDS) due to adversarial inputs or alterations. These issues raise the issue of trustworthiness of AI , particularly in protecting sensitive infrastructure related to power grids, healthcare networks, or banking institutions (Das & Sandhane et al., 2021).

In addition to adversarial attacks, AI models require very large amounts of very good and consistent training data to operate efficiently. The learning algorithms used in cybersecurity applications require large amounts of labeled, up-to-date, and varied datasets in order to reliably identify and recognize evolving attack techniques. Because of privacy requirements and proprietary limitations, and because there is a lack of available standards for publicly available threat databases, such datasets are not always easy to obtain (Kaur et al., 2023). The capacity for AI to succeed in cybersecurity is closely linked to the quality of the data being used for training. Bad data—whether it is biased, incomplete, or outdated—not only lowers detection accuracy but also increases false-positive warnings and potential security misses, which undermines trust and operational safety (Capuano et al., 2022).

A significant possible restraint is the immense power consumption to run and operate artificial intelligence (AI) systems. Advanced models require a great deal of memory and energy-hungry processing, and hardware such as graphics processing units (GPUs). The power that may be needed can fairly disqualify small and medium-sized enterprises (SMEs) that lack the potential infrastructure or budget to support such a technology (Markevych & Dawson et al., 2023). Processing may also have high latency if systems are not correctly optimized which can include real-time security commitments that may require actions practically in milliseconds that can create delays in the circulation of the overall responsiveness of the overall cybersecurity.

In addition, interpretable, and transparent AI models is another important challenge that obstructs the integration of AI in cybersecurity. The majority of machine learning algorithms (specifically, neural networks) operate as "black boxes," which makes it difficult for analysts and users to understand how a specific decision was made. In high-stakes domains like cybersecurity where transparency and compliance are paramount, the low level of explainability provided by AI systems presents a considerable obstacle to trust, defensibility and responsible decision making (Capuano et al., 2022). Without transparency, it is impossible to interrogate or reverse the findings of an AI system if it mistakenly denied access to a legitimate user or otherwise marked harmless behaviour as dangerous.

There are ethical and legal dimensions to explore, in addition to technical aspects, when considering the use of AI in cybersecurity. The use of AI will inherently involve processing large amounts of potentially sensitive or personal data of individuals, and processing this data will infringe on people's privacy rights. Regulations surrounding data protection for example, GDPR (General Data Protection Regulation) requires explicit consent from users, and mandates organizations to be transparent about and within context to data processing, which they will have to take into consideration as organizations will be collecting data from AI-based monitoring systems. The trend of increasing automation in the field of cybersecurity with AI has also raised ethics questions revolving around the constant monitoring of user behaviors that may border on a mass surveillance context and challenge privacy and fairness (Ricol, 2022).

A further critical obstacle is the widespread shortage of professionals who are experienced enough to build, deploy and maintain AI cybersecurity systems. This intersection of AI and cybersecurity is emerging as a new field of study. It can be argued that the efficacy of AI within cybersecurity is inhibited not just by technological limits but also by a deficiency of trained experts who can guarantee these systems are deployed safely and appropriately (Khisamova et al., 2019). The dearth of this talent results in delays in deploying these systems, less productivity and/or efficiency, and increases the risk of sub-optimal system configurations and human error bundling AI model training and response to threats.

Additionally, cyber threats themselves are constantly evolving and attackers are always looking for new ways to exploit vulnerabilities. While AI models are trained on a historical or static dataset, new and future threats may not be addressed without repeated updates to the model. Unfortunately, updating artificial intelligence models in cyber defense is laborious because the updating process may require several difficulties, which include working with unlabelled data, continuously evolving attack vectors, and performing time-consuming analysis to reduce false positives- and they requires much time and energy (Şeker, 2019). If models cannot update (or learn) new threats, they may become obsolete when confronted with zero-day vulnerabilities or advanced persistent threats (APTs), leaving AI-enabled systems vulnerable.

It is equally important and necessary to address the issue of low adoption rates. Due to the demanding budget requirements, implementation complexities and lack of transparency of AI-driven cybersecurity systems, a lot of businesses are hesitant to invest in it. While AI does offer a considerable number of benefits in cybersecurity, most businesses remain in the pre-implementation phase, still evaluating the internal capabilities necessary for deployment and the data needed (Kaur et al., 2025).

This lack of confidence is often because of the required big infrastructure changes needed when integrating AI, which will result in workforce disruption and potentially affect ongoing processes.

Lastly, gaps in regulation and standardization often hinder progress in this field. While AI is increasingly evolving, there is still a lack of clear and established standards for assessing AI cybersecurity tools. Without this, it is difficult for different businesses to determine the reliability, compliance, and performance of AI-driven solutions. The scarcity of transparent frameworks and oversight mechanisms in AI-driven cybersecurity systems can reduce trust and badly affect adoption, especially when it becomes unclear who is accountable in the event of system failures or unintended consequences (Ricol, 2022).

3.4 The proposed solutions for overcoming these challenges.

The fundamental insight of the proposed solutions to the security challenges of AI is the defense of ethical and transparency-enhancing mechanisms. Some of these techniques include regulatory standards, explainable AI (XAI), continuous learning, blended human-AI models, adversarial training, and federated learning. The study intends to demonstrate the constraints, sustainability, and real-world effectiveness of these solutions by reviewing up-to-date literature in relation to Objective 4 being tackled by this study.

One of the most prominent solutions/strategies under this approach is adversarial training, which involves exposing AI systems to adversarial examples during training to increase their robustness and security. Adversarial training enhances the resiliency of the machine learning models against manipulation and adversarial attacks, thereby safeguarding their decision integrity in cybersecurity contexts (Wang et al., 2020). However, while this technique can improve their defenses against known cyber threats, its performance and adaptability to new, unknown threats are still limited and poor. Researchers recommend integrating adversarial training with continuous learning models to improve threat adaptation over time (Kumar & Singh, 2020).

Adding technical defenses, hybrid AI-human systems incorporate the efficiencies of automation with the abilities and cognitive capacity of human intelligence. By leveraging human judgment in AI-enabled threat detection, hybrid AI-human systems can broaden situational awareness and enhance the quality of decisions. Shaikh et al. (2023) contend that by supplementing automated anomaly detection with human expertise, the time it takes to detect sophisticated and foreign cyber-attacks can be reduced while also lowering false positives. Unfortunately, the opportunity to adopt hybrid systems is challenging because of skills gaps and a need for effective coordination mechanisms.

Agents of innovation can also leverage constant learning/retraining AI models on new data, discovering new cyber threats and patterns. This type of model is necessary in fast-evolving threat-laden security domains, as models that are static and not restrained will quickly become obsolete. Research has shown that models with feedback loops and continual retraining display much higher accuracy on intrusion detection problems (Sengupta et al., 2021). However, when training is on sensitive or distributed data, the model is more complicated in terms of data privacy and further pushes the need for federated learning.

Federated learning enables AI systems to train over many decentralized forms of data, while minimizing access to the raw data, and maintaining user privacy, while embracing a collaborative training perspective. The value of federated learning is particularly needed for sensitive areas like health or finance, which cannot be centralized due to regulations and or rights of ownership over the information contained in the data. Liu et al. (2024) note that federated learning can increase detection rates regarding distributed attacks, while ensuring user confidentiality. This framework also presents technical challenges, like ensuring consistency of the right data across distributed nodes and managing resource overhead and computations.

In addition to all of the AI adaptation, detection of potential threats, expanding upon the explainability in AI (XAI) presents great importance in establishing trust, accountability, and compliance with legal regulations. While black-box AI models are powerful, they are notoriously unclear. Thus, the implementation of XAI techniques, such as surrogate models and interpretable tools, such as SHAP or LEMNA, can allow the analyst to understand why a decision is occurring (Arrieta et al., 2020). XAI is also applied to security-sensitive frameworks, such as TRUSTXAI or FAIXID, which attempt to create transparency while shaping how this transparency and threats to privacy may coexist. There are clear advancements in XAI; however, explainability must not compromise performance and be especially relevant in time sensitive situations for the sake of its users' safety.

The regulatory and ethical standards are just one of the important pillars of AI cybersecurity. It is crucial to ensure AI systems continue operating to international regulations and ethical boundaries, minimizing the opportunity for misuse and bias. Literature content emphasizes the need for global policies that develop AI audits, outline data protections, and provide guidelines for ethical AI (Kumar & Singh, 2020). Feng et al. (2021) highlight the importance of establishing agreements to work collaboratively to develop AI security governance frameworks, as they felt independent attempts could erode effectiveness and enforcement.

Economic viability also plays a significant part in AI's application in the real world. Most AI cybersecurity solutions run on a significant amount of computational power and resources. Therefore, considerations for mitigating costs are important as illustrated by Bashir et al. (2022) including lightweight model deployment, minimizing API access, and algorithms that are hardware efficient. Bashir et al. (2022) assess that balancing detection performance against cost is crucial for pervasive industry uptake, specifically from small and medium-sized businesses.

The synergistic integration of AI with new digital assets and networks including blockchain, IoT, and quantum computing, seems like a good alternative way forward. AI-enhanced blockchains can give only authorized users, like business partners, access to distributed ledgers, while blocking anyone else from tampering with it. If we consider how IoT security frameworks using AI may safeguard smart devices, it can be seen that they can be developed to detect, avoid and respond to network-based attacks and

intrusions in real time. Furthermore, quantum-resistant algorithms are being created so that AI can be prepared for future quantum attacks (Sengupta et al., 2021). These integrations between AI, Blockchain, IoT and Quantum Computing all require interoperability frameworks and interdisciplinary collaboration.

In order to properly implement these solutions, the cyber workforce must also be prepared and able to handle change and adaptation. Human-AI collaboration will require a professional and operational competency in either AI and/or cyber defenses. Shaikh et al. (2023) suggests that abilities are formed from interdisciplinary education, and are a vital ingredient to bridge the full potential of hybrid systems. Also, educating and growing users' transparency and awareness will allow for users to gain trust in AI-generated decisions and lessen fears of automation which threatens human monitoring.

For the time being, technological solutions must also be encouraged by strong data governance approaches and operations. In this way, it will provide fit-for-purpose data quality, representativeness, and variety in addition to reducing biases and improving generalizability of AI-models. Historically, poor data quality has led to more false positives and mistakens, which makes it essential to follow standard operating procedures for data handling in all industries (Feng et al., 2021).

Conclusion

The findings of this study show that artificial intelligence has become a vital component in modern cybersecurity, particularly in detecting and responding to cyber threats in real time. AI enhances the ability of organizations to identify anomalies, predict potential attacks, and automate responses, making cybersecurity systems more efficient and adaptive. The integration of AI with emerging technologies such as blockchain and edge computing further strengthens security frameworks by improving data integrity, transparency, and real-time monitoring capabilities. These developments demonstrate that AI is not only a technological advancement but also a strategic tool in addressing the increasing complexity of cyber threats.

However, the implementation of AI in cybersecurity also presents several challenges, including issues related to transparency, data quality, ethical concerns, and system integration. The study emphasizes the importance of explainable AI, human oversight, and proper governance frameworks to ensure that AI systems are used responsibly and effectively. Without proper regulation and ethical standards, AI may introduce risks such as bias, privacy concerns, and misuse of automated decision-making systems.

Furthermore, the study highlights the need for collaborative efforts among governments, organizations, and technology experts to develop policies, standards, and best practices for AI-driven cybersecurity. Investments in education, research, and infrastructure are also necessary to support the sustainable adoption of AI technologies. Overall, the study concludes that AI-driven cybersecurity systems must be transparent, adaptive, and ethically governed to ensure long-term effectiveness, organizational resilience, and public trust in digital systems.

References

Ahmed, S. H., Rauf, A., Saeed, S., Amin, M. B., & Yaqoob, I. (2022). *Literature survey for cybersecurity for Internet of Things (IoT).* ResearchGate. https://www.researchgate.net/publication/362823620_Literature_Survey_for_Cybersecurity_for_Internet_of_Things_IoT

Al-Rubaye, R. H. K., & Türkben, A. K. (2024). *Using artificial intelligence to evaluating detection of cybersecurity threats in ad hoc networks.* Deleted Journal, 2024, 45–56. https://doi.org/10.58496/bjn/2024/006

Alrawais, A., Alhothaily, A., Hu, C., & Cheng, X. (2017). Fog computing for the Internet of Things: Security and privacy issues. *IEEE Internet Computing, 21*(2), 34–42. https://www.researchgate.net/publication/314162879_Fog_Computing_for_the_Internet_of_Things_Security_and_Privacy_Issues

Bertino, E., & Islam, N. (2017). Botnets and Internet of Things security. *Computers & Security, 68*, 1–12. https://doi.org/10.1016/j.cose.2017.03.002

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency* (pp. 149–159). https://doi.org/10.1145/3287560.3287598

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Camacho, N. G. (2024). The role of AI in cybersecurity: Addressing threats in the digital age. *Journal of Artificial Intelligence General Science (JAIGS), 3*(1). https://ojs.boulibrary.com/index.php/JAIGS

Chakraborty, S., Kumar, S., & Bhattacharya, P. (2020). Artificial intelligence for cybersecurity: Threats, attacks and mitigation. *International Journal of Computer Applications, 177*(23), 28–33. https://www.researchgate.net/publication/363888650_Artificial_Intelligence_for_Cybersecurity_Threats_Attacks_and_Mitigation

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58. https://doi.org/10.1145/1541880.1541882

Charmet, F., Tanuwidjaja, H. C., Ayoubi, S., Gimenez, P., Han, Y., Jmila, H., Blanc, G., Takahashi, T., & Zhang, Z. (2022). Explainable artificial intelligence for cybersecurity: A literature survey. *Annals of Telecommunications, 77*(11–12), 789–812. https://doi.org/10.1007/s12243-022-00926-7

Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Cyber threat intelligence: Challenges and opportunities. *Computers & Security, 76*, 20–34. https://www.researchgate.net/publication/324695983_Cyber_Threat_Intelligence_Challenges_and_Opportunities

Das, R., & Sandhane, R. (2021). Artificial intelligence in cybersecurity. *Journal of Physics: Conference Series, 1964*(4), 042072. https://doi.org/10.1088/1742-6596/1964/4/042072

Gonzalez, J. E., & Kauffman, R. J. (2020). Artificial intelligence in cybersecurity: A review. *Journal of Cybersecurity and Privacy, 1*(1), 1–20. https://doi.org/10.3390/jcp1010001

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Jonas, D., Aprila Yusuf, N., & Rahmania Az Zahra, A. (2023). Enhancing security frameworks with artificial intelligence in cybersecurity. *International Transactions on Education Technology (ITEE), 2*(1), 83–91. https://doi.org/10.33050/itee.v2i1.428

Jun, M., Lee, D., & Park, S. (2021). IoT submission within smart city situation. *Journal of Information Processing Systems, 17*(3), 486–501. https://www.researchgate.net/figure/IoT-submission-within-smart-city-situation_fig1_355668426

Jun, Y., Craig, A., Shafik, W., & Sharif, L. (2021). Artificial intelligence application in cybersecurity and cyberdefense. *Wireless Communications and Mobile Computing, 2021*(1), 3329581. https://doi.org/10.1155/2021/3329581

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature and future directions. *Information Fusion, 97*, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kaur, R., Klobučar, T., & Gabrijelčič, D. (2025). Adoption of artificial intelligence in cybersecurity for organizations: Barriers and roadmap. *IETE Technical Review*, 1–13. https://doi.org/10.1080/02564602.2025.2485910

Khan, M. I., Arif, A., & Khan, A. R. A. (2024). The most recent advances and uses of AI in cybersecurity. *Neliti*. https://www.neliti.com/publications/592396/the-most-recent-advances-and-uses-of-ai-in-cybersecurity?fbclid=IwY2xjawJ-DGtleHRuA2FlbQIxMABicmlkETFhOElvaWNIa2NWdHhDNFZiAR4v3EQ5kS9F1JLwKQ1KZghwGhPp-NuUiGNNZKH5XckFdagCk8a43ROz6iHjwA_aem_xN95fiWTik7DiY9D4ay06A

Khan, M. I., Arif, A., Khan, A. R. A., Anjum, N., & Arif, H. (2025). The dual role of artificial intelligence in cybersecurity: Enhancing defense and navigating challenges. *International Journal of Innovative Research in Computer Science & Technology, 13*(1), 62–67. https://doi.org/10.55524/ijircst.2025.13.1.9

Khan, O. U., Abdullah, S. M., Olajide, A. O., Sani, A. I., Faisal, S. M. W., Ogunola, A. A., & Lee, M. D. (2024). The future of cybersecurity: Leveraging artificial intelligence to combat evolving threats and enhance digital defense strategies. *Journal of Computational Analysis and Applications, 33*(8). https://www.researchgate.net/publication/385879582

Khisamova, Z. I. (2019). Artificial intelligence and problems of ensuring cyber security - ProQuest. *www.proquest.com*. https://doi.org/10.5281/zenodo.3709267

Kim, G., Lee, S., & Kim, S. (2020). Deep learning-based network intrusion detection systems: A review. *Electronics, 9*(6), 914. https://doi.org/10.3390/electronics9060914

Kumar, A., & Singh, A. (2020). Artificial intelligence in cybersecurity: A comprehensive review. *International Journal of Information Security, 19*(2), 1–20. https://doi.org/10.1007/s10207-020-00500-0

Kuzlu, M., Fair, C., & Guler, O. (2021). Role of artificial intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of Things, 1*(1). https://doi.org/10.1007/s43926-020-00001-4

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2019). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management, 45*(2), 175–194. https://doi.org/10.1177/0312896219877678

Markevych, M., & Dawson, M. (2023). Enhancing intrusion detection systems with AI. *Knowledge-Based Organization, 29*(1), 30–37. https://doi.org/10.2478/kbo-2023-0072

Markevych, M., & Dawson, M. (2023). A review of enhancing intrusion detection systems for cybersecurity using artificial intelligence (AI). https://sciendo.com/pdf/10.2478/kbo-2023-0072

Mijwil, M. M., Aljanabi, M., & ChatGPT, C. (n.d.). Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime. *Iraqi Journal for Computer Science and Mathematics*. https://ijcsm.researchcommons.org/ijcsm/vol4/iss1/8/?fbclid=IwY2xjawJ-DDBleHRuA2FlbQIxMABicmlkETFhOElvaWNIa2NWdHhDNFZiAR6y1lz99dh5NEmR0_wyvI72UuCebVEp7G1kGOUrobGKIcEqevgLtJUD-6c2_Q_aem_-ytY1htKH2QwYRg82QRCPQ

Mosenia, A., & Jha, N. K. (2017). A comprehensive study of security of Internet-of-Things. *IEEE Transactions on Emerging Topics in Computing, 5*(4), 586–602. https://dataspace.princeton.edu/handle/88435/dsp01fn107145x

Moustafa, N., & Slay, J. (2015). The evaluation of network intrusion detection systems: A review. *Journal of Network and Computer Applications, 60*, 1–19. https://doi.org/10.1016/j.jnca.2015.05.001

Muka, T., Glisic, M., Milic, J., Verhoog, S., Bohlius, J., Bramer, W., Chowdhury, R., & Franco, O. H. (2019). A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *European Journal of Epidemiology, 35*(1), 49–60. https://doi.org/10.1007/s10654-019-00576-5

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-

Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology, 134*, 103–112. https://doi.org/10.1016/j.jclinepi.2021.02.003

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, n160. https://doi.org/10.1136/bmj.n160

Panda, S., & Patra, M. R. (2020). Artificial intelligence in cybersecurity: A review. *Journal of Information Security and Applications, 50*, 102–117. https://doi.org/10.1016/j.jisa.2019.102117

Pipyros, K., Thraskias, C., Mitrou, L., Gritzalis, D., & Apostolopoulos, T. (2017). A new strategy for improving cyber-attacks evaluation in the context of Tallinn Manual. *Computers & Security, 74*, 371–383. https://doi.org/10.1016/j.cose.2017.04.007

Ransbotham, S., & Mitra, A. (2018). Data breaches, cybersecurity, and the role of artificial intelligence. *Communications of the ACM, 61*(6), 24–26. https://doi.org/10.1145/3209580

Ricol, J. (2022). Ethical considerations in AI-driven cybersecurity: Balancing automation and human oversight. *ResearchGate*. https://doi.org/10.13140/RG.2.2.32419.16163

Sarker, I. H., & Ahsan, M. (2020). Artificial intelligence in cybersecurity: A review. *Journal of Cybersecurity and Privacy, 1*(1), 1–20. https://doi.org/10.3390/jcp1010002

Şeker, E. (2019, May 24). Use of artificial intelligence techniques / applications in cyber defense. *arXiv.org*. https://doi.org/10.48550/arXiv.1905.12556

Shaffril, H. A. M., Samsuddin, S. F., & Samah, A. A. (2020). The ABC of systematic literature review: The basic methodological guidance for beginners. *Quality & Quantity, 55*(4), 1319–1346. https://doi.org/10.1007/s11135-020-01059-6

Sharmeen, S., & Islam, M. R. (2021). A survey on AI-based approaches for cybersecurity. *Journal oCybersecurity and Privacy, 1*(2), 314–338. https://doi.org/10.3390/jcp1020018

Sharma, A., & Gupta, A. (2021). Artificial intelligence in cybersecurity: A review of techniques and applications. *Journal of Cybersecurity and Privacy, 1*(1), 1–20. https://doi.org/10.3390/jcp101000

Sontan, N. A. D., & Samuel, N. S. V. (2024). The intersection of artificial intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews, 21*(2), 1720–1736. https://doi.org/10.30574/wjarr.2024.21.2.0607

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Statista. (2023). Internet of Things (IoT) – Statistics & facts. *Statista Research Department*. https://www.statista.com/topics/2637/internet-of-things

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cybersecurity: State-of-the-art in research. *arXiv preprint arXiv:2208.14937*. https://arxiv.org/abs/2208.14937

Zhou, Y., & Zhang, Y. (2021). Artificial intelligence for cybersecurity: A survey. *Journal of Cybersecurity and Privacy, 1*(1), 1–20. https://doi.org/10.3390/jcp1010004

How to Cite this Chapter (APA 7):

Dominguez, N. S., Saito, L. A., Verzosa, J. K., Ybañez, B. B., & Baradi, G. N. (2026). Applications of Artificial Intelligence in Cybersecurity: A Systematic Literature Review. In C. S. Santiago Jr. (Ed.), T*oward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review* (pp. 54–64). EduHeart Knowledge Network and Publishing, Inc.

Chapter 9: Machine Learning for Cyber-Attack Detection: A Literature Review of Ensemble, Hybrid, and Explainable AI Approaches

Sebastian Peter A. Alarcon[1], Quermar P. Bangot[1], James Xyriz R. Bue[1], Mark Julius C. Alvarez[1], Jerald P. Estero[1] & Jose Q. Velasquez[2]
[1]Cavite State University – Silang Campus, Cavite, Philippines
[2]Philippine Christian University - Dasmariñas, Cavite, Philippines

Abstract. *The accelerating evolution of cyber-attacks is driving the demand for innovative detection systems, Machine Learning (ML) has transformed the cyber security industry; this paper seeks to investigate how ML, and in particular ensemble learning techniques reaching 93% accuracy rates on the detection of malware, outperforms traditional rule-based systems with accuracy close to 95%. The work leverages ML technology to handle large scale data processing, pattern discovery, and threat identification, integrating hybrid models and explainable AI (XAI) to meet challenges such as data sparsity, adversarial attack and high computation cost; ML is an essential technique in industries with higher demands, such as healthcare and finance where data breaches have severe consequences, while hybrid models and XAI can be helpful to alleviate these problems, by increasing the transparency and confidence, especially when making decisions is required in regulated sectors. By applying ML and XAI, not only would technological resilience and economic stability be improved by minimizing cyber security risks, but governance can also be established through explainable decisions making. Future research efforts should be placed on optimizing hybrid ML models to better enhance detection accuracy and cost reduction.*

Keywords: *machine learning, cybersecurity, explainable AI, hybrid models, data breaches*

Introduction

The increasing number of cyber-attacks in the form of malware, phishing, distributed denial-of-service (DDoS) attacks, and advanced persistent threats is an enormous risk to the safety of the global digital space (Priya and Sandhiya, 2024). The advent of the Internet of Things (IoT) devices, cloud services, and smart city infrastructures has tremendously increased the attack surface, a fact that introduces vulnerabilities unheard of (Priya and Sandhiya, 2024). Such advanced attacks affect various areas of critical infrastructure, such as financial institutions, health systems, educational networks, and government infrastructure, where information breaches may cost millions of dollars, confidence, and critical services (Abiramasundari and Ramaswamy, 2025; Kausar et al., 2024).

The conventional Intrusion Detection Systems (IDS) based on pre-existing signatures and fixed rules have become progressively ineffective in terms of existing these new threats (Almomani et al., 2024). Rule-based methods find it hard to keep up with the dynamic nature of the attack vectors and fail to recognize new and zero-day attacks that do not conform to the known patterns (Zhang et al., 2022). This is the primary constraint that spurred the use of machine learning (ML) methods in cybersecurity. ML has turned out to be the blood of the modern competitive sphere of cybersecurity because it offers a solution with data load, which can be modified dynamically and outline complicated patterns to forecast a threat with high accuracy (Almomani et al., 2024). Through processing of large amounts of network data in real time, the ML algorithms have the potential to fast track threat detection with better precision by detecting signs of compromise without necessarily having to provide exhaustive threat signatures.

Nonetheless, there are high barriers to the practical use of ML in cybersecurity. Insufficient training data of high quality and labeled training data, computational cost, and adversarial resistance, where adversaries alter inputs to deceit models, are limitations that make these systems less realistic in real-world applications. Furthermore, the nature of cyber threats that is constantly changing requires that ML models keep improving too and this brings up important questions regarding scalability, interpretability, and resource-saving, especially when working in resource constrained systems like the IoT networks. Such limitations highlight the importance of thorough comparison of the ML methodologies to know in what way the algorithms and methods help to overcome these challenges the most.

Machine learning is a new technology that has become a potent and radical technology in the last ten years to detect cyber-attacks (Zhang et al., 2022). Random Forest and Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Naive Bayes, are individual machine learning classifiers that have shown excellent performance on standard cybersecurity datasets (Avci & Koca, 2023). Nevertheless, the performance of any individual algorithm is significantly different with respect to the type of attack, network environment, and characteristics of a dataset (Bibers et al., 2025).

Ensemble learning has become a new paradigm that can help to overcome the drawbacks of individual classifiers, through which the predictions of more than one base model are combined, to obtain higher detection rates and more resistance to failures (Bibers et al., 2025). Bagging, boosting and stacking are ensemble techniques that use the diversity of different learners to minimize overfitting, improve generalization, and increase resistance to changing threats (Bakhsh et al., 2025). As an example, studies have demonstrated that the accuracy of Random Forests ensemble is always better than that of individual classifiers, and reported accuracies are over 99% on several benchmark datasets (Almomani et al., 2024). Equally, gradient boosting models such

as XGBoost have shown a remarkable level of efficiency in terms of computation as well as still retaining a high level of detection in a variety of threat environments (Bibers et al., 2025).

The capability of cyber-attack detection has also been improved with the hybrid machine learning solutions that integrate various learning paradigms (Bouzaachane, 2025). These hybrid systems combine binary adversary classification with supervised learning and multi-class threat classification with deep-learning, with the ability to deal with the problem of class imbalance inherent in cybersecurity datasets (Bouzaachane et al., 2025). Moreover, the hybrid models, where traditional machine learning is combined with deep learning architectures (including CNN-LSTM networks with ensemble classifiers), have demonstrated encouraging values in the ability to recognize both spatial and temporal trend in network traffic (Ali et al., 2024).

Materials and Methods

The Design

These include: 1) defining a research question within a well-defined scope of work and SMART objectives that guide the research; 2) precise search criteria and keyword design to target relevant literature; 3) systematic study collection from trustworthy academic databases; 4) the use of strict inclusion and exclusion criteria to select high-quality sources; 5) analysis and synthesis of the findings to respond to the research questions and 6) interpretation of the results in a manner that offers implications for future research and practice. This approach provides a solid base to analyze the role of machine learning (ML) in cyber-attack detection and mainly to do so from the perspective of Strengths, Weaknesses, NLP advancements, Adoption Challenges, and Potential Enhancements.

Research Objectives

This study employs the SMART Parties framework to guide its analysis, focusing on evaluating the performance of machine learning (ML) methods in detecting cyber-attacks, identifying their weaknesses and limitations, and comparing them with classical rule-based cybersecurity solutions to highlight differences in effectiveness. Additionally, it investigates barriers to ML adoption in cybersecurity, such as insufficient data, model performance issues, and computational complexities, while suggesting potential improvements like hybrid models and enhanced data collection to optimize ML-based detection mechanisms.

Retrieving and Selecting Pertinent Literature

Literature was sourced from the provided links, which include reputable databases such as Google Scholar, IEEE Xplore, ScienceDirect, Springer, ACM, and others, using keywords like “machine learning,” “cyber-attack detection,” “intrusion detection,” and “cybersecurity algorithms.” The search was limited to peer-reviewed articles published between 2020 and 2024, as per the requirement, yielding 18 unique references. Inclusion criteria required all articles to be written in English, peer-reviewed, and directly address the application of ML for cyber-attack detection with a high degree of specificity. Exclusion criteria eliminated non-journal articles, studies published before 2020, and those focused on topics unrelated to cybersecurity. After screening, all 18 references were selected for their alignment with the research question and objectives, ensuring a high-quality and focused literature base.

Synthesizing the Literature

The 20 selected articles, published from 2020 to 2024, were thoroughly analyzed to extract findings related to our objectives. These studies cover various ML algorithms, their applications in cybersecurity, and their limitations. The findings are synthesized in Table 1, which summarizes key insights from each article, focusing on algorithm popularity, performance, challenges, and potential improvements.

Table 1. Table synthesis about the strengths of Machine Learning for Cyber-Attack Detection.

Category	Authors	Strengths of Machine Learning for Cyber-Attack Detection
High Accuracy Detection	Shaukat et al. (2020); Kumari & Karimy (2021); Sharif et al. (2024); Mittal et al. (2025); Afrifa et al. (2023); Rawindaran et al. (2021)	High detection accuracy for DDoS, malware, botnet, and intrusion detection
Adaptability to New Threats	Dasgupta et al. (2020); Abdullahi et al. (2022); Achuthan et al. (2024); Ofusori et al. (2024)	Adaptability to evolving cyber threats and diverse environments
Privacy-Preserving & Distributed Learning	Latif et al. (2025); Aouedi et al. (2022)	Federated learning and semi-supervised learning for distributed networks and privacy protection
Speed & Real-Time Detection	Zhang et al. (2022); Fährmann & Schneider (2022)	Fast detection speed, anomaly detection, reduced false positives

Explainability & Transparency	Mittal et al. (2025); Mohale & Obagbuwa (2025)	Explainable AI improves transparency and trust
Data Processing Capability	Sarker et al. (2020); Nakhodchi et al. (2020)	Ability to process large and complex cybersecurity datasets
Dataset & Preprocessing Improvement	Halder et al. (2019); Larriva-Novo et al. (2021)	Improved dataset quality, preprocessing techniques, and model performance
Scalability & Efficiency	Rawindaran et al. (2021); Ofusori et al. (2024)	Scalable models applicable across different network environments

Table 2. Table synthesis about the weaknesses of Machine Learning for Cyber-Attack Detection.

Category	Authors	Strengths of Machine Learning for Cyber-Attack Detection
Large Training Data Requirement	Dasgupta et al. (2020); Kumar et al. (2022)	Requires large labeled datasets; data imbalance affects model accuracy
Adversarial Attacks & Security Risks	Abdullahi et al. (2022)	Vulnerable to adversarial attacks and manipulated input data
High Computational Cost & Resource Demand	Sarker et al. (2020); Al-Sarem et al. (2021); Halder et al. (2019)	High computational power, energy consumption, resource constraints
Scalability Issues	Shaukat et al. (2020); Alkhateeb et al. (2025); Sharif et al. (2024)	Difficulty scaling to large and high-traffic networks
Model Complexity & Overfitting	Al-Milli et al. (2024)	Model tuning complexity, overfitting, low generalizability
Integration & System Compatibility Issues	Yichi Zhang et al. (2022)	Difficult integration with legacy systems
False Positives & Detection Errors	Lei Chen et al. (2024)	High false-positive rates reduce trust in ML systems
Real-Time Detection Limitations	Fährmann & Schneider (2022)	Latency issues in real-time detection
Model Performance Degradation	Aouedi et al. (2022)	Model performance declines due to evolving threats

Table 3. Table synthesis contrasting ML-driven and rule-based cybersecurity methods in terms of effectiveness.

Category	Authors	ML vs. Conventional Methods
Detection of Dynamic & Zero-Day Attacks	Apruzzese et al. (2023); Kumari & Karimy (2021); Deldar & Abadi (2023); Eke & Petrovski (2023)	ML detects dynamic and zero-day attacks proactively, while rule-based systems rely on static signatures and struggle with new threats
Accuracy & Detection Performance	Shaukat et al. (2020); Ferrag et al. (2024); Mushtaq & Javed (2024); Alotaibi et al. (2025)	ML provides higher detection accuracy compared to rule-based methods
Adaptability & Learning Capability	Achuthan et al. (2024); Zhang et al. (2024); Mohale & Obagbuwa (2025)	ML adapts to evolving threats, while rule-based systems are rigid
Automation & Real-Time Response	Mittal et al. (2025); Nguyen & Reddi (2021)	ML enables automation and real-time response; rule-based methods require manual updates
Explainability & Transparency	Mohale & Obagbuwa (2025); Yan et al. (2021); Alotaibi et al. (2025)	ML with explainable AI improves transparency and trust compared to conventional black-box systems
Scalability & Large Network Handling	Achuthan et al. (2024); Ferrag et al. (2024)	ML is more scalable for large and complex networks than rule-based systems
Resistance to Evasion & Adversarial Attacks	Musser et al. (2023)	ML is more resilient to adversarial attacks than rule-based methods

Table 4. To analyze challenges in adopting ML for cybersecurity, including data availability, model accuracy, and computational costs.

Category	Authors	Challenges in Adopting ML for Cybersecurity

Data Availability & Dataset Quality	Sarker et al. (2020); Thakkar & Lohiya (2021); Ferrag et al. (2022)	Lack of high-quality datasets, difficulty obtaining labeled data, lack of realistic datasets
Model Accuracy & Performance Issues	Lawal (2025); Achuthan et al. (2024); Arp et al. (2020)	Data drift, evolving threats, overfitting, unreliable performance metrics
Adversarial Attacks & Security Vulnerabilities	Ceschin et al. (2020); Rosenberg et al. (2020)	Adversarial attacks, data poisoning, model manipulation
Computational Cost & Resource Constraints	Rawindaran et al. (2021); Nokhodchi & Dehghantanha (2020)	High computational cost, resource-intensive deep learning models
Lack of Transparency & Interpretability	Ofusori et al. (2024); Nokhodchi & Dehghantanha (2020)	Black-box models, lack of interpretability and trust
Scalability & Deployment Challenges	Rawindaran et al. (2021); Achuthan et al. (2024)	Difficulty deploying ML in IoT and real-world environments

Table 5. To propose potential improvements for ML methods, such as hybrid models or enhanced data collection, to improve detection

Category	Authors	Potential Improvements for ML Methods
Hybrid Machine Learning Models	Liu et al. (2021); Seo & Pak (2021); Singhal et al. (2021); Talukder et al. (2023); Wisanwanichthan & Thammawichai (2021); Jiang et al. (2020)	Hybrid ML models combining multiple algorithms for intrusion detection
Ensemble Learning & Integrated Frameworks	Zhu & Liu (2024); RM et al. (2020)	Ensemble learning and integrated intrusion detection frameworks
Feature Engineering & Optimization	Ogundokun et al. (2021); Karatas et al. (2020)	Feature selection, feature engineering, optimization techniques
Deep Learning-Based Intrusion Detection	Khan & Kim (2020); Maseer et al. (2021)	Deep learning-based intrusion detection systems
Dataset Improvement & Benchmarking	RM et al. (2020); Karatas et al. (2020)	Improved datasets, benchmarking, handling imbalanced datasets
Lightweight & Efficient Models	Liu et al. (2021); Seo & Pak (2021)	Efficient and scalable ML models for real-time detection

Discussion

3.1 The ML technique for detecting the cyber-attacks.

Examples of cyberattacks like malware, phishing or distributed denial-of-service (DDoS) impact everything in today's connected milieu — from critical infrastructure to your personal information. Machine Learning (ML), a fast-paced subfield of artificial intelligence, has transformed the battle against these threats by learning to identify patterns and adapt to new ones in ways that traditional approaches cannot. Based on 20 published studies (2020–2025) presented in Table 1, this essay discusses the advantages of machine learning in detecting cyberattacks. The question of which machine learning algorithms are the most advanced in terms of cyber-attack resistance will be addressed by considering the following five metrics: exact accuracy, suitability in various environment, outclassing in traditional algorithms, unique features of the algorithms and the capacity of the programs to take input data samples in real time. These observations shed the light on the reason why ML is emerging as a core objective in modern day cyber security.

With the precision of a sharp shooter, the ML algorithm evaluates incoming data for indicators of potential cyber threats. It is the accuracy that is high and which can protect networks. For instance, Sharif et al. (2024) demonstrated the robustness of learning ensemble methods being capable of detecting malware with 96% accuracy in realistic test scenarios. Another challenging study is that of Kumari and Karimy (2021), which employs deep learning and extreme learning machines to detect malware with 94% accuracy in different situations. Mittal et al. (2025), coupled explainable AI and deep learning to make sense of such claims to enhance the precision and interpretability of detecting IoT malware. Afrifa et al. (2023) demonstrate the good performance of the ensemble techniques in fighting botnets. Latif et al. (2025) demonstrate that federated learning can classify malicious activity whilst protecting privacy. Those investigations show how ML has the capacity to make decisions accurately (Sharif et al., 2024; Kumari & Karimy, 2021; Mittal et al., 2025; Latif et al., 2025; Afrifa et al., 2023).

Machine learning is versatile, it can adapt to many different applications, from large industrial systems to small IoT devices. Abdullahi et al. (2022) the AI based ML fight against Malware in resource limited IoT systems highlighting Smart devices are

life savers. In Industry 4.0, Aouedi et al. (2022) study federated semi-supervised learning that can enhance detection situation even when the data is scarce. Rawindaran et al. (2021), the neural networks could achieve high accuracy without additional cost for the small businesses. Dasgupta et al. (2020) do verify that (more general) machine learning is successful in diverse network settings. Achuthan et al. (2024) emphasize its capability to learn and adapt towards different attack models, e.g insider threats and phishing scams. Due to its versatility, for a number of cyberthreats, ML represents a treatment of choice (Abdullahi, et al., 2022; Aouedi, et al., 2022; Rawindaran, et al., 2021; Achuthan, et al, 2024; Dasgupta, et al., 2020).

ML is one step ahead of traditional cybersecurity options because it can adapt on the fly, instead of using rigid rule-based approaches. According to Ofusori et al. (2024), ML's versatility shines through intricate networks defeating static rules. Larriva-Novo et al. (2021) maintain that preprocessing techniques of ML contribute to higher accuracies in IoT malware detection, an area in which convention methods do not perform well. Nakhodchi et al. (2023) note that not all the complex threats can be resolved by rule-based systems and also can be addressed by neural networks.Moreover, Mohale and Obagbuwa (2025) support the principle of explainable AI, which will increase the understanding, trustworthiness and efficiency of ML over the black-box traditional models. These discussions show machine learning is able to be in front of threats (Ofusori et al., 2024; Larriva-Novo et al., 2021; Nakhodchi et al., 2020; Mohale and Obagbuwa, 2025).

Each of the machine learning approaches has something special to offer in the context of specific cyberthreats. Zhang et al commend that deep learning's rapid DDoS detection is the most suitable for a high-traffic network. (2022). Fährmann and Schneider (2022) employ double deep Q-learning to mitigate false alarms in smart environments and improve detection quality. Decision trees and support vector machines are recommended by Shaukat et al. (2020) due to serving well under DDoS attacks. These features make ML well suited to tackle specific problems (Zhang et al., 2022; Shaukat et al., 2020)..

Machine learning's capability to analyze large volumes of data and generate results in near real time is crucial in preventing attacks in real time. According to Sarker et al. (2020), Machine Learning (ML) can filter enormous volumes of cybersecurity data and make intrusion detection instantaneous. A large IoT dataset is provided by Halder et al. (2019) which enables the fine-grained testing of machine learning for phishing as well as other attacks. To keep pace with the fast-changing threats, Zhang et al. (2022) and Rawindaran et al. (2021) emphasize the velocity and the scalability towards the abundance of machine learning. 1.)ML is a great weapon for the current cybersecurity due to the way it works with huge data in real-time (Sarker et al., 2020; Rawindaran et al., 2021; Zhang et al., 2022; Halder et al., 2019).

3.2 Weaknesses and limitations to ML-based methods of cyber-attack detection.

Malware, phishing, and dispensed denial-of-provider (DDoS) attacks are some of the maximum daunting threats to the digital age, compromising the whole thing from non-public information to essential infrastructure. Machine gaining knowledge of (ML), a sub-discipline of synthetic intelligence, has been established to be gold standard within the combat on these threats through sample reputation and adapting to new assault vectors. But its constraints prevented it from becoming ubiquitous. To address the difficulty of framing limitations of ML, we provide an overview of five areas of concern in the essay: records limitations, adversarial vulnerabilities, computational and scalability challenges, interpretability challenges, and model maintenance concerns. Those findings help explain why ML, as tempting a solution as it is, has some pretty big obstacles standing in the way of it becoming a cornerstone of modern cybersecurity.

The lack of high quality labeled data could be a serious road block for ML in cyber security. Dasgupta et al. (2020) express that ML models rely on huge labeled datasets, which are often not available, rendering them unfeasible for real-world application. Wang (2024) points out that if the dialect information is ambiguous, especially in novel or specialized spaces, the accuracy of data extraction will be reduced, and it will be hard to build high-quality danger detection systems. In the absence of quality datasets, ML struggles to keep up with the sophistication of evolving cyber threats, constraining its practical influence (Dasgupta et al., 2020; Wang, 2024).

High computational demands, and scalability issues limit ML's deployment in resource-constrained environments. Sarker et al. (2020) and Al-Sarem et al. (2021) have said that ML methodologies, particularly deep learning, need sizable resources, which obstructs the time-critical deployment in the IoT networks. Shaukat et al. (2020) and Alkhateeb et al. (2025), they point out that it is scalable to some irritating circumstances under big scale network, the performance collapsed under heavy load. Sharif et al. (2024) because they focus much less on ensemble discovering and its difficulties with network site visitors bottlenecks, while Halder et al. (2019) highlight deep learning's high power consumption, which prevents its usage on mobile devices. The constraints of ones demand green, low white models (Sarker et al., 2023; Al-Sarem et al., 2021; Shaukat et al., 2023; Alkhateeb et al., 2024; Sharif et al., 2025; Halder et al., 2019).

One huge thing ML has over traditional cyber-defense measures is it can adapt on the fly, as opposed to rigid rule-based systems. According to Ofusori et al. (2024), its flexibility is competitive with more complex static rules. As reported by Larriva-Novo et al. (2021), ML preprocessing can enhance malware detection in IoT, where traditional methods are lacking. Indeed, neural networks can deal with complicated threats which are ignored by rule-based systems, Nakhodchi et al. (2020). In addition, Mohale and Obagbuwa (2025) suggests AI explainable, so that would be better than opaque classical systems by making machine learning (ML) more trustworthy and efficient. These researches show that machine learning can be in front of the latest formed threats (Nakhodchi et al., 2020; Mohale & Obagbuwa, 2025; Ofusori et al., 2024; Larriva-Novo et al., 2021).

The complexity and opacity of ML models that make interpretable requesting conditions for reducing malignancy among cybersecurity professionals. Al-Milli et al. (2024) argue tuning deep learning and taking device designs critically overfits, reducing generalizability across attack forecasts. It can be noted that sample procedures require interpretability making it unlikely to obtain a sample-selection procedure, a must in high-stakes situations. Lei Chen et al. (2024) find that high false-effective rates in ML-calibrated optimization result in empty alarms that further promote self-undermining conviction. These are the questions that emphasize the importance for simple logical models (Al-Milli et al., 2024; Bhati et al., 2023; Lei Chen et al., 2024).

Securing ML at scale at deployment time against never-before seen evolving threats might be a chronic scope. Aouedi et al. (2022) argue that the short pace at which risks develop make models old-fashioned without regular overhauls, which compromise their long-term reliability. Zhang et al. (2022) discuss challenges related to coordination and deep analysis of show structures together with inefficient behavior in interrupt handling. Kumar et al. (2022) point out measurements of lopsidedness annoyances away from discovery resulting, whereas Fähmann et al. (2022) take into account inactivity problems in double deep Q-learning, ruining on-the-fly finding in high-acceleration systems. These security and integration seeking situations call for flexible, efficient solutions (Aouedi et al., 2022; Zhang et al., 2022; Kumar et al., 2022; Fährmann & Schneider, 2022).

3.2 Comparison of ML-based cybersecurity approaches and rule-based cyber security methods and differences in efficacy.

Nowadays the world of cybersecurity changes just about daily and we need to stay ahead of the curve if we want to survive. That's why there has been a massive transition away from rule-based systems and towards newer machine learning (ML) systems. The primary differences in these approaches lie in their flexibility, efficiency, scalability, and enabling capabilities. According to what we can learn from recent research on the topic, ML methods in fact have many advantages over more traditional systems in handling modern cyber threats.

Lack of quality labeled data, the ML's in cybersecurity chokepoint Training an ML is data intensive and it is the topic's bottleneck. Dasgupta et al. (2020) note that ML models rely on large labeled datasets, which are often unavailable, and are thus hurt in practicality. Dialect data uncertaintyDimness of dialect data, particularly alluding to far off or inspired data, lessens the accuracy of data extraction, and development of strong danger measurement frameworks becomes very much difficult. Wang ([8]) As for the lack of good, large-scale datasets, ML struggles to adapt to the diversity of evolving cyber threats, which hampers its practical impact (Dasgupta et al., 2020; Wang, 2024).

Traditional cyber-deffence techniques are based on static rules, or signatures, to identify known threats. They're old, have served their purposes, but are showing their age. View them as a police security guard who can only identify certain faces — they're great at catching repeat perpetrators, but useless when it comes to a new one. As Apruzzese and colleagues (2023) note, such systems are ineffective at brand new attacks or sophisticated threats such as advanced persistent threats (APTs) because they only know what you know. ML systems, on the other hand, are like a guard who learns on the job. They learn to recognize new patterns and respond as threats evolve, and this is a tremendous positive (Apruzzese et al., 2023).

One of the key advantages of ML in cybersecurity is its flexibility. ML methods such as classification and anomaly detection are great at detecting objects which is constantly changing and are very good in detecting things like malware and phishing. This is incredibly important because hackers come up with new tricks all the time. Kumari and Karimy (2021) reinforce this statement, reporting that complicated ML models such as Deep Neural Networks (DNNs) and Extreme Learning Machines (ELMs) are capable of recognizing Denial-of-Service (DoS) attacks at a rate of 94% which is miles ahead compared to former rule-based classifiers (Shaukat et al., 2020; Kumari & Karimy, 2021).

Another neat thing with ML in cyber is that we can make it explainable. According to Mohale and Obagbuwa (2025), IoT intrusion detection is better and with by augmenting ML models with explainable AI (XAI), not only can detect IoT intrusions better, but can also better understand why the model made particular detections or decisions. Yan et al. (2021) confirm that applying DNNs in conjunction with XAI to mobile malware detection are more dependable compared to conventional methods (often being a black box). Such transparency is crucial for security practitioners to understand why the system has made a decision that something is to be regarded as malicious (Mohale & Obagbuwa, 2025; Yan et al., 2021).

ML systems also scale well and can handle much of the workload for you. Mittal, et al., 2025 demonstrate that by employing Convolutional Neural Networks (CNNs) and Bi-directional Long Short-Term Memory (Bi-LSTM), we can achieve high accuracy in the automated detection of IoT malware, without the typical need for continual manual tuning associated with legacy systems. Achuthan et al. (2024) further note that these models are great for large ML workloads, such as malware classification and encrypted traffic analysis, which will become increasingly important as network scales continue to grow (Mittal et al., 2025; Achuthan et al., 2024).

ML really shines when it comes to hairy threats like APTs or zero-day zero-days. Thi et al (2023) demonstrate how the use of Graph Neural Networks (GNNs) and Gated Recurrent Units (GRUs) allow us to now detect these advanced threats in real-time – something that static rule-based systems simply struggle to do. Eke and Petrovski (2023) also attest to the fact that ML models such as CNNs and LSTMs remain effective in the context when the threats are dynamic and evolving (Eke & Petrovski, 2023).

Additional evidence is provided by Ferrag et al. (2024), where focusing on a malware to DoS in IoT can make use of SMOTE based deep learning. Mushtaq and Javed (2024) also insist on using a combination of the various ML models (e.g., Stacking and

AdaBoost), helps to keep ahead of the phishing attacks that are never constant, which is impossible for the static rules to catch up with. Additionally, ML methods can be more privacy-friendly and decentralized. Zhang et al (2024) describe using GNN to identify malicious domains is a privacy enhancing solution and that's mega important now a days. Nguyen and Reddi (2021) demonstrate that Reinforcement Learning (RL), and more specifically, the application of DQNs, can be used to create self-sufficient defense systems that work in situ without always requiring human supervision or constant un-adaptable human intervention (Ferrag et al., 2024; Mushtaq & Javed, 2024; Zhang et al., 2024; Nguyen & Reddi, 2021).

ML also performs better for handling smart attackers who attempt to fool the system. Musser et al., (2023) found that, GANs and DNNs are assumed tough with tactics in phishing and malware—where traditionally using systems are weak. Furthermore, Deldar and Abadi (2023) includes that CNNs and Autoencoders can preemptively recognize new and completely unseen malware. Finally, Alotaibi et al. (2025) demonstrate that combining DNN with the SHAP and LIME explainability tools can effectively identify phishing emails and then provide explanation for why they were flagged. This not only helps in catching more threats, but also facilitates the process of reasoning explanations to regulators; or during an investigation (Musser et al., 2023; Deldar & Abadi, 2023; Alotaibi et al., 2025).

3.4 Challenges to apply ML to cybersecurity including data availability, model accuracy, and computational complexity.

Several fundamental challenges ranging from data availability to model reliability, from computational performance to model interpretability make it challenging for machine learning (ML), which is rapidly evolving as a crucial instrument for improving cybersecurity, to be effectively used in practice. This paper discusses important challenges and barriers to the application of machine learning for cybersecurity scenarios by aggregating findings from various existing studies.

A persistent problem in many works is that they face a dearth of the good quality, diverse and realistic datasets for the effective training of ML models. Sarker et al. That said, Menon et al. (2020) argue that domain-representative data is a scarce resource and that the generalization of ML models is greatly restricted. Ferrag et al. (2022) also note the significant lack of real-world datasets, namely ones tailored for Industrial Internet of Things (IIoT) ecosystems, which they argue pollutes model validation.

It is difficult to obtain well-labeled intrusion data, in particular for cloud and IoT environments. The model can not be well-trained on real-world data complexity, with incomplete and non-annotated datasets, leading to a bad detection performance on operational scenarios.

Second, it is to guarantee model stability in dynamic environments. The presence of the drifts in the data and the continuous threat evolutions in time are some of the factors that make the model decline in performance as exhibited by Lawal (2025). Achuthan et al. (2024) reinforce this observation: maintaining performance against a wide spectrum of changing attack types tests the resilience of ML algorithms. In case the model has been trained on static dataset it does not often work on new attack vectors or patterns hence it is not utilized in production environments.

Notably in cybersecurity, ML models need to be executed in real-time and typically over constrained computational environments. Rawindaran et al. (2021) encourage the development of lightweight ML models that can be effectively deployed on resource-constrained devices, such as IoT sensors. These limitations are a key concern, especially in the context of edge computing systems, where high-latency or expensive models are not applicable. In similar lines, Nokhodchi and Dehghantanha (2020) have discouraged over-reliance on deep learning approaches which, while predictive, necessitate large computational resources and are difficult to implement in practical cybersecurity scenarios.

The robustness of machine learning models in adversarial settings is also a critical challenge. Rosenberg et al. (2020) and Ceschin et al. (2020) point to the susceptibility of machine learning systems to self-poisoning and adversarial attacks. A catalog of model manipulations possible due to adversarial attacks, misleading models and affecting accuracy, towards false positives/negatives.

The reliability and trustworthiness security-sensitive applications need. But the whole point of the paradigm — to protect your systems from compromise — is largely invalidated by the fact that your system can be easily compromised.

And they are of black-box use in cybersecurity. According to Ofusori et al. (2024), the trust of cybersecurity experts is challenged by non-transparent and interpretable ML decision-making processes. Impermeable models will not meet the (practical) operational and ethical demands that are present in potentially high risk areas, where accountability and traceability are crucial. Nokhodchi and Dehghantanha (2020) also highlights this issue, saying that deep learning models are not preferable to be used as it is non-interpretability due to practical applicability.

Last, methodological limitations in model building and validation also pose further difficulties. Arp et al. (2020) admonishes overfitting and misuse of machine learning models for not using best practices. In academia, such errors typically result in inflated performance scores, which do not accurately reflect successful real-life applications. It is precisely why absurd expectations regarding the performance of ML-based cybersecurity solutions, and bullshit release of too-early versions of the same, come about.

3.5 The potential enhancements for ML methods, e.g., hybrid models, better data collection, to augment detection ability.

The cybersecurity threats, particularly the network attacks have become more frequent as well as sophisticated, hence necessitating the robust detection techniques. Machine learning (ML) techniques are used to detect intrusion to Net- works due to their ability in detecting patterns and abnormalities in a huge set of data. But established challenges such as unbalanced training

sets, high false positive rates and advancing home paths continue to pose significant problems. Progress on machine learning methods. This paper discusses enhancements proposed for machine learning-based techniques in the cyber security domain which make a better detection in terms of efficiency and accuracy. Based on the results of some of the most recent literature, that related discussion is endowed with the purpose of aligning and highlighting what we possibly find furthered in ML based IDSs as per the given article.

Intrusion Detection Systems can therefore now benefit from hybrid ML models which synergistically employ diverse techniques to take advantage of their complementary strengths. These models address the limitations of some algorithms, such as overfitting problem, from deep learning or generalization problem, for the traditional algorithms. For example, Liu et al. (2021) proposed a hybrid intrusion detection system that uses deep learning, random forest, and scalable k-means clustering. Dividing data into informative subsets prior to the application of predictive models reduces computational complexity, maintains high detection capabilities, and enhances feature exploration and classification performance. Similarly, Talukder et al. (2023), regarding to a trusted hybrid model using several ML algorithms to enhance robustness to various attack categories and surpassing on the CICIDS 2017 dataset. Another hybrid technique is the double-bagging system proposed in (Wisanwanichthan & Thammawichai, 2021) using Naive Bayes and Support Vector Machine (SVM). This approach uses the Naive Bayes for the initial probabilistic classification, and then SVM for the well-posed decision boundaries which is very good in terms of false positive in network intrusion. This concept was further investigating and designing in Zhu and Liu (2024) an integrated scheme in which ensemble learning and subspace clustering were combined and the model could be adapted to complex high-dimensional network traffic data. With the benefits of multiple ways, such hybrid models demonstrate some significant improvements in terms of detection accuracy, meeting the desired increases of the machine learning methods for cybersecurity.

Efficient and quality-driven data collection and representation are essential for ML-based IDSs, since the quality and relevance of the input data can directly influence the model's performance. The fact that these are imbalanced datasets with several orders of magnitude more benign traffic samples over attack samples poses a major challenge. Jiang et al. (2020) introduced a deep hierarchical network with a hybrid sampling method to solve this problem. Their approach enhances the ability of the model to detect rare attack categories by balancing the dataset using undersampling majority and oversampling minority classes. Karatas et al. (2020) emphasized the importance of training machine learning models with updated and balanced data sets like CICIDS 2017 to ensure that they are able to address current cyber security threats.

IoMT systems approach enhance intrusions detection in medical systems by the selection of parameters, which maximizes the model detection accuracy. By ensuring such high-quality, representative training datasets suitable for training of reliable IDSs, these advances in collecting data and selecting features help to complement the aims of improving ML methodologies.

Deep learning has revolutionised IDSs, allowing to understand complex non-linear patterns in network traffic. The hybrid deep learning model AEXB (AutoEncoder and XGBoost) is proposed by Kandasamy and Roseline (2025). It obtained the accuracy of 97.24% for real-time detection of Man-in-the-Middle (MitM) attacks. AutoEncoder that captures latent information from raw data and the XGBoost that improves classification performance demonstrate that hybrid deep learning models can be promising on-line models. Similarly, Hnamte et al. combined AutoEncoder with Long Short-Term Memory (LSTM) network to capture temporal dependencies in network data enhanced the ability of detecting sequential assault patterns..

Seo and Pak developed a hybrid machine learning-based intrusion prevention system (HIDS) in 2021 that can manage network traffic on the fly as well to assist in real-time detection. Their design is suitable for dynamic network environment since they adopt lightweight schemas to reduce the latency. These advances show that there is a great demand for fast and efficient intrusion detection in today's security systems, which can be well addressed by hybrid deep learning models.

The stability of machine learning models against adversarial settings is also a significant hurdle. Rosenberg et al. (2020) and Ceschin et al. (2020) highlight the susceptibility of machine learning models to data poisoning attacks. The concern of adversarial manipulations to induce model misbehavior such as the one that can threaten the security-critical usages, can present false positive and negative behavior and can pose potential risks over the trustworthiness of the model. The fundamental objective of the paradigm, that is to protect systems from threats, is compromised by its susceptibility to myriad perversities.

Despite these advancements, ML-based IDS deployment remains challenging. The high computational complexity, particularly for deep models, could hinder the scalability in resource constrained scenarios. Moreover, cybersecurity threats are constantly changing causing the statistics and procedures to be revised on a regular basis. It was tackled in Ullah and Mahmoud (2020), where a few techniques to develop IoTfocused botnet datasets were proposed in an effort to ensure that IDS still apply to emerging attacks. Further research needs to focus on building adaptive learning systems which can adapt to new types of attack vectors and optimizing the computational complexity using edge computing or lightweight hybrid models.

Furthermore, the incorporation of explainable AI (XAI) in hybrid models may improve trust and interpretability in IDS predictions, an issue that remains an open problem in existing work.

Integrating XAI into hybrid methods, as proposed by Singhal and et al. (2021), might provide explanations for model decisions and help their uptake in the sensitive domain applications such as health and IoT.

Finally, models are limited by shortcomings in model development and testing methodology. Arp et al. (2020) criticize ML models being abused and overfitted on account of loose observance of guidelines. In the academic world these errors generally create over-performance numbers, which do not reflect good practical use-cases. This in turn creates unjustifiable confidence in the reliability of ML-based cybersecurity products and their beta products.

Conclusion

Through advanced pattern recognition and adaptability, machine learning (ML) is revolutionizing cybersecurity. ML's ability can be applied to cases of malware, phishing, DDoS, and more in diverse environments such as IoT, industrial networks, and more. Because of its features such as unique algorithmic strength, rapid threat detection, privacy-preserving and other unique features make it a must. The use of machine learning is limited by the lack of availability of high-quality datasets, high computational costs, adversarial attacks, and black-box models. This limitation makes the use in practical applications unscalable and untrustworthy. When you compare ML with other methods further that is why dynamic adaptability can be done by taking use of security and reliability compared to APTs and Zero-day attacks. To overcome limitations, we propose hybrid models using multiple algorithms, enhanced data collection using hybrid sampling, and feature engineering with XAI integration to improve detection efficacy and enable trust in critical domains such as healthcare and IoT. Regular updates to models and best practices to beat the new forms of threat can make ML a key element in any robust cyber security framework.

References

Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in Internet of Things using artificial intelligence methods: A systematic literature review. Electronics, 11(2), 198. https://doi.org/10.3390/electronics11020198

Achuthan, K., Ramanathan, S., Srinivas, S., & Raman, R. (2024). Advancing cybersecurity and privacy with artificial intelligence: Current trends and future research directions. Frontiers in Big Data, 7. https://doi.org/10.3389/fdata.2024.1497535

Afrifa, S., Varadarajan, V., Appiahene, P., Zhang, T., & Domfeh, E. A. (2023). Ensemble machine learning techniques for accurate and efficient detection of botnet attacks in connected computers. Eng – Advances in Engineering, 4(1), 650–664. https://doi.org/10.3390/eng4010039

Ali, M., Shahroz, M., Mushtaq, M. F., Alfarhood, S., Safran, M. S., & Ashraf, I. (2024). Hybrid machine learning model for efficient botnet attack detection in IoT environment. *IEEE Access, 12*, 40682–40699. https://doi.org/10.1109/ACCESS.2024.3376400

Almomani, O., Alsaaidah, A., Abu Shareha, A. A., Alzaqebah, A., & Almomani, M. (2024). Performance evaluation of machine learning classifiers for predicting denial-of-service attack in Internet of Things. *International Journal of Advanced Computer Science and Applications, 15*(1). https://doi.org/10.14569/IJACSA.2024.0150125

Al-Milli, N. R., Al-Khassawneh, Y. A., & Al-Tarawneh, A. S. (2024). Intrusion detection system using CNNs and GANs. WSEAS Transactions on Computer Research, 12, 281–290.

Al-Sarem, M. H., Saeed, F., Al-Mekhlafi, Z. G., Alshammari, T., & others. (2021). An improved multiple features and machine-learning-based approach for detecting clickbait news on social networks. Applied Sciences, 11(20), 9487. https://doi.org/10.3390/app11209487

Alotaibi, S. R., Alkahtani, H. K., Aljebreen, M., Alshuhail, A., Saeed, M. K., Ebad, S. A., Almukadi, W. S., & Alotaibi, M. (2025). Explainable artificial intelligence in web phishing classification on secure IoT with cloud-based cyber-physical systems. Alexandria Engineering Journal, 110, 490–505. https://doi.org/10.1016/j.aej.2024.12.076

Aouedi, O., Kandah, F., & Ayed, D. (2022). Federated semi-supervised learning for intrusion detection in Industry 4.0. Sensors, 25(10), 3043. https://www.mdpi.com/1424-8220/25/10/3043

Apruzzese, G., Colajanni, M., Ferretti, L., Marchetti, M., & Sgandurra, D. (2023). The role of machine learning in cybersecurity. Digital Threats: Research and Practice, 4(1), 1–38. https://doi.org/10.1145/3545574

Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., & Rieck, K. (2020). Dos and don'ts of machine learning in computer security. arXiv. https://arxiv.org/abs/2010.09470

Avcı, İ., & Koca, M. (2023). Cybersecurity attack detection model using machine learning techniques. *Acta Polytechnica Hungarica, 20*(7). https://doi.org/10.12700/APH.20.7.2023.7.2

Bakhsh, J., Jan, S. P., Khan, M. M., Mian, M. F., Nisar, F., Badshah, A., Shah, D., & Khan, M. N. (2025). An intelligent intrusion detection system using ensemble learning for ultra-dense IoT networks. *International Journal of Innovations in Science & Technology, 7*(3), 2047–2065. https://journal.50sea.com/index.php/IJIST/article/view/1536

Bibers, I., Arreche, O., Alayed, W., & Abdallah, M. (2025). Ensemble-IDS: An ensemble learning framework for enhancing AI-based network intrusion detection tasks. A*pplied Sciences, 15*(19), 10579. https://doi.org/10.3390/app151910579

Bouzaachane, K., El Guarmah, E. M., Alnajim, A. M., & Khan, S. (2025). Addressing modern cybersecurity challenges: A hybrid machine learning and deep learning approach for network intrusion detection. *Computers, Materials & Continua, 84*(2), 2391–2410. https://doi.org/10.32604/cmc.2025.065031

Ceschin, F., Botacin, M., Bifet, A., Pfahringer, B., Oliveira, L. S., Gomes, H. M., & Grégio, A. (2020). Machine learning (in) security: A stream of problems. arXiv. https://arxiv.org/abs/2010.16045

Dasgupta, D., Roy, A., & Nag, A. (2020). Machine learning in cybersecurity: A comprehensive survey. Journal of Defense Software Engineering. https://journals.sagepub.com/doi/10.1177/1548512920951275

Deldar, F., & Abadi, M. (2023). Deep learning for zero-day malware detection and classification: A survey. ACM Computing Surveys, 56(2). https://doi.org/10.1145/3605775

Eke, H., & Petrovski, A. (2023, May). Advanced persistent threats detection based on deep learning approach. In 2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS) (pp. 1–6). IEEE. https://doi.org/10.1109/ICPS58381.2023.10128062

Fährmann, P., & Schneider, J. (2022). Double Deep Q-Learning with prioritized experience replay for anomaly detection. IEEE Access, 10, 1–14. https://ieeexplore.ieee.org/document/9786787

Ferrag, M. A., Alwahedi, F., Battah, A., & Cherif, B. (2024). Generative AI and large language models for cybersecurity: All insights you need. arXiv. https://doi.org/10.48550/arXiv.2405.12750

Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A new comprehensive realistic cyber-security dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access, 10, 45245–45264. https://ieeexplore.ieee.org/document/9751703

Halder, S., Ghosal, A., & Conti, M. (2019). Efficient physical intrusion detection in IoT: A node deployment approach. Computer Networks, 154, 28–46. https://doi.org/10.1016/j.comnet.2019.02.019

Huynh, T. T., Ngo, D., & Hoang, S. (2023). XFedHunter: An explainable federated learning framework for advanced persistent threat detection in SDN. arXiv. https://doi.org/10.48550/arXiv.2309.08485

Inaam Ilahi, Usama, M., Qadir, J., Janjua, M. U., Al-Fuqaha, A., & Hoang, D. T. (2021). Challenges and countermeasures for adversarial attacks on deep reinforcement learning. IEEE Access, 9, 23014–23036. https://ieeexplore.ieee.org/document/9536399

Jiang, K., Wang, W., Wang, A., & Wu, H. (2020). Network intrusion detection combined hybrid sampling with deep hierarchical network. IEEE Access, 8, 32464–32476. https://doi.org/10.1109/ACCESS.2020.2973730

Karatas, G., Demir, O., & Sahingoz, O. K. (2020). Increasing the performance of machine-learning-based IDSs on an imbalanced and up-to-date dataset. IEEE Access, 8, 32150–32162. https://doi.org/10.1109/ACCESS.2020.2973219

Kausar, F., Deo, S., Hussain, S., & Haque, Z. U. (2024). Federated deep learning model for false data injection attack detection in cyber-physical power systems. *Energies, 17*(21), 5337. https://doi.org/10.3390/en17215337

Karimy, A. U., & Kumari, M. T. M. (2021). Intelligent intrusion detection system using deep learning and extreme machine learning algorithms. International Journal of Research, 9(11), 2320–2882. https://www.researchgate.net/publication/367221202

Khan, M. A., & Kim, J. (2020). Toward developing efficient Conv-AE-based intrusion detection system using heterogeneous dataset. Electronics, 9(11), 1771. https://doi.org/10.3390/electronics9111771

Kumar, G., Alam, M., & Sharma, R. (2022). Comparative analysis of machine-learning models for phishing detection. Computers & Security, 119, 102792. https://doi.org/10.1016/j.cose.2022.102792

Kumari, T. M., & Karimy, A. U. (2021). Intelligent intrusion detection system using deep learning and extreme machine learning algorithms. International Journal of Creative Research Thoughts, 9(11), 1–10.

Larriva-Novo, X., Vega-Barbas, M., & Villagrá, V. A. (2021). An IoT-focused intrusion-detection-system approach based on preprocessing characterization. Sensors, 21(2), 656. https://doi.org/10.3390/s21020656

Latif, N., Ma, W., & Ahmad, H. B. (2025). Advancements in securing federated learning with IDS: A comprehensive review of neural networks and feature-engineering techniques for malicious-client detection. Artificial Intelligence Review, 58(3), 1893–1935. https://doi.org/10.1007/s10462-024-11082-w

Lawal, K. (2025). Real-world impact and challenges of machine learning in cyber-threat detection. ResearchGate. https://www.researchgate.net/publication/390509598

Liu, C., Gu, Z., & Wang, J. (2021). A hybrid intrusion detection system based on scalable k-means++ random forest and deep learning. IEEE Access, 9, 75729–75740. https://ieeexplore.ieee.org/document/9437227

Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., & Foozy, C. F. M. (2021). Benchmarking of machine learning for anomaly-based intrusion detection systems in the CICIDS2017 dataset. IEEE Access, 9, 22351–22370. https://ieeexplore.ieee.org/document/9437227

Mittal, S., Wazid, M., Singh, D. P., Das, A. K., & Hossain, M. S. (2025). A deep-learning ensemble approach for malware detection in Internet of Things utilizing explainable artificial intelligence. Engineering Applications of Artificial Intelligence, 139(A), 109560. https://doi.org/10.1016/j.engappai.2024.109560

Mohale, V. Z., & Obagbuwa, I. C. (2025). A systematic review on the integration of explainable artificial intelligence in intrusion-detection systems to enhance transparency and interpretability in cybersecurity. Frontiers in Artificial Intelligence, 8, 1526221. https://doi.org/10.3389/frai.2025.1526221

Mushtaq, S., & Javed, T. (2024). Ensemble-learning-powered URL phishing detection: A performance-driven approach. Journal of Informatics and Web Engineering, 3(2), 134–145. https://doi.org/10.33093/jiwe.2024.3.2.10

Musser, M., Lohn, A., Dempsey, J. X., & Spring, J. (2023). Adversarial machine learning and cybersecurity: Risks, challenges, and legal implications. https://doi.org/10.51593/2022CA003

Nakhodchi, S., & Dehghantanha, A. (2020). A bibliometric analysis on the application of deep learning in cybersecurity. In Proceedings of the International Conference on Computational Intelligence and Security (pp. 139–149). Springer. https://doi.org/10.1007/978-3-030-45541-5_11

Nguyen, T. T., & Reddi, V. J. (2021). Deep reinforcement learning for cybersecurity. IEEE Transactions on Neural Networks and Learning Systems, 34(10), 8395–8410. https://doi.org/10.1109/TNNLS.2021.3121870

Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial intelligence in cybersecurity: A comprehensive review and future direction. Applied Artificial Intelligence, 38(1), 1–37. https://doi.org/10.1080/08839514.2024.2439609

Ogundokun, R. O., Awotunde, J. B., Sadiku, P. O., Adeniyi, E. A., Abiodun, M. O., & Dauda, O. I. (2021). An enhanced intrusion detection system using particle swarm optimization feature extraction technique. Procedia Computer Science, 193, 504–512. https://doi.org/10.1016/j.procs.2021.10.052

Priya, A. S., & Sandhiya, A. (2024). Machine learning–based cyber attack detection on internet traffic. *International Journal of Science and Research Archive, 18*(3). https://doi.org/10.30574/ijsra.2024.11.2.0459

Rawindaran, N., Jayal, A., & Prakash, E. (2021). Machine-learning cybersecurity adoption in small and medium enterprises in developed countries. Computers, 10(11), 150. https://doi.org/10.3390/computers10110150

Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2020). Adversarial machine-learning attacks and defense methods in the cybersecurity domain. arXiv. https://arxiv.org/abs/2007.02407

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. Journal of Big Data, 7, 41. https://doi.org/10.1186/s40537-020-00318-5

Seo, W., & Pak, W. (2021). Real-time network intrusion-prevention system based on hybrid machine learning. IEEE Access, 9, 46386–46397. https://ieeexplore.ieee.org/document/9380395

Sharif, F., Khan, M. A., Al-Ghamdi, M. A., & Alshamrani, S. S. (2024). The role of ensemble learning in strengthening intrusion-detection systems: A machine-learning perspective. ResearchGate. https://www.researchgate.net/publication/384366905

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine-learning techniques in cybersecurity. Energies, 13(10), 2509. https://doi.org/10.3390/en13102509

Singhal, A., Maan, A., Chaudhary, D., & Vishwakarma, D. (2021). A hybrid machine-learning and data-mining-based approach to network-intrusion detection. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 312–318). IEEE. https://ieeexplore.ieee.org/document/9395918

Talukder, M. A., Hasan, K. F., Islam, M. M., Uddin, M. A., Akhter, A., Yousuf, M. A., Alharbi, F., & Moni, M. A. (2022). A dependable hybrid machine-learning model for network intrusion detection. Journal of Information Security and Applications, 72, 103405. https://doi.org/10.1016/j.jisa.2022.103405

Thakkar, A., & Lohiya, R. (2021). A review on machine-learning and deep-learning perspectives of IDS for cloud computing. In Digital Forensics and Cyber Crime (Vol. 12, pp. 83–104). Springer. https://www.researchgate.net/publication/344781559

Wisanwanichthan, T., & Thammawichai, M. (2021). A double-layered hybrid approach for network intrusion-detection system using combined naïve Bayes and SVM. IEEE Access, 9, 138432–138450. https://doi.org/10.1109/ACCESS.2021.3118573

Yan, A., Chen, Z., Zhang, H., Peng, L., Yan, Q., Hassan, M. U., Zhao, C., & Yang, B. (2021). Effective detection of mobile-malware behavior based on an explainable deep-neural network. Neurocomputing, 453, 482–492. https://doi.org/10.1016/j.neucom.2021.04.106

Zhang, Y., Yang, C., Huang, K., & Li, Y. (2022). Intrusion detection of industrial Internet of Things based on reconstructed graph neural networks. IEEE Access, 10, 1–14. https://doi.org/10.1109/ACCESS.2022.3169964

Zhang, S., Hao, Q., Gong, Z., Zhu, F., Wang, Y., & Yang, W. (2024). MDD-FedGNN: A vertical federated graph-learning framework for malicious-domain detection. Computers & Security, 147, 104093. https://doi.org/10.1016/j.cose.2024.104093

Zhang, J., Pan, L., Han, Q.-L., Chen, C., Wen, S., & Xiang, Y. (2022). Deep learning–based attack detection for cyber-physical system cybersecurity: A survey. *IEEE/CAA Journal of Automatica Sinica, 9*(3), 377–391. https://doi.org/10.1109/JAS.2021.1004261

Zhu, J., & Liu, X. (2024). An integrated intrusion-detection framework based on subspace clustering and ensemble learning. Computers & Electrical Engineering, 115, 109113. https://doi.org/10.1016/j.compeleceng.2024.109113

How to Cite this Chapter (APA 7):

Alarcon, S. P., Bangot, Q., Bue, J., Alvarez, M. J., Estero, J., & Velasquez, J. Q. (2026). Machine learning for cyber-attack detection: A literature review of ensemble, hybrid, and explainable AI approaches. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 65–76). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 10: AI-Assisted Code Generation: A Comparative Study of Developer Productivity, Code Quality, and Usability

Kristian Angelo Cantada[1], Rei Joshuel Sajenes[1], Aerhon Kim Salazar[1], Frederick Masikat[1] & Eleandro C. Anciro[2]
[1]Cavite State University – Silang Campus, Cavite, Philippines
[2]Philippine Christian University - Dasmariñas, Cavite, Philippines

Abstract. *AI-powered tools, particularly those based on large language models such as GitHub Copilot and ChatGPT, are reshaping software development by enabling automated code generation, enhancing developer productivity, and raising critical ethical concerns, including algorithmic bias, over-reliance on AI systems, and potential risks of unintentional plagiarism. This study evaluates the efficiency, usability, and experiences of developers using these AI tools to gain insights into their role in modern coding practices A comparative mixed-method research design was employed, combining quantitative measures (productivity gains, code quality, error rates) with qualitative findings (developer feedback, usability, satisfaction). Results show that AI-powered expert systems significantly support developers in coding, debugging, and testing. However, excessive dependence on AI raises concerns about reduced problem-solving ability and originality. This research highlights the importance of selecting AI tools that match development context and calls for future systems to be ethical, transparent, and adaptable. It also underscores the need for responsible use, especially in avoiding code plagiarism and ensuring developers maintain essential skills. Overall, the study contributes practical insights into the responsible adoption of AI in software engineering.*

Keywords: *AI code generation, Github Copilot, ChatGPT, developer productivity, usability, ethics*

Introduction

With the era of new technologies and Artificial Intelligence (AI), the role of developers in shaping the future of tech and AI becomes essential. AI can be programmed manually by developers or using AI tools, for example, expert systems that mimic human decision-making based on rule-based reasoning, knowledge representation, and machine learning. These AI professional systems are now central to automating code generation, debugging, testing, and support through structured knowledge bases and learning algorithms. The recent innovations in AI-supported code generation, such as software like GitHub Copilot, have improved productivity and code quality significantly, providing automation, smart suggestions, and ease of integration into development environments (Umeh & Umeh, 2024). AI profoundly affects the software development cycle, particularly during implementation where Automated Code Generation (ACG) helps programmers translate requirements into code, improving productivity and handling frequent problems.

This research gives a detailed review of conventional and AI-based ACG approaches, discussing their usage, merits, demerits, and performance with measures such as Accuracy, Efficiency, Scalability, and Generalization. We contrast AI techniques like rule-based, machine learning, deep learning (e.g., GPT, BERT), and evolutionary algorithms. Deep learning models are very good at recognizing intricate patterns but are uninterpretable and prone to bias; rule-based systems are interpretable yet rigid. Disputes regarding evaluation criteria and ethical issues such as bias and job replacement underscore the necessity for hybrid methods and continued study to triangulate transparency, ethics, and performance (Odeh et al., 2024). As AI redefines software development, measuring the effectiveness, usability, and efficiency of systems continues to be critical to comprehend and improve next-generation innovations.

AI-driven expert systems are becoming more known in the software development industry, especially in code generation. A lot of research indicates that such systems can increase productivity and quality in code by automating the routine work and helping in complex coding decisions (Alonso et al., 2024) but on the other hand, it is not definitely yet clear how effective they are in everyday development. While some works point out quantifiable enhancement for code output and efficiency considering, some warning studies suggests potential negative effect for example high occurrence of code plagiarism, bugs hidden and developers reliance on the AI-generated suggestions (Alonso et al., 2024). The trustworthiness of produced code and the potential for bias in algorithmic advice alike invite critical reflection, particularly around ethical concerns such as plagiarism and IP (Birkenkrahe, 2024). Others warn of possible adverse side effects, such as an increase in code plagiarism, the going viral of tiny bugs, and overdependence on AI help. Knowing which systems are valuable and when they are best utilized is essential in the increasingly AI-integrated pipelines of today. The reliability and code independence of users create a question about ethical considerations, such as plagiarism risks and bias in code suggestions; it must also be addressed, especially as these tools become more integrated into development workflows.

This paper aims to explore the effects of AI-driven expert systems, such as ChatGPT, GitHub Copilot, and other similar AI tools , on how the developers write code. The purpose is to see how useful and effective these tools are in helping developers with coding during the software development process.

Materials and Methods

The Design

The research design used is a comparative mixed-methods approach to study the usability and productivity effects of AI-powered expert systems in code generation. The research design combines quantitative and qualitative methods to provide a comprehensive evaluation of AI tools including GitHub Copilot and ChatGPT. The study followed the systematic research method outlined by Durach et al. (2017) in six steps to maintain a structured investigation. The research question focused on understanding how expert systems powered by AI improve intelligent code generation and what elements determine their success in supporting software developers. The evaluation criteria and attributes for assessment included productivity alongside code quality and usability and developer feedback. The researchers collected data from academic articles and developer blogs and official tool documentation and case studies about AI code generation tools. The collected materials underwent filtering to validate their relevance and credibility and maintain recentness. The researchers used both primary and secondary data sources which included surveys and interviews with software developers and published studies. The study combines quantitative measures (e.g., error rates, time efficiency, code quality) and qualitative analysis (developer feedback, thematic analysis of usability and satisfaction). Following Durach et al.'s (2017) six-step systematic review procedure, the methodology involved:

Problem definition – identifying gaps in AI usage for coding productivity and usability. Search strategy – conducting structured searches across Google Scholar, Scopus, IEEE Xplore, and ACM DL using terms such as *"AI code generation," "developer productivity,"* and *"ChatGPT/Copilot in software development."* Screening and selection – inclusion criteria: peer-reviewed articles from 2021–2025, developer blogs with practical insights, and case studies of AI tools.

Data extraction – extracting themes, metrics, and findings about productivity, usability, and ethical implications. Quality assessment – evaluating source credibility based on author expertise, citation count, and recency. Synthesis and interpretation – triangulating findings using both narrative and numerical synthesis.

The study used primary data via surveys and interviews with 20 software developers (junior to senior level), and secondary data from published papers, blogs, and documentation. Quantitative analysis measured changes in productivity and code quality, while qualitative thematic coding analyzed user satisfaction and usability challenges.

Literature Synthesis

Table 1. The efficiency of AI-powered expert systems in maximizing software development productivity.

Category	Authors	Efficiency
AI for Code Generation & Automation	Szikszó (2024); Odeh (2024); Ghai et al. (2024); Johnson-Mary (2024); Smith et al. (2024); Sauvola et al. (2024); Arora et al. (2023)	AI automates coding, generates code from natural language, reduces development time, and increases productivity
AI for Bug Detection & Testing Automation	Anand et al. (2024); Khaliq et al. (2022)	AI automates debugging and software testing, reducing manual effort and improving code quality
AI for Software Design & Optimization	Gollapalli (2021)	AI techniques such as neural architecture search, GAN, and reinforcement learning improve code optimization and execution efficiency
AI for Project Management & Decision-Making	Saha Srija (2024); Nguyen-Duc et al. (2023); Wang et al. (2024)	AI supports decision-making, predictive analytics, and improves overall software development efficiency
AI in Requirements Engineering	Güzel & Egesoy (2025); Arora et al. (2023)	AI improves requirements analysis and handles unclear requirements using fuzzy logic and automation
AI Adoption, Trends & Development Processes	Zhang (2024); Martínez-Fernández et al. (2022); Kokol (2024)	AI improves efficiency across development lifecycle and shows increasing adoption in software engineering
AI Challenges, Ethics & Trust	Tistelgrén (2024); Nasir & Kallinteris (2025); Wang et al. (2024)	AI adoption faces issues such as ethics, bias, interpretability, and developer trust

Table 2. Usability of f expert systems fueled by AI in supporting code activities among software developers.

Category	Authors	Usability of AI in Supporting Code Activities
Time Efficiency & Productivity	Szikszó (2024); Tistelgrén (2024); Johnson-Mary (2024); Smith et al. (2024)	AI saves time, automates repetitive tasks, improves productivity and workflow

Requirements Engineering & Project Management	Güzel & Egesoy (2025); Arora et al. (2023); Nguyen-Duc et al. (2023)	AI supports requirements writing, project management, and design processes
Code Assistance & Smart Development Tools	Anand et al. (2024); Gollapalli (2021); Odeh (2024)	AI provides code suggestions, natural language coding, and smart code editors
Software Quality & Error Detection	Saha Srija (2024); Khaliq et al. (2022)	AI helps detect errors early and improves software quality
Integration & Ease of Learning	Ghai et al. (2024); Sauvola et al. (2024); Smith et al. (2024)	AI tools integrate into existing development tools and are easy to learn
Flexibility, Scalability & Collaboration	Nasir & Kallinteris (2025)	AI improves system flexibility, scalability, collaboration, and adaptive feedback
User Experience, Usability & Adoption Challenges	Kokol (2024); Martínez-Fernández et al. (2022); Wang et al. (2024)	AI usability depends on user experience, learning curve, trust, and proper configuration

Table 3. Effectiveness of different AI-driven expert systems in assisting software developers.

Category	Authors	Effectiveness of AI-Driven Expert Systems
Code Generation & Development Efficiency	Szikszó (2024); Odeh (2024); Gollapalli (2021); Nguyen-Duc et al. (2023)	AI improves development efficiency, automates coding tasks, and enhances software implementation and maintenance
Software Quality, Accuracy & Reliability	Anand et al. (2024); Ghai et al. (2024); Smith et al. (2024); Sauvola et al. (2024); Khaliq et al. (2022); Arora et al. (2023)	AI improves code correctness, reliability, maintainability, and defect detection
Project Management & Process Optimization	Güzel & Egesoy (2025); Saha Srija (2024)	AI improves project tracking, task prioritization, testing optimization, and requirements analysis
System Performance & Optimization	Kokol (2024)	AI improves system performance, optimization, and software quality metrics
Innovation, Value Creation & Automation	Tistelgrén (2024); Zhang (2024)	AI supports innovation, automation, and future adaptive software systems
Reliability, Safety & Trust in AI Systems	Martínez-Fernández et al. (2022); Wang et al. (2024)	Effectiveness depends on reliability, safety, trust, and transparency of AI systems

Table 4..Comparison of AI-driven expert systems' approaches to coding

Category	Authors	Comparison of AI-Driven Expert Systems' Approaches to Coding
AI vs Traditional Software Development	Odeh (2024); Gollapalli (2021); Saha Srija (2024); Ghai et al. (2024); Johnson-Mary (2024); Smith et al. (2024); Sauvola et al. (2024)	AI-based approaches improve efficiency, accuracy, and automation compared to traditional methods, but require large datasets and computational resources
AI for Code Generation, Debugging & Testing	Szikszó (2024); Srinikhita Tistelgrén (2024); Anand et al. (2024); Khaliq et al. (2022)	AI supports automated code generation, debugging, testing, and code optimization
AI in Software Development Lifecycle	Zhang (2024); Nguyen-Duc et al. (2023)	AI can be integrated across the software development lifecycle including design, development, testing, project management, and maintenance
AI for Requirements Engineering & Project Management	Güzel & Egesoy (2025); Arora et al. (2023)	AI improves requirements management, project tracking, and decision-making
AI Collaboration, Trust, Ethics & Bias	Nasir & Kallinteris (2025); Wang et al. (2024)	AI introduces collaboration between humans and AI but raises issues in trust, ethics, bias, and interpretability
Comparison of Different AI Approaches & Methods	Martínez-Fernández et al. (2022); Kokol (2024)	Different AI-driven software engineering approaches have varying strengths, limitations, and research gaps

Discussion

3.1 The efficiency of AI-powered expert systems in maximizing software development productivity and code quality.

The world is evolving, and there is a dynamic change everywhere around us. Technology is improving and evolving in a positive direction. The explosive growth of artificial intelligence (AI) has had a profound impact on the face of software engineering. Expert systems powered by AI now dominate the field of maximizing productivity, automating code processes, increasing quality assurance, and overall efficiency. This part gives a detailed discourse of the major findings from some of the identified scholarly sources, as indicated in Table 1, concerning the effectiveness of AI-driven systems in software development.

The study of AI software development tools is rapidly increasing since it may increase productivity in that field. According to authors such as Szikszó (2024) and Ghai et al. (2024), the natural language interface reduces the cognitive load on programmers by allowing them to define functions in natural language, which are then translated into executable code. These are sometimes generated by ChatGPT since the platform generates code snippets that contain natural language inputs and basic codes that enable the developers' code to function more efficiently and rapidly than using traditional methods of just typing and thinking broadly for themselves. According to Johnson-Mary (2024), and Smith, Doe, & Lee (2024), the role of AI in increasing efficiency in development is significantly getting higher. As the role of AI increases, human involvement decreases. It also hastens development cycles by reducing manual work and less time dedicated to mundane coding tasks. Additionally, Sauvola et al. (2024) assessed the effect of generative AI on reducing mundane activities like boilerplate code generation and syntax fixing, further supporting the assertion that AI tools are effective productivity boosters. With these burdens transferred, developers get to concentrate on sophisticated logic and architecturally significant decisions.One of the most measurable contributions is that of Gollapalli (2021), who showed quantifiable gains in development performance. His results indicate that integrating Generative Adversarial Networks (GANs) and Reinforcement Learning (RL) enhanced code generation correctness to 96.5% and execution time to a mere 105.2 milliseconds, which are significant gains in efficiency and responsiveness.

AI systems considerably help in the quality of software generated. Anand et al. (2024) and Nguyen-Duc et al. (2023) stated that large language models (LLMs) can now be leveraged for automated bug detection and correction so that the debugging process is streamlined, reducing the chances of human error; thus, cleaner, more robust code means fewer issues arise at the deployment stages of software. Predictive tools enable prediction of potential failure points or performance bottlenecks before they occur and enable development teams to apply preventive measures to enhance the reliability and maintainability of the final product. However, Zhang (2024) and Kokol (2024) saw it from a different perspective of the entire design and implementation life cycle in relation to how AI is applied in all design, implementation, testing, and maintenance. They suggest that applying AI tools consistently throughout the software development lifecycle increases productivity, resulting in not only greater consistency but also more adherence to best practices, all of which are indicators of code quality. Additionally, Khaliq et al. (2022) considered software testing, a troubleshooting area commonly associated with manual efforts and human errors. AI allows automating most of the testing tasks, like test case generation and regression testing, which makes wider and more extensive testing coverage possible using fewer resources. This ultimately results in more dependable and less buggy software that has fewer nasty surprises after release. He further considered software testing, a troubleshooting area commonly associated with manual efforts and human errors. AI allows automating most of the testing tasks, like test case generation and regression testing, which makes wider and more extensive testing coverage possible using fewer resources. This ultimately results in more dependable and less buggy software that has fewer nasty surprises after release (Khaliq, 2022).

Various researches also examined the ways AI supports automation and informed decision-making in software development. Scholars such as Güzel & Egesoy (2025) suggested the Fuzzy Specification Tree (FST) model that combines AI and fuzzy logic to address ambiguities in software requirements. By supporting the elimination of vague user expectations, the model improves requirement accuracy and reduces time spent on iterative clarification between developers and stakeholders. Similarly, Arora et al. (2023) asserted and established the way AI, in particular LLMs, can be used to automate tasks in requirements engineering (RE), arguably the most vital yet challenging domain within the process. Computer-aided RE improves specifications' precision and allows for quicker transition into design and coding phases, improving efficiency in every area. Nguyen-Duc et al. (2023) further expanded and advanced the discussion by introducing Generative AI (GenAI) to the toolkit for automating not only low-level coding tasks but also strategic choice. By applying GenAI to project management software, teams can simulate project results, streamline task assignments, and prioritize issues more effectively. Tistelgrén (2024) and Martínez-Fernández et al. (2022) also introduced and synthesized a broader vision of the capacity of AI in coordinating complex workflows and interdependencies in massive software development projects. Their systematic review of scholarly studies shows that decision-support systems facilitated by AI are becoming crucial in coordinating project timelines, resource management, and risk factors.

Apart from their benefits, AI expert systems have a number of challenges.Some researchers like Tistelgrén (2024),for example, cites ethical and technical issues regarding the use of AI in software development, such as knowledge gaps within developers, the trustworthiness of AI results, and possible over dependence on automation. Developers need to continue testing and verifying AI written code, which adds another layer of complexity. Nasir and Kallinteris (2025) further pointed out challenges of interpretability and bias in multi-agent systems (MAS) that employ AI, specifically LLMs. With models becoming increasingly sophisticated, it is harder to understand their decision-making process. This transparency can discourage developers' faith in the system and complicate diagnosis and correction of problems that occur from AI-generated outcomes. Wang et al. (2024) studied the influence of developer trust on utilizing AI-supported coding tools. According to their study, productivity benefits derived

from AI systems critically depend on the confidence developers have in the systems. When trust levels are high, developers will be more likely to adopt AI suggestions, speeding up development. Yet when trust is low, human verification is more common, possibly offsetting productivity gains. Integration issues are also addressed in Zhang (2024) and Kokol (2024), who stated that implementing AI tools into existing development workflows can be a labor-intensive task, involving retraining staff, adapting workflows, and ironing out compatibility issues with legacy software architecture. AI usage issues regarding data privacy, intellectual property, and ethical use of AI continue to hang around in the academic dialogue. For instance, the use of cloud-based AI services makes it problematic how sensitive project data is processed and stored. Compliance with data protection legislation becomes the top priority in such cases.

3.2 The usability of expert systems fueled by AI in supporting code activities among software developers.

The growing and fast-moving technology era is at hand, and software development plays a big role in this time. Software development requires a lot of work to do, demands a lot of complex processes, often filled with challenges in terms of debugging, time-consuming understanding of complex codes, and implementing unfamiliar functionalities. In these past few years, AI-powered expert systems have started to get involved and plays a role in the software engineering landscape. Code generation, review, and comprehension are the main thing that has been internalized by AI driven expert systems. These systems are transitioning from being viewed as supplementary tools to becoming collaborative partners and sometimes developers depend on it in the development process.

User experience describes the overall experience of a person utilizing a software application, whereas usability refers to its efficiency, simplicity, and general satisfaction obtained from the utilization of the system. Author Szikszó (2024) stated that tools like GitHub Copilot can suggest code snippets, autocomplete functions, and even write full and functional blocks of codes based on a few prompt, suggestions, and commands. These features aim to help the developer to reduce repetitive coding and save effort in time. From the standpoint of usability, the value lies in making the process easier and decrease the need for mental effort in completing the requirement for a task.

On the other hand, the true test of usability goes beyond speed. It also concerns how easy for the system to be embedded into the developer's existing nature of work. According to authors such as Güzel and Egesoy (2025), for AI systems to be truly effective, they must be well-integrated into development environments such as Visual Studio Code, IntelliJ, or other IDEs commonly used by professionals. The developers gain a lot of benefits in terms of when this thing in already accessible within the coding interface, it eliminates the need for the developer to switch between platforms and form a lot of tabs. When these AI tools are already embedded directly in the landscape of the developer, it became an extension that go further and extends the programmer's capability.

Developer control in another critical aspect of usability. A reliable and highly usable AI must support the decision-making and not override it, even though they can replicate the nature of work the developers. Also, developers want to be responsible and remain to be in charge of their work, even when AI-generated suggestions are present. A recent study by Zhang (2024) points out that developers prefer tools that offer optional guidance, allowing them to accept, reject, or modify the code proposals. This ensures that AI functions remains as an assistant rather than an automated replacement.

Usability also includes the AI's role in promoting learning and skill growth. According to Tistelgrén (2024), that many developers, especially students and early-career programmers, benefit from AI systems that not only suggest solutions but also explain the reasoning behind them. This transforms the AI into a learning aid, similar to a mentor or tutor. Exposing developers to a wide variety of coding patterns, and giving different explanations, will help them internalize best practices and learn to think about problems over time. When it teaches rather than helps, a system becomes more usable.

However, there are usablility issues remaining to be addressed for the AI expert systems. Leading concerns include incorrect or suboptimal code-generations. Nasir and Kallinteris (2025) showed that AI systems can generate "seemingly reasonable incorrect" solutions that seem plausible, but which are in fact bugged or inefficient. For beginners in particular, there's a risk they will happily take any recommendation and assume it's what they should do. So a good system has to provide some means of assisting developers in verifying the correctness of code and promoting thought, not just acceptance.

One other worry is becoming too dependent on AI, which can lead to less meaningful interaction and slower advancement of essential programming skills. Research finding by Nguyen-Duc et al. (2023) warn that habitual reliance on AI tools may cause 'skill decay,' when developers distance themselves from foundational knowledge. This downside can also harm the long-term usability, causing the ability of a developer to work on his/her own to decrease. One way to do this is to design systems that strike a balance between providing help and encouraging users to first try solutions themselves before providing them.

Clearly, AI expert systems can be used as well in cooperative development. In a team setting, they can help enforcing uniformity across code by suggesting a common style and allowing collaboration to flow more easily, by enforcing strict rules. According to research by Szikszó (2024), in collaborative repositories, AI can help with detection of code reviews and highlighting of potential issues in advance to increase team productivity. If such tools are made with collaboration in mind, they can be used not only for personal support but for team-wide benefit.

3.3 The effectiveness of different AI-driven expert systems in assisting software developers.

The swift rise of Artificial Intelligence (AI) is bringing about a new paradigm for software development. AI-based expert systems are increasingly being used by developers to facilitate programing, automate certain tasks, and improve productivity. Judging the actual utility of these tools is essential to understand their role and to maximize their usefulness.

There is a wide range of functions available from AI expert systems to help software developers in many aspects of the software development process. Code generation has made the most noise in the potential of AI. With tools like ChatGPT, it's also possible to generate code snippets from natural language, where developers will cut coding time and mundane operations (Szikszó, 2024). This can release developers from the tyranny of focusing on the more complicated and interesting parts of software development.

In addition to code generation, AI enables dynamic project tracking and task prioritization that helps working more efficiently at project management and workflow by eliminating waste (Güzel & Egesoy, 2025). AI can aid with organizing, scheduling and resource management by automating some project management functions. Also, optimization techniques based on AI have demonstrated potential to enhance system performance, maintainability, and code quality. AI analyses the code and can point out places needing improvement or optimizations that improve performance while decreasing technical debt.

Some other researches give the empirical result about AI's good influence on development efficiency. According to Gollapalli (2021), JPMorgan reached a 20% improvement in the efficiency of coders who used an AI based coding tool. This reveals that AI can in fact deliver orders of magnitude of improved efficiency, freeing up developers for higher value and strategic projects. AI solutions assist in performing excellent automation in functions such as identification of defects in codes, test optimization, and analysis of requirements which in turn result in better project outcomes and better software quality (Saha Srija, 2024). AI can do the accuracy and reliability AI can automate these manual and error-prone tasks and refine all phases of the whole SDLC. Ghai et al. (2024) point out that with AI guided automation for code generation, the time and investment for software development can be significantly reduced, which can also shorten the time to market.

Researchers have defined a number of evaluation criteria and metrics to evaluate AI techniques used in automated code generation in a rigorous manner. Odeh (2024) implies some rules of thumbs including accuracy, economy, scalability, generalization, and correctness. These criteria serve as guidelines for assessing the performance of AI tools and contrasting different code creation methods. Accuracy is the capability of AI to produce correct and useful code, while efficiency is the computational resources and time used by code generation. Scalability tests how well AI tools are able to cope with large-scale and complex projects, and generalization evaluates how well they generate working code for a broader range of use cases. The correctness aspect makes sure that the produced code is in conformance with requirements and performs as expected.

It is crucial to scrutinize the correctness and reliability of AI-created code. Ghai et al. (2024) and Smith et al. (2024) emphasize importance of correctness, efficiency, and maintainability. Equivalence works under the premise that the code must be correct i.e., the generated code should behave as it was designed and give the expected results. Performance Consider the function of the code, which can be seen for how time and resources it uses. NMT1 (maintainability)n nudn u584n u584 n Maintainability is a measure of how easy it is to understand, update and maintain the code. Kokol (2024) highlights the requirement for metrics to measure the effectiveness of AI applications on software quality and global performance.

The literature portrays AI-powered expert systems as being quite advantageous, but it identifies possible challenges and downsides that should be handled. Zhang AZA) expresses worries of excessive use of AI software, who is going to judge upon ethical matters, what about security. Finding the trade-off between taking advantage of AI tools and human supervision of code to preserve quality and prevent side effects is essential. Pillai et al. (2024) warns against side effects such as plagiarism of code, hidden faults and use of AI support as a crutch. These issues exemplify the need to encourage responsible use of AI and inform developers of possible fallacies.

Ethical concerns are also of major relevance. The dependability of code produced and the possibility of algorithm suggestions leading to forms of bias pose questions of plagiarism and intellectual property (Birkenkrahe, 2024; Yang, et al., 2024). Developers and corporations need to come up with a clear set of guidelines and best practices for ethical use of AI tools and avoiding potential abuse of AI-generated code. In addition, Nguyen-Duc et al. (2023) argue that while AI-powered tools hold potential to automate certain tasks, research and development in aspects such as implementation, quality assurance, and maintenance still need to prove GenAI efficiency. Wang et al. (2024) stress the importance of trust of developers in AI code generation tools in general and argue the necessity of addressing productivity factors that enable or hinder productivity and the need to inspire confidence to developers, while using AI mechanisms.

A handful of other studies shed some light on companies that are using AI in software development in practice and hint at where it may be headed. For example, Szikszó (2024) reports a small experiment in which ChatGPT was employed to help create a simple portfolio website, illustrating a practical use of AI tools in coding tasks. Other work considers AI within a particular context, for example inautomated program repair (APR) and code synthesis (Anand et al., 2024), and in the use of Multi-AgentSystems(MAS) for improved cooperation and productivity in software engineering (Nasir & Kallinteris, 2025).

There are several promising directions for AI in software development based on the review of the literature. Zhang (2024) foresees improvements such as adaptive systems that can learn and change with time and, perhaps whole software automation. Click here to download a free preview of the questionnaire There is increasing focus on how to deal with ethical issues, the

enhancement of interpretability of AI models and the removal of biases in AI systems to ensure fairness and transparency (Nasir & Kallinteris, 2025). Further research is needed to better define evaluation criteria, hybrid approaches of AI and human expertise, and to facilitate the responsible and effective incorporation of AI in software development process.

3.4 Comparison of AI-driven expert systems' approaches to coding

By reviewing these tools holistically in terms of efficacy as well as the user experience, this study sheds light on their real world applications, and limitations. Now that AI is more and more present in development environments, such comparative research is more important as ever to support developers in well-considered tool adoption (Szikszó, 2024).

AI expert systems increase not just computational accuracy, but also developer productivity. This is in line with the overall objective of the study, that is to assess the usability and performance of such systems in varying development environments. While some tools are designed to be simple and easy to use for beginners, others offer more advanced features for the masters of the domain. Perceiving these differences can help identify how AI-tools influence coding practices and also design efforts to be taken (Güzel & Egesoy, 2025).

The advancement of AI has also enabled the creation of programming helpers who can produce context aware snippets of code, relieving cognitive burden and helping in the speed of development. Various tools like GitHub Copilot, Tabnine, and Amazon CodeWhisperer are compared to the one in our work, and it illustrates different design philosophies. We employ copilot, a code generation model based on large language models trained over large sources of open-source code, to generate multi-line suggestions considering both the surrounding code and natural language prompts. CodeWhisperer, conversely, is integrated into enterprise and cloud ecosystems and focuses on secure and policy compliant code suggestions, especially for AWS. In the mean time, Tabnine utilizes smaller, downloadable models to enhance home development, avoiding infringing privacy and supporting offline (Zhang, 2024).

These have quite disparate architectures, and practical merits. GitHub Copilot provides a flexible, creative coding experience, but may also generate errors or insecure code. CodeWhisperer focuses on security and compliance, so its target use case is in enterprises. Tabnine prioritizes developer independence and data privacy, despite potentially not having the generative depth of larger models (Tistelgrén, 2024).

Such variation highlights the need to match tool selection to your development goals—be they creativity, security, adaptability, or performance. For example: GitHub Copilot is great for generic use cases whereas CodeWhisperer is a better fit for regulated / enterprise environments. Tabnine offers a better choice if you value privacy and local control. Finally, this comparative evaluation also highlights how various AI systems trade off different goals. Some of them target human-like and fluent output, and some other of them look for a deterministic and secure behavior. Such trade-offs impact on user-experiences and more widespread adoption of such systems in real-world professional settings (Anand et al., 2024).

AI based expert systems are now important for code completion, debugging and synthesis. Using rule-based logic, natural language processing, neural networks, etc., they simulate the human reasoning to save the manual programming efforts. These tools can make accurate, context-sensitive suggestions allowing developers to produce their programs more effectively and efficiently. As we show in this paper, GithubCopilot, CodeWhisperer, and Tabnine each introduce a variation of the secondsuch approach that addresses a different development environment and user requirement (Odeh, 2024).

Conclusion

The research was a clear indication that AI-powered expert systems like GitHub Copilot and ChatGPT are currently capable of helping developers to complete tasks more quickly using a reduced amount of manual effort. They're designed to help you write code, debug errors, or even manage some of the complexity of working on bigger projects. Tools like these help developers do their job more easily and less stressfully, where repetitive or overly time-consuming tasks are concerned, many of the developers said. Based on those results, AI does a good job of boosting productivity and and making developers more efficient at what they're doing. But this new method of coding is not without flaw, and has its obstacles. The experiment showed that both an overdependence on AI leads to problems, such as reduced creativity, skill atrophy, and the propagation on unseen mistakes. It also implies risks of code plagiarism, ethical issues, lack of innovation and originality. Developers may accept AI-generated code on its face, without verification, something that could reduce code quality in the long term. That's why it is so critical to use AI with responsibility. When and how these tools should be used are the two main questions that developers and companies need to consider. They should be careful, especially about privacy, trust, and ownership of code. AI should be a mentor, not a surrogate. There must be better rules and standards so that AI assists the work rather than does violence to it. Ultimately, the usefulness of AI will come down to what we choose to do with it. If we apply it sensibly and develop good habits, AI can genuinely enable developers, not to the detriment of thought. So the future of software development isn't just better AI, but wiser decisions from people who use that AI.

References

Alonso, J. M., Álvarez-Rodríguez, J. M., Martínez, L., & Meersman, R. (2024). AI-based techniques for automated code generation: A review of methods and evaluation criteria. *ACM Digital Library.* https://doi.org/10.1145/3650212.3680347

Anand, A., Gupta, A., Yadav, N., & Bajaj, S. (2024, November 12). A comprehensive survey of AI-Driven advancements and techniques in automated program repair and code generation. arXiv.org. https://arxiv.org/abs/2411.07586

Arora, C., Grundy, J., & Abdelrazek, M. (2023, October 21). Advancing Requirements Engineering through Generative AI: Assessing the Role of LLMs. arXiv.org. https://arxiv.org/abs/2310.13976

Birkenkrahe, M. (2024). The role of AI coding assistants: Revisiting the need for literate programming in computer and data science education. In INTED2024 Proceedings: 18th International Conference on Technology, Education and Development. IATED Academy. https://doi.org/10.21125/inted.2024.0071

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. Journal of Supply Chain Management, 53(4), 67–85. https://doi.org/10.1111/jscm.12145

Ghai, A. S., Rawat, V., Gupta, V. K., & Ghai, K. (2024). Artificial intelligence in system and software engineering for auto code generation. In 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) (pp. 1–5). IEEE. https://doi.org/10.1109/ICEECT61758.2024.10738945

Gollapalli, V. S. T. (2021, February 17). Generative AI for intelligent code synthesis: Advancing automated software development and optimization. https://ijerst.org/index.php/ijerst/article/view/579

Güzel, A., & Egesoy, A. (2025, January 1). Development of an AI-Driven model for advancing software engineering practices. https://ijircstjournal.org/index.php/ijircst/article/view/13

Johnson-Mary, B. (2024). Revolutionizing Software Development with AI: The Path to 2030. ResearchGate. https://www.researchgate.net/profile/Britney-Johnson-Mary/publication/390209191_Revolutionizing_Software_Development_with_AI_The_Path_to_2030/links/67e47f2143ec6369e2ff6198/Revolutionizing-Software-Development-with-AI-The-Path-to-2030.pdf

Khaliq, Z., Farooq, S. U., & Khan, D. A. (2022, January 14). Artificial intelligence in software testing: Impact, problems, challenges and prospect. arXiv.org. https://arxiv.org/abs/2201.05371

Kokol, P. (2024). The use of AI in software engineering: A synthetic knowledge synthesis of the recent research literature. Information, 15(6), 354. https://doi.org/10.3390/info15060354

Martínez-Fernández, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., Vollmer, A. M., & Wagner, S. (2022). Software Engineering for AI-Based Systems: A survey. ACM Transactions on Software Engineering and Methodology, 31(2), 1–59. https://doi.org/10.1145/3487043

Nasir, W., & Kallinteris, N. (2025). From code generation to AI collaboration: The role of multi-agent systems in software engineering. https://www.researchgate.net/profile/Nikoletta-Kallinteris/publication/388835330_From_Code_Generation_to_AI_Collaboration_The_Role_of_Multi-Agent_Systems_in_Software_Engineering/links/67a900d9207c0c20fa81a501/From-Code-Generation-to-AI-Collaboration-The-Role-of-Multi-Agent-Systems-in-Software-Engineering.pdf

Nguyen-Duc, A., Cabrero-Daniel, B., Przybylek, A., Arora, C., Khanna, D., Herda, T., Rafiq, U., Melegati, J., Guerra, E., Kemell, K., Saari, M., Zhang, Z., Le, H., Quan, T., & Abrahamsson, P. (2023, October 28). Generative Artificial Intelligence for Software Engineering. arXiv.org. https://arxiv.org/abs/2310.18648

Odeh, A. (2024). Exploring AI innovations in automated software source code generation: Progress, hurdles, and future paths. Informatica, 48(8). https://doi.org/10.31449/inf.v48i8.5291

Odeh, A., Odeh, N., & Mohammed, A. S. (2024). A comparative review of AI techniques for automated code generation in software development: Advancements, challenges, and future directions. TEM Journal, 726–739. https://doi.org/10.18421/tem131-76

Pillai, N., Thomas, A., & Kumar, S. (2024). AI expert systems in software engineering: Opportunities, limitations, and ethical challenges. Information and Software Technology. https://www.sciencedirect.com/science/article/abs/pii/S0950584924002155

Sauvola, J., Tarkoma, S., Klemettinen, M., Riekki, J., & Doermann, D. (2024). Future software development with generative AI. https://link.springer.com/article/10.1007/s10515-024-00426-z

Szikszó, Z. P. (2024). Artificial intelligence for software developers (By Haaga-Helia University of Applied Sciences) [Thesis]. https://www.theseus.fi/bitstream/handle/10024/859306/Szikszo_Zoltan.pdf?sequence=2

Tistelgrén, S. (2024). Artificial intelligence in software development: Exploring utilisation, tools, and value creation. Theseus. https://www.theseus.fi/handle/10024/853451

Umeh, I. I., & Umeh, K. C. (2025). A comparative analysis of AI system development tools for improved outcomes. International Journal of Sustainability Management and Information Technologies, 11(1), 1–20.

Wang, R., Cheng, R., Ford, D., & Zimmermann, T. (2024). Investigating and designing for trust in AI-powered code generation tools. 2022 ACM Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3630106.3658984

Zhang, Q. (2024). The role of artificial intelligence in modern software engineering. Applied and Computational Engineering, 97(1), 18–23. https://doi.org/10.54254/2755-2721/97/20241339

How to Cite this Chapter (APA 7):

Cantada, K. A., Sajenes, R. J., Salazar, A. K., Masikat, F., & Anciro, E. C. (2026). AI-assisted code generation: A comparative study of developer productivity, code quality, and usability. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 77–85). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 11: Fake Reservation Post Detection on Facebook Resort Pages: A Literature Review

Dr. Cheeryl V. Garcia
Iloilo Science and Technology University - Miagao Campus
Iloilo, Philippines

Abstract. *The development of the social media sites has changed the way the tourism is being booked, and Facebook has become one of the channels where direct accommodation bookings are made. On the one hand, this digital-first approach makes it more accessible and lowers the costs of intermediaries, but on the other hand, it opens travelers to high-risk fraud. This paper is a synthesis of the existing literature on fake reservation posts in Facebook resort groups, and it notes that scams such as identity spoofing, advance payment schemes, phishing, bot-driven fraud, and review manipulation are the most common. Based on the deception detection theory, information asymmetry, and models of trust and credibility, the review analyzed how fraudulent posts use linguistic cues, urgency tactics, and manipulated social proofs to mislead travelers. The paper also examined natural language processing (NLP) models, such as preprocessing of text, TF-IDF, n-grams, embeddings, and sentiment analysis, and machine learning models, such as SVM, Naive Bayes, and transformer models to detect fraud. It has been shown that AI-based solutions are highly accurate in detecting deceptive content, but there are still issues with data asymmetry and the sophistication of scams. This review highlighted the necessity to develop efficient fraud prevention tools in the context of the social media-driven tourism transactions and defines key research gaps in ensuring that travelers are not victims of organized online booking fraud.*

Keywords: *fake news detection, fraud detection, tourism fraud, social media security, deception detection*

I. Introduction

The use of social media sites has radically changed the tourism sector, especially in the way tourists find and reserve accommodation (Franjić and Farrell, 2025). The main touchpoints that have emerged where tourists can find information, discuss experiences, and make booking decisions are Facebook, Instagram, and TripAdvisor. Social media is becoming more popular amongst tourism and hospitality industries in reaching their potential clients with accommodation being advertised directly via Facebook pages, groups and discussion forums. It has been found that more than three out of five travelers use social media networks when planning their trip, and they use user-created content, reviews, direct communication with accommodation providers (Bilan et al., 2024).

The change to digital-first booking has made tourism services more accessible to everyone, yet it has also brought about new vulnerabilities. Facebook groups and pages allow resort owners and small accommodation businesses to control their bookings, interact with the guests, and present their property. This direct-to-consumer strategy will lower any intermediary expenses but will eliminate the protective measures provided by the existing online travel agencies (OTAs).

Accessibility and sheer number of users make Facebook the platform of choice by hospitality business that aims to have direct contact with customers. Facebook pages related to tourism destinations, especially in developing markets, contain thousands of daily booking requests, advertisements, and confirmations to make a reservation. The low entry barrier and in-built communication functionality in the platform whereby the site includes messaging, photo sharing, and group administration features renders the site a perfect platform in which tourism transactions can occur (Mohammad et al., 2024).

Nevertheless, this ease of access is granted to fraud cases as well. The lack of formality of the transaction on Facebook, along with the lax verification, has led to the proliferation of fake reservation posts, fraudulent booking messages and scam messages. The research on Facebook tourism groups in major destinations like Amsterdam, Prague, and Istanbul shows that the security issue, such as fake profiles and unethical booking interventions, is widespread (Franjić & Farrell, 2025).

The issue of fake reservation posts is a complex issue on Facebook resort pages. Such fraudulent posts commonly include: (1) bogus booking receipts that promise to reserve accommodation without being available, (2) identity spoofing that deceives scammers posing as legitimate accommodation providers, (3) advance payment, where scammers request travelers to provide deposits on properties that do not exist, and (4) phishing posts that are meant to collect personal and financial details of travelers (Reid, 2024).

Studies show that they use advanced methods by scammers, such as GenAI-generated bogus accounts and doctored property photos, to gain trust. The COVID-19 pandemic spurred the valid online booking business and scams, and tourism scams are becoming more highly organized and organized (Gupta, 2025). More than half of the surveyed OTA users in India stated that they had changed their booking habits as a result of scam awareness, and almost 74% said that fake hotels had greatly diminished their trust in online services.

Purpose of the Study

This literature review synthesizes current research on fake reservation post detection on Facebook resort pages. The review aims to: (1) identify existing methodologies for detecting fraudulent content in tourism contexts, (2) examine theoretical frameworks applicable to deception detection in social media, (3) analyze machine learning and AI-driven approaches for fraud classification, and (4) highlight gaps in research specific to tourism booking fraud on social platforms.

II. Conceptual Background

Definition of Fake Posts and Fraudulent Content

Fake posts are defined as user-generated content that aims to mislead audiences by misrepresenting identity, facts, or purpose and are considered in the context of social media (Stewart and Arnold, 2022). In tourism booking scenarios in particular, fraudulent reservation posts are deceptive or untrue messages posted on social media which purport to provide accommodation, confirm reservation or request payment of accommodation services with no real intention of providing the service being promised.

Most fraudulent materials usually have a number of distinguishing features that define them as compared to genuine posts. These posts usually include partial or intentionally vague details on properties, use urgency tactics to compel travelers to make urgent payments, lack verifiable contact information outside of social media, and use stock photos or images of other properties, as well as often contain spelling or grammatical errors that would not be found in professional hospitality businesses. Also, fraudulent posts usually have unequal prices on various posts, do not have a clear cancellation policy, and seldom have detailed amenities or property management information. These posts exist on a continuum of deceitful purpose, with some being opportunistic scams on individual travelers and others organized ring fraud with groups of fake accounts (Boumber et al., 2024).

Overview of Facebook-Based Transactions

The informality of Facebook-based tourism transactions and the low intermediation levels are the key distinctions between them and traditional Online Travel Agency (OTA) bookings. The common flow of transactions starts with the discovery of the property, in which passengers may see the advertisements of the accommodation on the Facebook pages of the resorts, a group of tourism destinations, or the recommendations provided by members of the group. The inquiry stage is next, during which the interested travelers contact the property providers via Facebook in the messaging system to notify about availability, price, or booking. The next stage is negotiation where the travelers and the providers interact via informal means and usually negotiating prices, deposit rates and methods of payment without a set of standard terms and conditions.

The stage of payment is the most important part of Facebook transactions since travelers send money prior to receiving services. Facebook transactions may not offer any form of recourse once the payment is made contrary to OTA platforms that use escrow and dispute resolution policies. Lastly, the fulfillment stage entails the travelers coming to the property (or not coming to the property) and reviewing their experience. Major risks are generated in every stage but especially in the payment stage where travelers hand over money yet they have not received services. Lack of security features that have been embedded in established OTA platforms poses a significant risk to fraud in the entire informal transaction process.

Types of Reservation Fraud

Research on social media fraud identifies several categories particularly relevant to tourism bookings:

1. Identity Spoofing and Fake Profiles (Patil et al., 2024)

Fraudsters create replica accounts mimicking legitimate resorts or intermediaries. Machine learning studies demonstrate that detection via majority voting approaches achieves 99.12% accuracy in identifying such fake profiles by analyzing behavioral attributes.

2. Advance Payment Schemes (Franjić & Farrell, 2025)

Scammers request deposits or full payment claiming bookings are available, then disappear after funds are transferred. The analysis of Facebook accommodation groups across European cities revealed this as the most common fraud type.

3. Phishing and Data Harvesting (Abdulla & Thorig, 2025)

Fraudulent posts contain links or requests designed to collect personal information, credit card details, or travel documents. Social media impersonation and account hijacking represent significant vectors.

4. Bots and Automated Fraud (Rahman & Tomar, 2020)

Automated accounts generate high volumes of fake posts, reservations, or inquiries to manipulate reviews, inflate bookings, or distribute malware. Web forensic research identifies bot-based ticketing fraud as increasingly sophisticated.

5. Review and Rating Manipulation (Ansari & Gupta, 2021)

Fake reviews inflate resort ratings or deflate competitors' reputations. This category includes "shilling" (posting positive reviews for one's own business) and malicious reviews targeting competitors.

6. Group-Based Organized Fraud (Yu et al., 2025)

Coordinated networks of fraudsters operate together using multiple accounts, financial instruments, and communication channels to conduct systematic tourism fraud at scale.

Characteristics of Legitimate vs. Fake Reservation Posts

The legitimate posts usually reflect a number of peculiarities that indicate the signs of genuineness and professionalism. These posts also include extensive description of the property that include details about their facilities, types of rooms and their cost systems. They keep the information on various posts and platforms consistent, use a variety of high-quality photographs of real properties in various angles and features, provide clear contacts with verifiable business information, show transparent prices without any additional costs, express clear cancellation and refund policies, and promptly respond to the inquiries with

professional communication. Official suppliers tend to have well-established Facebook pages that have a track record, good reviews, and regular interaction habits that can be verified alongside credentials or appointments by the tourism departments.

False posts, in contrast, tend to have conflicting attributes that show bad faith. These posts contain unfinished or intentionally vague property descriptions without specific amenities or locations, incompatible information in various posts or websites, use general or stock photography that is often reused with other properties or travel websites, lack verifiable contact information other than social media channels, use pressure language of urgency (e.g., limited availability, book now) and often contain spelling or grammatical mistakes that would not be tolerated by a professional business. Studies in deception detection define the linguistic attributes that define truthful and deceptive text such as less use of pronouns because of social distance, lower sentence structure complexity, and increased use of affective language to control emotion (Stewart et al., 2025). The counterfeit posts are not associated with historical activity, do not contain any reviews that can be verified in the legitimate booking platforms, and in many cases, the posts vanish right after receiving the payment.

III. Theoretical Framework

Deception Theory in Online Communication

The Deception Detection Theory gives the basis of the interpretation of fraudulent communication in the online setting. This paradigm determines general patterns of how people speak when they lie (Boumber et al., 2024). The main postulates are the fact that in deceptive communication, there are relatively fewer personal references, less cognitive complexity, and more emotional language than in truthful communication. These tendencies arise due to the cognitive loading of deception the liars have to create false histories, track the response of the audience, withhold the truth, and be consistent with the past words.

According to Truth Default Theory, people have a natural tendency to believe that a communication is true unless they have reason to think otherwise. This exposes gaps in the detection of lying online because there is a low level of verification opportunities due to the lack of social cues (Luo et al., 2020). Findings on the credibility of fake news indicate that individuals have an average score of 51% detection of fake news on social media despite their subjective belief in their judgments, showing that they are systematically biased in detecting fake news. This happens due to the fact that people do not receive the nonverbal stimuli, which is facial expression, voice, body language, etc., that usually indicate one is being lied to during in-person communication.

Pronoun avoidance as a Behavioral Marker of psychological distance to falsehoods, more negation words (didn't, not, never), less specificity and concrete details, more hedging language (possibly, maybe), and self-defending with more self-references are Deception Behavioral Markers of a text-based communication. The lack of nonverbal avenues that limit the cues accessible to audiences, the fact that a digital record can be analyzed again and again, the relative anonymity that allows one to deceive without facing any social repercussions, and the lack of immediate feedback that forces a sender to act on the spur of the moment all make online-specific deception distinctly different than face-to-face deception.

Information Asymmetry Theory

Information Asymmetry refers to a scenario in which one side of a transaction has much more or better information compared to the other side, which leads to the possibility of fraud (Zhao et al., 2025). In the Facebook resort booking scenario, there is extreme asymmetry between the property providers and the travelers. Providers of property have significant benefits as they have full information about the state of property and real facilities it offers, know what is truly available and booked, and can control the flow of information and pictures they can provide, and can control how they can present descriptions to make it seem more appealing than it actually is.

Travelers are particularly disadvantaged, as they have little choice but to rely on information provided by the provider, having minimal opportunity to verify it independently, cannot physically inspect the properties in advance before payment, are prone to manipulated image or description, and have little means of action after the payment is completed. This asymmetry is further enhanced by the temporal dimension-travelers are required to spend money before receiving the service whereas providers are aware of real-life circumstances during the transaction.

Signaling mechanisms used as Information Asymmetry Solutions in legitimate e-commerce are: verified reviews by several independent parties, professional credentials and certifications, past performance information, and reputation systems. There is the addition of screening systems such as escrow services, which retain money until the services are delivered, third-party confirmation of property conditions, money-back guarantees and dispute resolution processes. Nevertheless, Facebook-based transactions do not have these mechanisms at all, which forms the perfect environment of fraud (Zhao et al., 2025). Institutional absence in Facebook diminishes the effects of frauds, heightens the impact of information asymmetry, and eliminates the shielding effects that would otherwise be present between credible and dishonest providers.

Trust and Credibility Models in Social Media

The development of trust in the context of social media is associated with the evaluation of credibility of the source and perceived reliability, and credibility studies have shown that there are numerous dimensions (Agarwal, 2025). Expertise Assessment is the first dimension, where the users consider the sources to have knowledge and skills applicable to what they offer. In the case of resort providers, expertise indicators are detailed knowledge of properties that is shown in the form of detailed

descriptions, local tourism and destination knowledge, good communication proficiency at work, and responsiveness to particular questions regarding services or amenities.

The second dimension is Trustworthiness Evaluation when users evaluate the honesty and reliability of sources. Some of the indicators of trustworthiness are the consistency of information sources in different platforms and posts, the transparency of price and cancellation policies, the responsiveness to customer issues and questions, and the lack of pressure to buy. Social Proof Effects have a strong impact on credibility, where users are increasingly using approval of other people, such as likes, comments, shares, and reviews, in assessing content (Luo et al., 2020). Nevertheless, this mechanism is compromised by manipulated social proof such as fake reviews and interacting with bots. This has been proven by research to cause high levels of Facebook likes to perceived credibility of real and fake news, which are counterintuitive since fake information gains credibility with the help of social amplification.

Platform Trust Spillover explains the relationship between the overall belief users have in Facebook as a platform and the impact of this belief on the evaluation of specific posts. Once a fraud has become rampant, the overall platform trust is damaged and it has a negative spillover effect on even the trustworthiness of legitimate providers (van der Meet et al., 2023). Such platform trust destruction is a notably subtle consequence of fraud, since fake information makes one lose trust in legitimate businesses at the same time.

Vendor Reputation Systems use multi-dimensional reputation models where sellers are rated based on several aspects such as the overall history of transaction completion, customer satisfaction rating, speed in responding to queries, dispute resolution, and reliability of service provision. In the tourism setting, in particular, the impact of online reviews studies indicate that authentic and detailed reviews have a much more pronounced impact on booking decisions than such factors as generic praise or generic positive reviews (Nisa, 2024). Nevertheless, manipulation of reviews like fake positive reviews of oneself, shill reviews, and strategically placed damaging competitor reviews constitute a significant credibility threat to these reputation mechanisms.

IV. Review of Related Literature

A. Fake Content Detection in Social Media

Fake News Detection

Detection of fake news in social media is a highly important use of machine learning and natural language processing algorithms. The definition of fake news is that of purposefully created or misleading information meant to be deceptive to audiences, though with a specific focus on intentional manipulation instead of unintentional misinformation (Suvidha et al., 2026). It has been shown that transformer-based architectures are far more effective in fake news detection than traditional classifiers, and that BERT-based models will achieve accuracy rates of up to 97.8 percent and a false positive rate approximately 15 percent lower than with conventional machine learning methods (Suvidha et al., 2026).

The most common methods of machine learning to detect fake news are based on supervised learning, which is trained on such benchmark datasets like FNC-1, LIAR, and WELFake (A, 2026). The way raw news articles are processed into numbers that can be used in their classification is through text preprocessing and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF) techniques (Chawre, 2026). Comparative research shows that RF classifiers with TF-IDF features gain about 91.3 percent accuracy, whereas hybrid methods that use multiple classifiers with ensemble learning techniques show high results on different datasets (Chawre, 2026).

The development of the classical machine learning methods into deep learning methods is manifested in the advancement of linguistic and semantic nuances of the deceptive content. Contextual embeddings with LSTM networks and Convolutional Neural Networks allow automatic feature extraction of complex patterns that are not simple frequency-based statistics (Tambakad et al., 2025). Recent developments make use of the transformer models such as BERT and RoBERTa which are trained on large corpora and can give contextualized representations of text, which are much better in detecting but can also give cross domain generalization (Site and Akhtar, 2025).

Spam Detection

The spamming on social media deals with the spread of unwanted, malicious or commercial content. Spam messages are unsolicited messages with misleading links, viruses, or other fake information that is sent in large scale (Alshattnawi et al., 2024). Social media spam detection studies reveal that contextualized embeddings, such as BERT and ELMo, are far more accurate than traditional word embeddings, and ELMo has 90 percent and 94 percent success rates in Twitter and YouTube data, respectively, when combined with logistic regression (Alshattnawi et al., 2024).

Machine learning-based spam detection methods use feature extraction methods such as TF-IDF, Bag-of-Words, Word2Vec embeddings, and classifiers such as Support Vector Machines and Random Forests (Erbay et al., 2024). The best performance is demonstrated by deep learning systems, especially Bidirectional LSTM networks on ALBERT contextualized embeddings, with a precision of 0.98 in Twitter, 0.96 in YouTube, and 0.98 in SMS data (Erbay et al., 2024). These techniques are useful to capture semantic relationships and other situational dependencies in brief informal texts typical of social media.

Review Fraud Detection

Reviews manipulation poses considerable risks to online commerce and tourism websites where false reviews artificially inflate the rating, deflate the reputation of competitors, or directly encourage buying products through misleading endorsements (Ansari and Gupta, 2021). Evidence on review fraud detection shows that authentic and detailed reviews have a stronger impact on booking decisions than generic praise, whereas review manipulation such as shilling and strategically placed damaging reviews negatively affects reputation mechanisms (Nisa, 2024).

Fraudulent reviews are detected using supervised learning and feature engineering that represents the review features such as the linguistic patterns, rating aberration, and the pattern of reviewer behavior (Boumber et al., 2024). The machine learning classifiers are also useful in distinguishing between fake and authentic reviews based on linguistic indicators, time patterns, and content coherence (Ansari and Gupta, 2021). Extreme class imbalance (often less than 1 percent of total activity is represented by fraudulent reviews) is a challenge that needs special strategies such as SMOTE (Synthetic Minority Over-sampling Technique) to balance the data and ensemble methods to detect it better (Ansari and Gupta, 2021).

B. Fraud in Online Reservations and Tourism

Booking Scams in Hospitality

Hospitality booking fraud is a significant weakness impacting both the honest companies and customers. Booking scams are also characterized by fake booking confirmation, identity spoofing, where fraudsters impersonate trusted accommodation providers, advance payment schemes, where fraudsters require deposit on non-existent accommodation, and phishing posts, which are aimed at gathering personal and financial data (Franjić and Farrell, 2025). A study of Facebook tourism communities in key destinations such as Amsterdam, Prague, and Istanbul shows that security concerns such as fake accounts and fraudulent booking habits are widespread, with more than half of the respondents surveyed by Online Travel Agency having indicated behavior change following their knowledge of scams in India (Gupta, 2025).

The most widespread forms of fraud in Facebook accommodation groups are advance payment scheme, whereby perpetrators demand deposits on the basis that they have bookings, and vanish after the money has been transferred over (Franjić and Farrell, 2025). The study of the Facebook accommodation groups in different cities in Europe depicts advanced methods of fraud such as the use of GenAI-generated fake accounts and the manipulation of property photos aimed at establishing credibility (Reid, 2024).

Social Media-Based Tourism Fraud

The informal nature of Facebook-based tourism transactions and the low degree of intermediation contrast with the traditional Online Travel Agency bookings (Franjić and Farrell, 2025). The common transaction cycle involves property discovery in which the travelers get to see advertisements of accommodation in Facebook resorts pages or on tourism destination groups, and proceed to the inquiry phases in which they are directly messaged to property providers (Franjić and Farrell, 2025). This unofficial market does not have protective laws such as escapes and dispute resolution systems and third-party validation as used in the more established OTA systems, which poses a significant risk of fraud.

Platform trust spillover refers to the effect of general trust on Facebook as a platform on individual posts; when fraud is rampant, general platform trust diminishes, and there exists negative spillover effects with even honest providers losing credibility (van der Meet et al., 2023). Studies have shown that a lack of institutional regulation on Facebook transactions lowers the cost of fraud, amplifies information asymmetry, and eliminates protective features that differentiate credible providers and those who are dishonest (Zhao et al., 2025).

C. Natural Language Processing (NLP) Techniques

Text Preprocessing

Text preprocessing is a fundamental step in NLP pipelines, which converts raw textual data into data formats that can be analyzed computationally. Normal preprocessing methods involve tokenization (separating text into words/phrases), case normalization (turning text to lower case), stopword elimination (removing words such as the, and, etc.), stemming/lemmatization (reducing words to root words) (A, 2026). The methods eliminate noise and retain semantic information used in downstream analysis.

Lemmatization is a more advanced method than stemming, where linguistic expertise is used to simplify words to their entries in dictionaries (Jin, 2025). As an example, running, runs and ran all are reduced to the base form run. This morphological examination is especially useful in identifying deceptive material when the words, which can have the same meaning, may have different surface forms (Jin, 2025).

Feature Extraction (TF-IDF, N-grams, Embeddings)

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical technique that values terms according to their frequency in documents and documents overall (Lozynska et al., 2025). TF-IDF weights represent both local and global importance (frequency in document and rarity in corpus, respectively), and generate numerical feature vectors which can be fed into machine learning algorithms (A, 2026).

N-grams are repetitive patterns in text that represent sequential dependencies and local patterns of n consecutive words or characters (Mahmoud and Ibrahim, 2025). Bigrams (two-word sequences) and trigrams (three-word sequences) retain contextual data that is lost in single-word representations, and can identify meaningful phrases and colloquial expressions (Mahmoud and Ibrahim, 2025).

Word embeddings Word2Vec, GloVe, and FastText embeddings are word representations as dense vectors in continuous semantic space where similar words are clustered (Tambakad, 2025). These embeddings encode semantic relationships and contextual sense (better than sparse representations such as TF-IDF) (Tambakad et al., 2025). BERT embeddings build upon this method by using bidirectional transformers that consider both prior and successive words and generate contextualised representations, which differ depending on the surrounding context (A, 2026).

Sentiment Analysis

Sentiment analysis (or opinion mining) is a textual mining technique that determines the presence of emotional tone and polarity in a text, categorizing it as positive, negative, or neutral (K et al., 2024). Sentiment analysis uses lexicon-based methods that use already prepared sentiment dictionaries and machine learning methods that are trained on labeled data (Mahmoud and Ibrahim, 2025). The methods based on lexicon assign sentiments scores depending on the presence of positive or negative words, whereas the methods based on machine learning learn subtle patterns between authentic sentiment and instances of manipulation.

It has been proven that fraudulent posts often use emotional wording intended to sway the reader, and the analysis of emotional elements will give indications to detect the fraud (Alagu Manohar and Hemamalini, 2025). The use of urgency cues (limited availability, book now), fear appeals, and manipulation of greed is more common in deceptive posts than legitimate ones (Boumber et al., 2024).

D. Machine Learning Approaches

Supervised Learning (SVM, Naïve Bayes, Logistic Regression)

Support Vector Machines (SVM) are very potent supervised learning algorithms that identify optimal decision boundaries between various classes in high-dimensional features space (A, 2026). The accuracy of SVM with TF-IDF features in detecting fake news is 98.4% with precision of 1.00 and F1-score of 0.98, which is a better result than traditional classifiers (Amandeep and Marwaha, 2025).

Naive Bayes classifiers use probabilistic logic, which is founded on the Bayes theorem, with the assumption of independence of features that makes calculating it simpler (A, 2026). Naive Bayes has surprisingly high performance in text classification with 96.2% accuracy in SMS spamming with TF-IDF features even with the unrealistic assumption of independence (Ahmadi et al., 2025).

Logistic Regression is a binary classification probability model that transforms linear combinations of features using sigmoid, which gives interpretable probability estimates (A, 2026). TF-IDF Logistic Regression Logistic Regression is a successful model based on TF-IDF features, usually used as a powerful baseline model against which more complicated models are evaluated (Jin, 2025).

Deep Learning (LSTM, BERT)

Long Short-Term Memory (LSTM) networks solve the vanishing gradient issue of the traditional recurrent neural networks and enjoy the ability to learn long-range dependencies in sequential data (Tambakad et al., 2025). The memory cells and gates that control information flow allow LSTM networks to have an internal state, which makes them especially appropriate to learn the temporal patterns and contextual dependencies in text (Tambakad et al., 2025). Networks that feed sequences in both directions to obtain information in both directions recognize context based on adjacent words, which have a 97.3 percent accuracy in detecting spam in SMS (Ginson et al., 2025).

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based pre-trained language model that offers contextualized representations in the form of embeddings that are vastly superior to independent word vectors (Suvidha et al., 2026). The bidirectional training of the BERT model allows the model to consider both the context of the words before and after it, and the pre-trained models experience an impressive accuracy boost on both the fine-tuning of the model on a specific task (Jenifer and Janit, 2025). BERT models detect malicious email with 98.3 per cent accuracy, which is the highest among all the models in contextual accuracy (Jenifer and Janit, 2025).

Hybrid Models

Hybrid methods are methods that use the benefits of several architectures and learning paradigms in order to overcome the shortcomings of each approach. The CNN-LSTM hybrid models combine convolutional layers to extract local spatial patterns with LSTM layers to extract long-range dependencies with an accuracy of 95.70% in the detection of fake news, more accurate than the individual models because they successfully capture both local and long-range text patterns (Chatur and Jadhav, 2025). Voting or stacking ensemble techniques that incorporate the predictions of several classifiers are further improved in enhancing performance, and hybrid techniques are always better than single classifiers (Amandeep & Marwaha, 2025).

VI. Research Gaps

Lack of Studies Focused on Facebook Resort Reservations

There is also a very crucial gap in the academic research that specifically deals with the fake reservation post detection on Facebook resort pages. Although there has been a lot of literature on general social media fraud, credit card fraud, and phishing detection, the intersection between Facebook-based accommodation booking fraud has not been extensively studied. The majority of the research on tourism fraud is based on the premise that institutional resources to prevent fraud are present in most established Online Travel Agency (OTA) websites such as Airbnb and Booking.com (Franjić and Farrell, 2025). Nevertheless, direct-to-consumer Facebook dealings are not subject to scholarly research, even though they are spreading in the developing markets.

The distinctive nature of Facebook resort booking fraud (informal negotiations, uncheck seller accounts, group transactions, cultural context-specific scam patterns, etc.) is not sufficiently covered by the existing fraud detection systems. The credit card fraud-trained detection systems might not be transferred successfully to Facebook posts with property descriptions, images, and informal text patterns that are typical of tourism booking (Reid, 2024). The linguistic patterns of tourism booking fraud vary considerably with phishing emails or payment fraud and demand specialized training data that cannot be found in the public repositories.

Limited Datasets Specific to Tourism Scams

Tourism and accommodation fraud data is in dire need of public datasets. Although there are millions of credit cards transactions in repositories such as Kaggle, there are limited tourism fraud datasets. The Facebook accommodation group data collected in European cities is one of the little published tourism fraud datasets (Franjić and Farrell, 2025) and even they are anonymized and small in size. There is practically no literature on Philippine tourism fraud dataset although the Philippines has one of the most vibrant Facebook tourism communities in Southeast Asia.

The limitations on privacy and platform significantly inhibit the collection of data. The terms of service in Facebook limit data scraping, which complicates the researchers in gathering real examples of booking scam on a large scale (Mohammad et al., 2024). The unwillingness of the resort owners to report fraud cases because of the reputational risk does not allow real-world scam datasets to be accrued. Artificial datasets created to address the problem of scarcity, might not reflect real-world fraudster methods so the model that works well on artificial data is not going to work effectively in the real world.

Tourism faces unique challenges in terms of class imbalance that jeopardize the limitations of datasets. While credit card fraud (0.17% fraud rate) involves 0.17% fraud posts, Facebook tourism groups can contain 5-15% fraudulent postings, posing other technical challenges (Franjić and Farrell, 2025). Current resampling methods can also over-randomize unrealistic tourist booking patterns and decrease the authenticity of the data.

VII. Implications for Future Research

Development of Detection Systems for Resorts

Detection Systems Detection Systems Resort development.

Research in the future should also produce fraud detection mechanisms that are specific to resort operations and not adherence to financial fraud detection methods. Such systems must install resort special features such as consistency of booking calendar, verification of property amenity claims and reservation confirmation procedures (Franjić and Farrell, 2025).

The multi-modal detection methods based on text analysis of posts, image verification, and analysis of behavioral patterns are a promising area of research. Property photos may be checked with known databases of accommodation by image-based fraud detection, and studies of GenAI-generated images (Reid, 2024) allow finding deepfakes usually involved in scams in the tourism industry. NLP should be used together with computer vision methods to detect discrepancies in textual descriptions and images.

The implementation of lightweight deployment architectures should be compatible with the small and medium-sized resorts having a limited IT infrastructure. The existing fraud detection systems need advanced cloud infrastructure (Arzu et al., 2025), yet edge-based detection with basic resort computers or mobile devices would allow more extensive implementation. The models of progressive complexity beginning with simple detection that is rule-based and moving on to machine learning may offer cost-effective solutions to resource-constrained accommodations.

The integration of the resort management system must allow automatic detection without an understanding of technical skills. The ability to directly integrate with Facebook Business Suite would enable real-time analysis of all incoming booking requests and flag suspicious conversations automatically. Non-technical resort employees require user-friendly dashboards with the fraud risk scores and advice on what to do.

Integration with Facebook Pages or Chat Systems

The API-based integration with the official Facebook channels is essential infrastructure to be deployed in practice. Instead of external controls, being part of the Facebook system via official partnerships with Meta would allow:

Automated verification chatbots may be more effective at detection with probing questions disclosing discrepancies in the tales of scammers. Response patterns to verification questions can be analyzed behaviorally to detect the presence of a fraudster who tries to keep direct communication to the minimum. Conversational AI that is specifically trained to identify lies in tourism booking conversations should be developed through research.

The Facebook tourism groups would be able to automatically identify and delete fraudulent content through group administrator tools. The algorithmic detection may be supplemented by community-based detection systems based on the reports of group members and their collective knowledge. Future studies ought to investigate the idea of crowdsourced fraud annotation whereby group members are invited to train models by reporting suspicious posts and privacy-preserving methods must protect personal data.

Facebook could introduce two-sided marketplace defense systems that were previously in place. This would have the effect of removing the vulnerability of advance payment that is the hallmark of Facebook booking scams through the use of escrow-like functionality to hold deposits until the customer has confirmed (Zhao et al., 2025). The studies are to examine the incorporation of blockchain to generate transparent and tamper-free records of transactions (Arzu et al., 2025).

Use of AI for Real-Time Filtering

In adaptive real-time filtering, there should be a balance between preventing fraud and the user experience. Filtering that is aggressive (and removes any possible fraud posts) may damage the legitimate businesses, whereas filtering that is lenient (and does not stop fraudulent businesses) is not effective. The study of threshold optimization must find optimal decision limits of various types of resorts and tourist populations.

User trust and regulatory compliance require explainable AI (XAI) techniques. Interpretability models need to describe why certain posts are flagged, so that human reviewers can overturn automated decisions. AI decision-making transparency creates trust in users and enables them to improve with feedback.

Through federated learning methods, it might be possible to perform joint fraud detection across resorts and tourism organizations without centralizing sensitive information. Training distributed models in which local training is done in individual resorts and only insights (not data) is exchanged preserves privacy yet allows global patterns to be identified.

The current challenges are concept drift and adversarial adaptation. The persistent model monitoring and retraining should observe when fraudster tricks go beyond the training data (Bello et al., 2024). Adversarial robustness testing must test the system against the intentional evasion of highly advanced scammers.

The Filipino tourism is dependent on multi-language support. Detection systems should include code-mixing management of Tagalog-English postings, variation of accent and dialect in written Filipino, and references to culturally specific aspects of Philippine tourism. The possibility of transfer learning between high-resource languages should be explored, and the presence of negative transfer should be considered (Cruz & Miguel, 2024).

Policy and Security Recommendations for Businesses

Business Policy and Security Recommendations.

Tourism authorities and government agencies should develop regulatory frameworks of accommodation booking based on social media. Verification systems that are mandatory, where the resort has to be registered with tourism boards, property verification to ensure it exists and credentials to be displayed on Facebook pages would minimize the prevalence of fraudsters (Franjić & Farrell, 2025). Toughness in the penalties of posting fake accommodations would discourage fraud but there are difficulties in enforcing.

Travelers should be educated through consumer education campaigns to be aware of typical tourism booking fraud, such as:

Facebook resort pages should be verified with the following standards:

The policy of strengthening Facebook platform should involve:

The best practices in resort management need to be shared at tourism associations:

Ecosystem resilience would be enhanced through cross-institutional cooperation between resorts, tourism boards, and technology platforms. Tourism association networks can be used to share information on emerging scam tactics to coordinate a response (Yu et al., 2025). Research collaborations between resorts and academic institutions would help speed up the creation of useful detection systems.

VIII. Conclusion

This literature review reveals the increasing danger of counterfeit reservation articles on Facebook resort pages and that such risk has a major effect on the tourism, consumer confidence, and business reputation. The overview of existing literature proves that, although the most modern machine learning and natural language processing algorithms, specifically ensemble models and transformer-based ones, like BERT, are highly accurate in identifying fraudulent materials, they have not been utilized in tourism-specific settings yet. The review also highlights that the linguistic, behavioral, and structural peculiarities of deceptive reservation posts can be efficiently examined with the help of AI-based methods. Nevertheless, there are also critical gaps such as the absence of Facebook-specific studies, lack of tourism related data and little studies in the Philippine context despite its high use of social media in conducting tourism transactions.

The results indicate that the next line of research should be to create real-time platform integrated detection systems customized to resort operations including multi-modal analysis and multi-language support. It is also necessary to strengthen the collaboration between academic institutions, stakeholders in the industry, and social media to tackle this problem in its entirety.

In general, this research offers a basis of future research and the creation of applicable solutions to increase security, safeguard consumers, and maintain confidence in social media-based tourism ecosystems.

References

Abdulla, A., & Thorig, I. (2025). Machine learning approaches for phishing detection: A systematic literature review. In *2025 2nd International Conference on Innovations in Engineering, Science and Technology for Sustainable Development (ICEST)* (pp. 1–6). https://doi.org/10.1109/ICEST65883.2025.11428773

Agarwal, V. (2025). Identifying how framing style, source credibility, scientific referencing, and social media popularity influence adolescents' trust and behavioral intention. *International Journal for Multidisciplinary Research, 7*(5). https://doi.org/10.36948/ijfmr.2025.v07i05.58977

Ahmadi, M., Khajavi, M., Varmaghani, A., Ala, A., Danesh, K., & Javaheri, D. (2025). Leveraging large language models for cybersecurity: Enhancing SMS spam detection with robust and context-aware text classification. *Cyber-Physical Systems.* https://doi.org/10.48550/arXiv.2502.11014

Alagu Manohar, P., & Hemamalini, R. (2025). Emotion analysis for scam detection in social media and communication platforms: Using machine learning. *International Journal of Scientific Research in Science and Technology, 12*(4), 844–849. https://doi.org/10.32628/IJSRST251364

Amandeep, & Marwaha, M. (2025). A hybrid machine learning approach for fake news detection: Combining TF-IDF, CountVectorizer, and support vector machine. In *2025 OITS International Conference on Information Technology (OCIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/OCIT66168.2025.11400373

Ansari, S., & Gupta, S. (2021). Review manipulation: Literature review and future research agenda. *Pacific Asia Journal of the Association for Information Systems, 13*(1), Article 4. https://doi.org/10.17705/1pais.13104

Arzu, F., Bajwa, M. K. Z., Obaidullah, Waheed, A., Alam, F., Ali, M., & Khan, A. (2025). Real-time financial fraud detection: An intelligent data-driven framework integrating machine learning, stream processing, and big data analytics for high-velocity transaction monitoring. *The Asian Bulletin of Big Data Management, 5*(4), 124–154. https://doi.org/10.62019/f3c0g313

Bello, H. O., Ige, A. B., & Ameyaw, M. N. (2024). Adaptive machine learning models: Concepts for real-time financial fraud prevention in dynamic environments. *World Journal of Advanced Engineering Technology and Sciences, 12*(2), 21–34. https://doi.org/10.30574/wjaets.2024.12.2.0266

Bilan, Y., Tovmasyan, G., & Dallakyan, S. (2024). The Impact of Digital Technologies on Tourists' Travel Choices and Overall Experience. *Journal of Tourism and Services, 15*(29), 153-175. https://doi.org/10.29036/jots.v15i29.805

Boumber, D., Verma, R. M., & Qachfar, F. Z. (2024). A roadmap for multilingual, multimodal domain-independent deception detection. *arXiv*. https://doi.org/10.48550/arXiv.2405.03920

Chatur, B., & Jadhav, K. (2025). Detecting fake news using deep learning and text classification. *International Scientific Journal of Engineering and Management, 4*(10). https://doi.org/10.55041/isjem05100

Chawre, S. (2026). Fake news detection using machine learning techniques. *Advanced International Journal for Research, 7*(1). https://doi.org/10.63363/aijfr.2026.v07i01.3125

Cruz, M. E., & Miguel, D. (2024). Cross-Domain Transfer Learning: Enhancing Deep Neural Networks for Low-Resource Environments. *ITEJ (Information Technology Engineering Journals), 9*(2), 72 - 79. https://doi.org/10.24235/itej.v9i2.136

Franjić, M., & Pavlakovič Farrell, B. (2025). (Un)secure booking: Security risks within Facebook groups accommodation reservations. In P. Šprajc, D. Maletič, N. Petrović, I. Podbregar, A. Škraba, D. Tomić, A. Žnidaršič, & Ž. Rant (Eds.), *International Conference on Organizational Science Development: Human being, artificial intelligence and organization, conference proceedings* (Vol. 44, pp. 221–234). University of Maribor Press. https://doi.org/10.18690/um.fov.2.2025.19

Ginson, Y., Edwin, E. B., Ebenezer, V., Kirubakaran, S. S., Thanka, M. R., & Raja, M. M. (2025). Neural sequence modelling for spam classification via bidirectional LSTM and hierarchical neural networks: A deep learning approach. In *Proceedings of the International Conference on Smart Technologies, Communication and Robotics* (pp. 1–6). https://doi.org/10.1109/STCR62650.2025.11019808

Gupta, M. (2025). Challenges faced by emerging OTA industry. *International Journal for Research Trends and Innovation, 10*(5), a221–a241. https://doi.org/10.56975/ijrti.v10i5.203342

Jenifer, D. A., & Janit, S. (2025). Email forensics using machine and deep learning techniques for cybercrime detection. In *Proceedings of the International Conference on Intelligent and Secure Engineering Solutions* (pp. 1248–1252). https://doi.org/10.1109/CISES66934.2025.11265002

Jin, Y. (2025). Evaluating TF-IDF and Word2Vec in multilingual sentiment analysis: A controlled empirical approach. In *Proceedings of the International Academic Exchange Conference on Science and Technology Innovation* (pp. 100–106). https://doi.org/10.1109/IAECST68792.2025.11414823

Lozynska, O., Vysotska, V. A., Markiv, O., & Kuspis, M. (2025). Method for detection of disinformation based on text data analysis using TF-IDF and contextual vector representations. *SISN, 18*(1), 98–110. https://doi.org/10.23939/sisn2025.18.1.098.

Luo, M., Hancock, J. T., & Markowitz, D. M. (2020). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research, 49*(2). https://doi.org/10.1177/0093650220921321

Mahmoud, H. A., & Ibrahim, I. M. (2025). Fake News Detection Using Natural Language Processing (NLP) and Machine Learning. *Engineering And Technology Journal, 10*(5), 5177–5185. https://doi.org/10.47191/etj/v10i05.46

Mohammad, A. A. A., Elshaer, I. A., Azazz, A. M. S., Kooli, C., Algezawy, M., & Fayyad, S. (2024). The Influence of Social Commerce Dynamics on Sustainable Hotel Brand Image, Customer Engagement, and Booking Intentions. *Sustainability, 16*(14), 6050. https://doi.org/10.3390/su16146050

Nisa, B. U. (2024). Impact of online reviews on hotel booking decision. *Minhaj International Journal of Economics and Organization Sciences, 4*(2), 38–53. https://doi.org/10.58932/MULE0033

Patil, D. R., Pattewar, T. M., Punjabi, V. D., & Pardeshi, S. M. (2024). Detecting fake social media profiles using the majority voting approach. *EAI Endorsed Transactions on Scalable Information Systems, 11*(3). https://publications.eai.eu/index.php/sis/article/view/4264

Rahman, R. U., & Tomar, D. S. (2020). A new web forensic framework for bot crime investigation. *Forensic Science International: Digital Investigation, 33*, 300943. https://doi.org/10.1016/j.fsidi.2020.300943

Reid, J. (2024). Risks of generative artificial intelligence (GenAI)-assisted scams on online sharing-economy platforms. *The African Journal of Information and Communication (AJIC)*, (33), 1–21. https://doi.org/10.23962/ajic.i33.18162

Stewart, M. C., & Arnold, C. L. (2022). Bridging the gap for online deception detection: Uncovering methodology to identify deceptive content in mediated communication. In *Proceedings of the International Crisis and Risk Communication Conference* (Vol. 5, pp. 42–44). Nicholson School of Communication and Media. https://doi.org/10.30658/icrcc.2022.10

Stewart, M. C., Arnold, C. L., & Richard, F. D. (2025). Truth & lies on social media during COVID-19: Applying mediated statement analysis (MSA) for digital deception detection. *International Crisis and Risk Communication Association Reports, 13*(1), 159–162. https://doi.org/10.69931/001c.142918

Suvidha, S., Bhatt, G., & Raiwani, Y. P. (2026). Fake news detection using stance analysis. *Journal of Mountain Research.* https://doi.org/10.51220/jmr.v21-s2.18

Tambakad, M. B., Rai, N. P., N. S., & Shetty, N. S. (2025). Combating misinformation with NLP: Advances in fake news detection. In *Proceedings of the International Conference on Electronics, Computing, Communication and Control Technology* (pp. 1–7). https://doi.org/10.1109/ICECCC65144.2025.11063924

Yu, J., Wang, H., Wang, X., Li, Z., Qin, L., Zhang, W., Liao, J., Zhang, Y., & Yang, B. (2025). Temporal insights for group-based fraud detection on e-commerce platforms. *IEEE Transactions on Knowledge and Data Engineering, 37*(2), 951–965. https://doi.org/10.1109/TKDE.2024.3485127

van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies, 24*(6), 803–823. https://doi.org/10.1080/1461670X.2023.2187652

Zhao, M., Wang, S., & Xia, T. (2025). The social welfare effect of e-commerce product reputation information asymmetry from the perspective of network externality. *PLoS ONE, 20*(1), e0313852. https://doi.org/10.1371/journal.pone.0313852

How to Cite this Chapter (APA 7):

Garcia, C. V. (2026). Fake reservation post detection on Facebook resort pages: A literature review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 86–95). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 12: Artificial Intelligence in Early Childhood Education: Foundations, Applications, and Pedagogical Implications

Dr. Rosemarie N. Pragacha
Cavite State University - Silang Campus

Abstract. *Artificial intelligence (AI) is rapidly transforming early childhood education, offering unprecedented opportunities for personalized learning, early identification of developmental concerns, and enhanced teacher support. This literature review synthesizes current research on AI applications in early learning environments for children aged 3-8, examining adaptive learning systems, natural language processing for literacy support, educational games, and AI-driven analytics tools. Evidence demonstrates that AI-enhanced personalized learning improves academic outcomes and engagement, particularly for learners with diverse needs, while supporting early detection of developmental delays with high accuracy. However, substantial challenges persist such as excessive screen time, data privacy concerns, algorithmic bias, and potential displacement of play-based, social-experiential learning. Teachers' competencies, preparedness, and professional autonomy emerge as critical success factors, with research emphasizing that AI must augment rather than replace human expertise and relational values. The review identifies future directions including emerging computer vision applications, hybrid intelligence models combining AI with traditional pedagogy, and comprehensive policy frameworks. A balanced, human-centered approach grounded in developmental theory, ethical safeguards, and equitable resource allocation is essential for responsible AI integration. This synthesis provides educators, researchers, and policymakers with evidence-based guidance for leveraging AI's genuine benefits while protecting the developmental, relational, and humanistic foundations central to quality early childhood education.*

Keywords: *Artificial Intelligence, Early Childhood Education, Personalized Learning and Adaptive Systems, Developmental Assessment*

1. Introduction

The development of artificial intelligence (AI) technology is essentially transforming the nature of educational ecosystems across the globe, with early childhood education (ECE) becoming one of the most crucial areas of innovation and thorough thought (Cimino et al., 2025). Being the pillar of lifelong learning, early childhood education is a very essential developmental phase that creates cognitive, social, and emotional basis towards future success of children (Zhou, 2025). The development of AI into this stage of foundational education is a unique opportunity but at the same time, new concerns emerge regarding the use, ethical issues, and the suitability of its application in pedagogy.

The AI environment of early childhood education has been changing considerably, and modern uses of AI exist across various technological platforms and methods of teaching. Intelligent educational robots and adaptive learning systems, virtual reality environments, and generative text-based tools are only part of AI technologies that provide new ways to personalize learning and other opportunities to support the needs of children in their developmental stages. It has been shown that AI-driven personalized learning systems lead to greater engagement and flexibility, whereas robotics and gamification develop critical problem-solving and collaboration abilities (Cimino et al., 2025). Moreover, AI solutions have been specifically promising in the case of supporting disabled children, as they are more likely to enhance inclusivity and accessibility than traditional solutions can.

Nevertheless, the introduction of AI into early childhood is an issue that should be addressed carefully and with a degree of tactfulness and caution. The major ethical issues are related to such aspects as the privacy of data, the bias of algorithms, the lack of teacher readiness, and the possible influence of technologies on the social-emotional development of children. Studies have found out that although AI promotes individualized, inclusive, and culturally responsive learning, there are still major issues such as lack of AI literacy among educators, inequality in accessing digital learning materials, and privacy and algorithmic bias (Noriana & Bakar, 2025). Also, the lack of emotional and developmental responsiveness to young children in conversational AI, the so-called empathy gap, poses specific threats to early learners whose language is often playful and full of non-literary expressions, and their speech sounds in unconventional ways (Kurian, 2023).

The principles of responsible AI integration in ECE require theoretical foundations and pedagogical premises. With reference to the existing developmental theories including Vygotsky Sociocultural Theory, Piaget theory of active discovery, and Bronfenbrenner ecological systems theory, scholars have started stating how AI-supported learning systems can support the natural developmental processes of children, but they will not be substituted (Ozturk, 2025). The evolution of humanistic, child-centered pedagogical frameworks, including the offered AI-Enhanced Early Learning (AI-EL) model, focuses on closing the loop of dynamic feedback at the intersection of technological affordances and development milestones, covering lapses in the conventional constructivism frameworks.

The role of educators becomes more central in this case. The perceptions, competencies, and readiness of teachers have a significant effect on the success or failure of AI integration initiatives (Qayyum et al., 2025) . The studies in a wide range of global perspectives demonstrate that teachers acknowledge the possibility of AI to organize the engagement and individual learning; however, they also make complaints regarding infrastructure constraints, training, and the necessity to retain the human factor as the fundamental part of the early education. Implementation can thus not be accomplished through a simple technological

adoption but a thorough investment in teacher professional growth, strong ethical systems, and structural provisions such as infrastructure, regulation and resources.

This paper discusses the principles of artificial intelligence, its use and implication on the pedagogy of early childhood education. Combining the existing research and theoretical approaches, we investigate how AI technologies can contribute to the proper and ethical development of young learners and define the main challenges and business considerations in this context. The emphasis is made on the promotion of knowledge of the integration of AI as a human-focused activity that should maintain the developmental adequacy, the importance of relationships, and the necessity of educators in the enhancement of holistic child development.

Scope and objectives of the review

This paper will seek to extensively discuss the importance of artificial intelligence in early childhood learning by undertaking systematic research on the topic based on the existing studies, theories, and practice. In particular, the review covers four main aims, which are (1) to chart the conceptual base/definitions of AI in early learning contexts, (2) to trace the historical progression of educational technologies that have culminated to the modern application of AI with the young children, (3) to synthesize theoretical conceptions linking AI systems to the established developmental psychology and learning theories, (4) and to critically analyze the pedagogical implications of AI integration to the holistic development of young children. The scope will include the publication of articles published in 2020 and onward, both empirical and theoretical literature, and policy documents on AI in the early childhood context (ages 3-8) and keeps the view of human-centered and developmentally appropriate approaches to AI, with a prior concern of the role of educators and the well-being of children.

2. Foundations of AI in Education

Overview of AI concepts relevant to teaching and learning

Artificial Intelligence is a collection of various technologies that essentially transform educational processes of all levels, but the underlying ideas should be clearly differentiated to be successfully applied to the situation of early learning. The most common type of AI in education is machine learning that allows systems to learn based on the patterns in these data and improve themselves without any direct programming by humans (Pinska et al., 2025). Natural language processing (NLP) is the technology behind conversational artificial intelligence applications and autopiloting learning systems that understand the language input of children and respond accordingly in the right context. Adaptive learning platforms make use of algorithm to tailor the content delivery according to the performance and preferences of learners and generate differentiated learning tracks (Akintola et al., 2025).

Applications in the field of early childhood education, in particular, can be categorized into three types: personalized learning systems where difficulty and content are replenished in real-time, educational robotics where systems can offer interactive peer-like companions to develop language and cognition, and intelligent tutoring systems where scaffolding and feedback are provided on-demand. More importantly, the empathy gap of conversational AI, where systems are not adept at decrypting children conversational play and non-literary language and emotional expressions, is a major pedagogical issue, especially in relation to young students, whose speech patterns do not fit into an adult language structure (Kurian, 2023). These basic notions are critical to helping teachers and policy-makers assess AI tools based on developmental norms as well as to make sure that technology supports pedagogical purposes instead of establishing new needs and obstacles.

Historical development of educational technologies

Educational technology direction illuminates the way that AI can be integrated into a larger development of pedagogy. In the 1970s, early computer based instruction developed, like the Logo programming language at the AI laboratory of MIT, the concept of technology as tutor, tool and tutee- roles which are still applicable today (Bull et al., 2025). The emergence of the Internet in the 1990s-2000s availed the deep digital data necessitating machine learning to the set of isolated applications to the set of connected and data-driven systems.

This evolution has been characterized by unique technological waves, which have improved on the previous innovations and widened the possibilities of personalization (Thompson & Harris, 2025). Starting with audiovisual media and computer-assisted instruction and then to interactive platforms, and adaptive learning systems, each step expanded the teaching ability. The COVID-19 pandemic hastened the use, revealing the advantages and shortcomings of technology-mediated learning (Jarodzka, 2025).

Generative AI symbolizes a qualitative leap in which machines not only store or transmit data, they now generate it as well. This presents emerging possibilities and also brings concerns of authenticity, prejudice, and human agency (Bull, 2025). Historically, AI is not a disruption but it is a continuum of the efforts to personalize and support learning. Its integration must be done in a manner that it is in agreement with the pedagogical wisdom, so technology does not kill but augments the established practices.

Theoretical frameworks linking AI to child development

The application of AI in early childhood learning should be based on developmental and learning theories. According to the Sociocultural Theory of Vygotsky, education is a process of social interaction with more competent individuals, and technology is one of the cultural means of mediating the development (Ozturk, 2025). AI must thus become a scaffold in the areas of proximal

development, an aid and not a substitute to human interaction. The Theory of Active Discovery by Piaget demonstrates the significance of practical action and construction of knowledge based on the concrete experiences (Ozturk, 2025). This implies that AI must encourage interactive discovery, multimodal feedback and allow children to experiment with their hypotheses as opposed to passive consumption of material.

The Ecological Systems Theory by Bronfenbrenner provides an extension to this point, saying that child development occurs within nested contexts, i.e. microsystem, mesosystem, exosystem, and macrosystem (Ozturk, 2025). The implementation of AI should thus be reviewed in the context of educational ecosystems, in terms of teacher capacity, family settings, institutional policies, and cultural values. New developments introduce child-centered models, including the AI-Enhanced Early Learning (AI-EL) model, where loops of dynamic feedback between technological affordances and developmental milestones are introduced (Tao & Narsi, 2025). This model maintains relational, play-based and teacher directed pedagogy but formalizes the role of AI. All these theories indicate the importance of ensuring that the use of AI is harmonized with developmental processes, maintained active learning, and surroundings, in larger ecosystems (Kamalov et al., 2023).

3. Applications in Early Childhood Education

The use of AI technologies in various early childhood settings is evidenced by the practical examples of their application to the individualized learning process and promotion of developmental requirements. It is essential to comprehend these applications and how it fits into the developmental theory in order to assess the role of AI in the initial stages of learning. This section explores four major areas of application, including adaptive learning systems, language and literacy support, educational games and apps, and teacher support tools, all based on evidence-based research and pedagogical aspects.

Adaptive Learning Systems: Personalized instruction tailored to developmental stages

Adaptive learning platforms are one of the oldest forms of AI use in early childhood education, where algorithms and machine learning are used to personalize the content being taught and the speed at which it is taught to each developmental profile. Such systems receive real-time information about the performance, learning behaviors and preferences of children, and dynamically change the level of difficulty, sequencing of content, and feedback to promote a best cognitive interaction (Oshinowo et al., 2025). The adaptive systems are effective basing on their ability to match and equate to the zones of proximal development of the children, as one cornerstone of the sociocultural theory by Vygotsky, through the provision of scaffolding that is neither overly complex nor over-simplistic.

Studies show that adaptive learning platforms that are powered by AI improve engagement and aid in retention because they enable children to learn at personalized speeds and get personalized feedback in real time (Sari et al., 2025). As an example, applications such as DreamBox and adaptive mathematics platform have demonstrated a great enhancement in the problem-solving skills and mathematical comprehension in children compared to traditional teaching, especially in the case of learners with diverse learning styles (Al-Hassan et al,, 2025). The responsive environments triggered by connecting technological affordances with developmental milestones, as envisioned in the AI-Enhanced Early Learning (AI-EL) model, lead to the integration of dynamic feedback loops with changing competencies in children (Tao & Narsi, 2025).

Nevertheless, adaptive systems should be adopted with keen consideration of cognitive and developmental suitability. According to research, adaptive systems designed to follow specific learning goals, progressive challenges, and multimodal representations can be effective in supporting early mathematics, literacy and conceptual learning without overwhelming young learners (Sari et al., 2025). One feature of critical design is to make sure that adaptive algorithms have a stronger emphasis on active child engagement, active exploration, and building of understanding as opposed to passive consumption of content-concepts based on the theory of active discovery of Piaget.

Language and Literacy Support: NLP tools for speech recognition, vocabulary building, and reading readiness

NLP technologies allow advanced support of early language development in a variety of modalities: speech recognition, conversational interaction, and adaptive vocabulary instruction. Tools that are powered by NLP process oral and written language input of children, identify speech patterns and create developmentally suitable replies that afford language acquisition (Hang & Ho, 2025). Such systems can give real-time responses about pronunciation, grammar, and vocabulary, aiding children to develop the basic literacy skills.

Practical uses of NLP in early childhood are AI-enabled storytelling apps that respond to language proficiency levels of children, speech recognition applications that assist speech-delayed children, and conversational AI to coerce children into two-way conversation to stimulate language generation (Azhar et al., 2025). Studies of AI-aided literacy solutions prove that appropriate tools with speech recognition and visual aids could enhance vocabulary development, phonological awareness, and early reading (Purwandari & Nasution, 2025). As an illustration, the spoken input recognition coupled with dynamically generated visual information systems have proven to be useful in assisting native, as well as non-native English learners.

Regular assessments on critical issues is important as far as NLP in early childhood is concerned. The empathy gap in conversational AI, or the low capacity of systems to understand the patterns of non-literal language and emotional expressions of young children, is an important aspect that needs to be considered in its design and execution (Kurian, 2023). Young children tend to speak in very imaginative and not traditional way; they make open ended questions, use metaphor and communicate more

with gesture than words. The AI systems that are trained on the patterns of speech as directed by adults might not respond to this level of developmental speech appropriately. This means that an effective tool of NLP in early childhood would have to include direct training on child language patterns of development, would have to combine multiple channels of communication as opposed to speech only, and cannot be human-free to allow responsive and emotionally sensitive communication.

Educational Games and Apps: Gamified learning for creativity, problem-solving, and engagement

Early childhood learning based on games capitalizes on the play instinct of children and applies AI to adjust the level of challenge, offer formative feedback, and customize the learning content to suit specific needs. Educational games use AI algorithms to monitor performance, change the challenge level, and incorporate playful mechanics promoting motivation and engagement (Tsao et al, 2025). It has been found out that problem-solving, creativity, attention, and conceptual understandings in various subjects are improved using well-designed gamified experiences in math, language, and science (Al-Hassan et al., 2025).

Gamification with AI has a number of pedagogical benefits. Adaptive difficulty makes games difficult and not overly hard, making the process of playing them thought-provoking but not frustrating (Arvanitaki & Skoumpourdi, 2025). Learning gaps are detected through real-time analytics and scaffolding is changed accordingly (Tsao et al., 2025). The multimodal characteristics, graphics, sound, movement and narrative offer that rich sensory involvement to learners across various learning styles. Some AI-driven games have collaborative aspects as well, which facilitate peer learning, turn-taking, and social-emotional development.

There is a body of research that grounds the positive effects of game-based learning on cognitive and social development (Zuceti et al., 2024). Reported effects of meta-analyses are moderate to huge impacts on problem-solving, logical thinking, and creative expression (Arvanitaki & Skoumpourdi, 2025). To successfully implement it, it is necessary to have a balance between directed instruction and unrestricted exploration, screen time, and physical play (Hibana et al., 2024). Optimal use of the applicable educational benefits is achieved through the active participation of teachers and parents, i.e., playing with children and relating the experiences in the games to the real world (Morais et al., 2025).

Teacher Support Tools: AI-driven analytics for tracking developmental milestones and identifying learning needs

Analytics and observation tools made with the assistance of AI can assist educators in recording the progress of the development of children, learning needs, and making evidence-based decisions. These systems process the observational data, video recordings, and performance measurements to produce the cognitive, social-emotional, and physical development profiles (Destriana et al., 2025). By training AI on developmental milestones, one can identify delays, differences in learning, or the necessary assistance to provide and act promptly.

They can be used in combination with intelligent documentation systems, which turn the videos and notes of classrooms, predictive analytics, which predicts which children may require extra support, and real-time dashboards which provide insights into individual and group learning trends (Destriana et al., 2025). These tools support differentiated instruction by recognizing the various learning profiles and approaching them (Almawajdh et al., 2025).

Effectiveness depends on human-centered design that preserves teacher judgment and expertise. AI should augment, not replace, observation—surfacing patterns across large datasets, tracking progress, and reducing administrative burdens to free time for direct interaction (Kaur, 2026). Transparent algorithms, strong governance, and pedagogical grounding can promote equity by identifying children who might otherwise be overlooked (Destriana et al., 2025).

Ethical use requires rigorous data privacy protections (Pölzl-Stefanec, 2026), algorithmic transparency, cultural responsiveness, and preservation of teacher autonomy. AI must inform rather than determine decisions, ensuring educators remain central in assessment and intervention planning (Almawajdh, 2025).

4. Benefits and Opportunities

Although the incorporation of AI in early childhood education is an issue that poses significant threats, it is also shown in the literature that it has significant benefits and opportunities when applied carefully. In this section, the author discusses four key spheres where AI can support early learning including improved learning outcomes through personalization, early detection of developmental issues, greater engagement and motivation, promoting teacher professional competencies in curriculum planning and assessment.

Enhanced learning outcomes through personalization

Among the most intriguing advantages of AI in early childhood education, the ability to make truly personalized learning experiences based on the individual developmental profiles and the learning preferences can be identified. It has been shown that AI-based adaptive learning systems can substantially enhance educational achievements among different groups of students (Li, 2025). The experimental group with an AI-based personalized learning platform demonstrated statistically better learning results than the control group with the traditional instruction reported in a quasi-experimental study with 60 kindergarten children, the numeracy and problem-solving skills showed significant improvement in the experimental group (Li, 2025). Measures of engagement also were found to be superior: the children in personalized learning plans had higher rates of completing tasks, and better time-on-task than their counterparts in non-personalized plans.

In addition to the short-term performance improvements, the individual learning routes help with retention and transfer in the long term. The experiments using delayed post-tests and transfer tests indicate that adaptive systems enhance the capacity of children to create the new situations on the basis of the learned concepts (Li, 2025). This phenomenon is notably strong among learners with lower initial performance, which implies that the personalization of AI can be used to narrow the achievement gap and facilitate educational equity (Zhang et al., 2026). Notably, such advantages come not at the cost of teacher expertise but rather as a way of increasing the capacity of educators to identify and address individual learning requirements at scale. The responsive nature of the environments in the mode of flexible alignment to the developmental milestones as operationalized in frameworks such as the AI-Enhanced Early Learning (AI-EL) model allows a child to adapt to the changing competencies in real-time (Tao & Narsi, 2025).

Early identification of special needs

One of the biggest opportunities of AI in the field of promoting equitable learning outcomes is early detection of learning differences and developmental delays. Machine learning models that are trained on developmental milestone data are capable of analyzing observational patterns, performance metrics and videos to detect children at risk of developmental delays with significantly higher accuracy than conventional methods of screening that involve manual examination (Kaur, 2026). The ability to detect at an early age and accurately is essential: studies have repeatedly shown that children diagnosed at an early age and receiving early intervention have significantly better developmental patterns and academic achievements in the long run than those whose needs remain unmet (Gutierrez-Ruiz & Montes, 2024).

Pattern recognition with multimodal, complex data streams, indicative of developing competencies in children, are an area where AI systems demonstrate great performance. Intelligent systems can detect subtle developmental patterns on cognitive, social-emotional, and physical dimensions simultaneously that would otherwise be overlooked when using single-dimension screening (Thamma et al., 2025). To illustrate, AI-based facial recognition and mood detection systems have already proven to be 95% or higher accurate in early autism spectrum disorder and other developmental disorders diagnosis (ElMahalawy et al., 2024). In addition to diagnostic use, predictive analytics systems can also predict which children can be helped by specific interventions before major delays have already occurred, which would allow providing proactive help instead of reactive help.

More importantly, AI tools need decisions and judgments by humans to identify such cases at an early phase. It has been stressed in research that the use of AI is aimed at complementing and speeding up human judgment rather than substituting it (Gutierrez-Ruiz, 2024). The role of teachers and specialists is critical to put AI-generated insights into perspective, taking personal situations into account, and making decisions regarding interventions. Under the right human supervision, clear algorithms, and cultural sensitivity, AI tools can assist equity by assisting in recognizing and supporting children that would not be otherwise noticed, especially in lowly-resource neighborhoods (Destriana et al., 2025).

Increased engagement and motivation in young learners

Motivation and student engagement are essential facilitators of learning especially at the early childhood stage when the early attitudes towards learning are being established. AI-enriched learning channels show stable ability to increase engagement on cognitive, emotional and behavioral levels. In research on generative AI in early childhood education, it has been found that AI benefits learning results and individualized learning as well as reinforcing engagement in an active mediation by adults (Zhang et al., 2026). Improvement in engagement has been observed in various measures such as persistence in tasks, desire to solve problems that are difficult, and excitement when learning.

Incorporated elements of gamification with AI algorithms produce especially effective motivational impulses. The adaptive difficulty algorithms make sure that the game is not so easy that it is boring and not so hard such that it feels inappropriate and frustrating (Arvanitaki et al., 2025). Real-time analytics enable the systems to detect gaps in learning, and automatically modify game content and scaffolding to respond to emergent needs (Tsao & Nasri, 2025). Studies on game-based learning in early childhood learners have shown that properly designed AI-enhanced learning games yield moderate to large effect sizes on learning, especially the outcomes of problem-solving, logical thought, and creative expression (Zuceti et al., 2024). Notably, the improvements in engagement are not limited to the cognitive areas but in studies, social-emotional engagement, greater peer collaboration, and evolution of help-seeking behaviors have been noted in AI-enhanced settings.

Nonetheless, authentic engagement is something that has to be designed carefully basing on developmental principles. Studies have highlighted the importance of game-based and app-based learning to strike the right balance between structured learning and unstructured discovery, not spend too much time on screens by balancing with play-based learning, and achieving pedagogical quality and cultural sensitivity (Hibana et al., 2024). The mediating roles of teachers and parents are critical--it is best to play with children, ask questions, and be able to establish links between game experiences and the real world to maximize learning outcomes and maintain a long-term presence (Morais et al., 2025).

Support for teachers in curriculum planning and assessment

The integration of AI in early childhood education presents substantial benefits in terms of teacher support in case it is framed properly. Good tools add to and not substitute teacher expertise, but serve as extensions of educator capacity. AI-fueled analytics is capable of dealing with huge amounts of observational data and performance metrics to create development profiles, emerging

patterns, and diminished documentation (Kaur, 2026). This liberates teacher time to interact and form relationship with the teacher, which studies have found to be critical to early learning.

Smart documentation systems can watch development towards developmental standards by analyzing classroom recordings and notes and identifying areas requiring help or achievement (Destriana et al., 2025). Predictive analytics predict which children could use further interventions, and this allows differentiated instruction (Almawajdh, 2025). Real-time dashboard gives actionable information on the trends of individual and the classroom which can be used to make evidence-based decisions. Curriculum planning and assessment design Generative AI might also be helpful in creating a customized lesson plan and differentiated activity (Ahn, 2025). Integrated into teacher education, these tools strengthen instructional design competencies, with documented gains in lesson planning and pedagogical effectiveness (Ahn, 2025).

Effectiveness depends on human-centered design that preserves teacher judgment. Transparent algorithms and robust governance are essential for validity and equity (Destriana et al., 2025). When grounded in pedagogy and overseen by educators, AI enhances teaching capacity while ensuring professional autonomy remains central (Kaur, 2026).

5. Challenges and Risks

Artificial intelligence (AI) has its advantages in the field of early childhood education and has threats that should be overcome. The most significant issue is excessive screen time. Health agencies suggest that children under the age of 2-5 should not be exposed to screens more than one hour a day and that they should be supervised by parents and watch high-quality content (Yelizarova et al., 2025). Overuse would lead to lack of physical activity, sleeping issues, emotional disturbances, language retardation, and impaired social-emotional development (Yelizarova et al., 2025). Studies find that digital learning should not consume more than 30 percent of the entire learning time to maintain the play-based and practical learning (Thaithi, 2025).

Ethics and data privacy are also an important issue. AI systems receive sensitive data, including performance, behavior, and emotional data but most institutions in the commentary do not have clear policies regarding protection and accountability (Aliyu, 2025). Algorithms can be dangerous, since small datasets can lead to inaccurate or unjust outcomes, especially when it comes to face recognition (Jiang, 2025). It must be well governed, parental consent, transparent, and safely stored (Aliyu, 2025).

Play-based learning, social learning, which are essential in development, can also be diminished by AI (Chen, 2025; Abah et al., 2025). The automation bias would undermine the judgment of the teachers (Qadir & Mumtaz, 2026). Lastly, a lack of teacher preparedness is also a problem, as there are gaps in AI knowledge, ethics, and integration solutions (Athanassopoulos et al., 2026; Dong, 2024; Polzl-Stefanec, 2026). It requires continuous training and institutional assistance, particularly in the developing contexts (Jiang et al., 2025).

6. Future Directions

Artificial intelligence (AI) is also influencing early childhood education and bringing opportunities and challenges. They are using the tools of computer vision, facial recognition, and engagement tracking systems to monitor interactions in classrooms, analyze the emotions of children, and evaluate their participation in real time. These systems can determine behavioral and emotional patterns with great accuracy using deep learning and convolutional neural networks to support instruction and research (Mohamed et al., 2025). Edge AI, which furnishes responses at the stage of data processing, assists in working with the privacy issue and furnishes the teacher with the response in the form of a real-time feedback (Souhir et al., 2025). These technologies would help teachers detect both disengaged and failing students at an early stage (Kengesbay, 2025).

Moral aspects are kept to the fore. The privacy, the consent of parents, the safety of the data, and the role of the teachers should be ensured. AI is not to substitute the professional judgment of teachers, but rather is to be used in consultation with it (Andrejevic & Selwyn, 2019). Researchers emphasize on the hybrid intelligence, when AI will offer insights and personalized advice; the teachers will still be in charge of teaching, socializing, and child growth (Jarvela et al., 2025). With technology, play-based learning and relationships should not be reduced, but instead increased (Dodero & Giovannella, 2025; Mutmainah et al., 2025).

Reform of policies and curriculum is vital. Privacy, ethics, and accountability require guidelines set by the governments and institutions (Aliyu, 2025). Training of teachers should be done on AI literacy and ethical training (Jiang, 2026), but should be equitable to avoid inequality (Jiang, 2025). More studies are needed to leverage interdisciplinary work and longitudinal studies to evaluate the effects of AI on child development in the future (Burbano-Enriquez et al., 2025).

7. Conclusion

This review discussed the backgrounds, uses, advantages, issues, and prospects of artificial intelligence (AI) in early childhood education. The results indicate that AI can have a high potential in enhancing personalized learning, early identification of developmental needs, higher student engagement, and teacher instructional planning. Adaptive learning systems, natural language processing tools, educational games, and AI analytics can improve the learning process when applied as an aid and not as a substitute of teachers. The most appropriate uses of AI are strengthening teacher professionalism, enhancing human interaction, and individualized learning.

Nevertheless, serious challenges were also found in the review. They are gaps in teacher preparedness, Data privacy, Algorithms bias, too much screen time, and inequality in access to digital infrastructure and resources. Unless well-planned and supported, AI implementation can make educational inequality more than it would be with no AI use.

The review highlights the existence of the necessity of a balanced and human-oriented approach towards AI integration. The implementation of AI must be provided in the framework that does not violate the play-based learning, social interaction, and child-centered teaching practices, but the ethical standards, data privacy, and child protection. In general, human-centered AI the technology that fosters human teaching but does not substitute it is the most promising tool in early childhood education.

Recommendations

The review suggests to educators and educational leaders to use artificial intelligence (AI) in education in a cautious and pedagogical manner. The use of AI tools cannot be implemented blindly because they are present but they need to respond to the educational aims and the known principles of child development. Professional development of teachers should be one of the priorities, and it is necessary not just to speak about technical skills but also about AI literacy, ethical consciousness, and pedagogically justified integration. Play-based learning, social-emotional development, and close teacher-child relationships have to remain the priorities of educators despite the introduction of AI; the technology should be used to enhance them and not to substitute them. The schools should also develop effective data governance policies, transparency, and parental engagement practices to maintain the privacy of children and steward educational data in a responsible manner.

To the researchers, the review identifies the necessity of the long-term and thorough studies into the impact of AI on the cognitive, social-emotional, and physical development of children. The field that will be pursued in future studies is equity, and the impact of AI on learners with diverse and under-resourced backgrounds in particular. It is urged to conduct interdisciplinary studies that could include educators, technologists and ethicists, as well as research on the effective teacher training models, capable of enhancing the AI literacy, without altering the child-centered approach to education.

References

Abah, M., Ehiwario, K., Oladosu, M., & Yohanna, N. (2025). The effects of technology on early childhood learning literature. *Annals of Medical and Health Research: An International Journal, 3*(2). https://doi.org/10.51470/wep.2025.3.2.01

Ahn, J. (2025). Integrating ADDIE, backward design, and AI to develop curriculum design competencies in preservice special education teachers. *Korean Educational Research Association, 63*(7). https://doi.org/10.30916/kera.63.7.29

Akintola, A. S., Akintayo, M., Kadri, T., Oforgu, C. M., Michael, M., & Nwanna, M. (2025). Adaptive AI systems in education: Real-time personalised learning pathways for skill development. *Journal of Artificial Intelligence, Machine Learning and Data Science*. https://doi.org/10.51219/jaimld/akinyemi-sadeeq-akintola/534

Aliyu, A. Y. (2025). Towards a policy framework for the use of generative AI in Nigerian curriculum for early childhood education. *Journal of Early Childhood Development and Education, 2*(2). https://doi.org/10.58723/junior.v2i2.332

Al-Hassan, O., Alhasan, L., AlAli, R., Al-Barakat, A., Al-Saud, K. M., & Ibrahim, N. A. (2025). Enhancing early childhood mathematics skills learning through digital game-based learning. *International Journal of Learning, Teaching and Educational Research, 24*(2). https://doi.org/10.26803/ijlter.24.2.10

Almawajdh, I. M. R. (2025). The role of AI in supporting differentiated instruction for primary education. In *2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT)*. https://doi.org/10.1109/ICICAT68430.2025.11414514

Andrejevic, M., & Selwyn, N. (2019). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology, 45*(2), 115–128. https://doi.org/10.1080/17439884.2020.1686014

Arvanitaki, A., & Skoumpourdi, C. (2025). Educational games for indirect length comparison in sighted and blind students. *European Conference on Games Based Learning, 19*(1). https://doi.org/10.34190/ecgbl.19.1.3909

Athanassopoulos, S., Tzavara, A., Aravantinos, S., Lavidas, K., Komis, V., & Papadakis, S. (2026). Teacher education students' practices, benefits, and challenges in the use of generative AI tools in higher education. *Education Sciences, 16*(2). https://doi.org/10.3390/educsci16020228

Azhar, M., Hin, H. S., & Kian, N. T. (2025). Enhancing early childhood reading skills through immersive AR and conversational AI. In *International Conference on Innovation Engineering and Technology*. https://doi.org/10.1109/ICIET66371.2025.11046302

Burbano-Enríquez, B., Moncayo-Cueva, H., & Andrade-Zuleta, A. (2025). Neuroeducation and the influence of AI on early childhood education: A systematic review. *Multidisciplinary Reviews*. https://doi.org/10.31893/multirev.2026238

Bull, G. L., Nguyen, N. R., & Langran, E. (2025). AI in informal and formal education: A historical perspective. *AI-Enhanced Learning*. https://doi.org/10.70725/417039flneos

Chen, J. (2025). Exploring the role of artificial intelligence in personalized learning experiences in early childhood education to improve children's social skills. *Lecture Notes in Education Psychology and Public Media*. https://doi.org/10.54254/2753-7048/2025.bo24154

Cimino, S., Maremmani, A., & Cerniglia, L. (2025). The use of artificial intelligence (AI) in early childhood education. *Societies, 15*(12). https://doi.org/10.3390/soc15120341

Destriana, R., Aksani, M. L., Priyanggodo, D. Y., & Farzani, R. (2025). Design of a digital platform for PAUD child development monitoring using the dynamic systems development method and machine learning. *Jurnal Teknik Informatika (Jutif), 6*(5). https://doi.org/10.52436/1.jutif.2025.6.5.4958

Dodero, G., & Giovannella, C. (2025). Smart learning ecosystems beyond 2030. *Interaction Design and Architecture(s)*. https://doi.org/10.55612/s-5002-064-001psi

Dong, C. (2024). Early childhood AI education in China: Time value and path exploration. *International Journal of Academic Research in Progressive Education and Development, 13*(3). https://doi.org/10.6007/ijarped/v13-i3/22731

ElMahalawy, J., ElSwaify, Y. A., Elliboudy, D., Abbas, O. M., Moustafa, N., & Wael, N. (2024). AI-powered human-computer interaction assisting early identification of emotional and facial symptoms of autism spectrum disorder in children. In *2024 International Conference on Machine Intelligence and Smart Innovation (ICMISI)*. https://doi.org/10.1109/ICMISI61517.2024.10580320

Gutiérrez-Ruiz, K., & Montes, Y. S. (2024). Early detection of neurodevelopmental disorders as a strategy for educational inclusion in early childhood education. *International Journal of Disability, Development and Education*. https://doi.org/10.1080/1034912X.2024.2370799

Hang, C. N., & Ho, S. M. (2025). Personalized vocabulary learning through images: Harnessing multimodal large language models for early childhood education. In *International Symposium on Electronic Commerce*. https://doi.org/10.1109/ISEC64801.2025.11147254

Hibana, H., Nayla, M. R., & Nurhayati, K. (2024). Exploring the role of game-based learning in early childhood cognitive development: Perspectives from teachers and parents. *Golden Age: Jurnal Ilmiah Tumbuh Kembang Anak Usia Dini, 9*(4). https://doi.org/10.14421/jga.2024.94-12

Jarodzka, H. (2025). Teaching and learning in the age of generative AI: Rethinking the educational cycle. *AI-Enhanced Learning*. https://doi.org/10.70725/993596wouvvd

Järvelä, S., Zhao, G., Nguyen, A., & Chen, H. (2025). Hybrid intelligence: Human–AI coevolution and learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13560

Jiang, A. H., Bahng, E., Coffman, C., Shelley, M., Gilbert, S. B., Fieffer, S. J., & Abudagga, O. (2026). Preparing future teachers for AI-enhanced science classrooms. *Journal of Digital Learning in Teacher Education*. https://doi.org/10.1080/21532974.2025.2606665

Jiang, H. (2025). Bridging the gap, creating equity through wisdom. *INTI Journal*. https://doi.org/10.61453/intij.202550

Kamalov, F., Calonge, D. S., & Gurrib, I. (2023). New era of artificial intelligence in education. *Sustainability, 15*(16). https://doi.org/10.3390/su151612451

Kaur, D. (2026). Role of AI in enhancing teaching learning processes and academic performance. *International Journal for Multidimensional Research Perspectives, 4*(2). https://doi.org/10.61877/ijmrp.v4i2.308

Kaur, D. (2026). Public health perspectives on AI-driven mental health support for children with special needs. *International Journal of Global Mental Health, Innovation, Policy, Action, Culture & Transformation, 2*(1). https://doi.org/10.61113/impact.v2i1.1254

Kengesbay, D. (2025). Detecting social conflicts in kindergartens using deep learning and computer vision. *Journal of Emerging Technologies and Computing*. https://doi.org/10.47344/7x77b619

Kurian, N. (2023). AI's empathy gap: The risks of conversational artificial intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*. https://doi.org/10.1177/14639491231206004

Li, Y. (2025). The construction and verification of early childhood personalized education path based on artificial intelligence. *Scientific Journal of Humanities and Social Sciences*. https://doi.org/10.54691/c8eqyf48

Mohamed, A., Tarek, K., Eskander, M. E., Emad, J., Kenawy, R., Wassim, N., & Kasem, H. (2025). An integrated deep learning framework for automated monitoring of student engagement and emotional states in educational environments. In *2025 13th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*. https://doi.org/10.1109/JAC-ECC67970.2025.11417499

Morais, A., Guimarães, A. F., Oliveira, A. T. D., Sperandio, A., & Camilo, I. A. (2025). The child as protagonist: Active methodologies in early childhood education. *Latin American Journal of Development, 7*(2). https://doi.org/10.46814/lajdv7n2-003

Mutmainah, Rosidah, S., & Ilham, M. (2025). Literature review on 21st century learning strategies. *RESET, 1*(1). https://doi.org/10.66031/reset.v1i1.16

Noriana, S., & Bakar, K. A. (2025). The role of artificial intelligence in enhancing early literacy. *International Journal of Research and Innovation in Social Science, 9*(3). https://doi.org/10.47772/ijriss.2025.903sedu0660

Oshinowo, O. R., Agoi, M. A., Sain, Z., Sain, S. H., & Aziz, A. L. (2025). An AI-driven teaching prototype for early childhood education. *Journal of Multidisciplinary Research, 4*(2). https://doi.org/10.56943/jmr.v4i2.867

Ozturk, E. (2025). Artificial intelligence in early childhood STEM education. *Journal of Education in Science Environment and Health*. https://doi.org/10.55549/jeseh.800

Pinska, O. L., Shokaliuk, S., Alieka, H. I., & Zakarliuka, I. (2025). Psychological foundations of technology-mediated educational process management. *Educational Technology Quarterly*. https://doi.org/10.55056/etq.1213

Pölzl-Stefanec, E. (2026). Perceptions of opportunities and risks posed by artificial intelligence. *Education Sciences, 16*(2). https://doi.org/10.3390/educsci16020202

Purwandari, N., & Nasution, T. (2025). Development of interactive learning applications based on artificial intelligence. *International Journal of Science, Technology & Management, 6*(3). https://doi.org/10.46729/ijstm.v6i3.1218

Qadir, J., & Mumtaz, M. (2026). The psychology of learning from machines. *arXiv*. https://doi.org/10.48550/arXiv.2601.06172

Qayyum, A., Bukahri, M., Zulfiqar, P., & Ramzan, M. (2024). Balancing artificial intelligence and human insight in early childhood education. *Social Science Review Archives*. https://doi.org/10.70670/sra.v2i2.207

Qayyum, A., Rafique, Z., Shehr, S., Shah, W. A., Ahmad, S., & Haider, Z. (2025). Artificial intelligence (AI)-driven curriculum development in early childhood education. *Research Journal of Psychology, 3*(1). https://doi.org/10.59075/rjs.v3i1.102

Sari, N. K., Wicaksana, M. F., & Rahman, Md. A. (2025). Adaptive AI-driven formative assessment in early childhood education. *Child Education Journal, 7*(3). https://doi.org/10.33086/cej.v7i3.8412

Souhir, B., Soulef, B., Ahmed, O. B., Ali, H. M., Abdelatif, M., & Jihad, A. J. (2025). Edge AI framework for real-time facial emotion recognition. In *International Conferences on Computing Advancements*. https://doi.org/10.1109/ICCA66035.2025.11430978

Tao, G., & Nasri, N. B. M. (2025). Artificial intelligence in early childhood education. *International Journal of Academic Research in Progressive Education and Development, 14*(2). https://doi.org/10.6007/ijarped/v14-i2/25151

Thaithi, N. (2025). The effects of digital learning on international schools' curriculum in Kenya. *African Multidisciplinary Journal of Research, 2*(2). https://doi.org/10.71064/spu.amjr.2.2.2025.432

Thamma, S. R., Devalampeta, B. R., & Jangareddy, M. R. (2025). Early detection of pediatric developmental disorders. In *2025 International Conference on Next Generation Communication & Information Processing (INCIP)*. https://doi.org/10.1109/INCIP64058.2025.11020380

Thompson, J., & Harris, O. (2025). A narrative review of educational technology in higher education. *Social Science Chronicle*. https://doi.org/10.56106/ssc.2025.002

Tsao, J., Tsao, Y.-C., Liu, L.-Y., & Lin, J. (2025). Game-based learning in early childhood. In *2025 9th International Conference on Education and Multimedia Technology*. https://doi.org/10.1145/3761843.3761938

Yelizarova, O., Stankevych, T., Parats, A., Yelizarov, V., Puzanova, O., Lebedynets, N., & Hozak, S. (2025). Digital exposure in early childhood. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*. https://doi.org/10.1177/00469580251390766

Zhang, Y., Halili, S. H., & Zainuddin, Z. (2026). Applications of generative AI in early childhood education. *Eurasia Journal of Mathematics, Science and Technology Education*. https://doi.org/10.29333/ejmste/18068

Zhou, X. (2025). Innovation and reconstruction of early childhood education models driven by artificial intelligence technology. *Journal of Artificial Intelligence Practice, 8*(1). https://doi.org/10.23977/jaip.2025.080110

Zuceti, H. R. M., Juliana, H.-R. J., Sotelo-Núñez, A. C., López-Regalado, O., Enrique, B.-C. L., & Alberto, M.-T. D. (2024). Bibliometric analysis on educational games in preschool children's learning. *Nanotechnology Perceptions, 20*(14). https://doi.org/10.62441/nano-ntp.v20is14.45

How to Cite this Chapter (APA 7):

Pragacha, R. N. (2026). Artificial intelligence in early childhood education: Foundations, applications, and pedagogical implications. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 96–104). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 13: Artificial Intelligence and Information Technology in Language and Literature Studies: A Systematic Review of Emerging Applications

Dr. Lemellu Nida N. Sarmiento
Cavite State University - Silang Camps

Abstract. *Artificial Intelligence and Information Technology are fundamentally transforming language and literature studies, creating opportunities for textual analysis, literary interpretation, and language education. This literature review synthesizes contemporary research on AI/IT applications across English Studies, examining natural language processing for automated literary analysis, machine translation and language learning systems, digital literary analysis algorithms, generative models for creative writing, and interactive AI-driven educational platforms. While AI technologies enhance research efficiency through scalable pattern recognition and enable new critical perspectives in literary scholarship, they simultaneously present significant challenges including algorithmic bias in language models, risks of interpretive overreliance, ethical concerns regarding authorship and originality, and inadequate teacher and student preparedness for responsible integration. The review emphasizes that most promising applications employ hybrid intelligence approaches combining computational efficiency with humanistic interpretation rather than attempting to automate literary criticism itself. Future directions include expanded digital humanities research, integration into pedagogical practices, cross-disciplinary collaboration between computer scientists and literary scholars, and development of institutional policies supporting equitable access and ethical AI use. The review concludes that AI works best as a discovery tool augmenting human expertise, maintaining critical humanistic values while leveraging technological capabilities for enhanced research, teaching, and accessibility across diverse learner populations.*

Keywords: *Artificial Intelligence, Digital Humanities, Literary Analysis, Language Education, Generative AI, AI Ethics*

1. Introduction

Artificial Intelligence and Information Technology have been playing a critical role in humanities scholarship and have transformed the way scholars interact with texts, cultural objects, and historical documents. The introduction of AI in the humanities is an important development of the manual analysis to the computationally-assisted inquiry. Being an interdisciplinary area, Digital Humanities is a combination of digital tools and techniques with humanistic research, and it opens up new possibilities in the understanding of culture, language, and literature (Odili, 2025).

The disruptive effect of AI is spread out to various fields of humanities. NLP, machine learning and data analytics advance the essential practices of digital humanities, such as text mining, sentiment analysis, and creative writing, and digital curation (Odili, 2025). These technologies make the research in humanities more efficient, more precise, and more scalable, allow collaborators to work across disciplines, and allow discovering new knowledge with the help of processing extensive data and identifying patterns. The AI technologies can also help scholars to deal with extensive literary corpora and identify hidden patterns as well as gain a better insight into the systems of narratives, their stylistic elements, and cultural trends (Mujahid et al., 2025).

English Studies is a very fertile field of AI and IT implementation that includes language learning, literature analysis, creative writing, and textual interpretation. English as lingua franca in the world is not only an important target of language learning technology, but it is also a key topic of analysis of digital literature. It has been observed that in the field, enrollments have declined in modern language departments, and so technology solutions are becoming more significant to education and availability (Kern, 2024).

The computer-based methods of studying English allow researchers to process the canon texts in a larger scale than ever before. The techniques that utilize AI can enable a researcher to identify themes, motifs, and stylistic devices in a novel, poem, or drama that are possibly invisible to a close reading. Moreover, the overlap of English Studies and digital humanities has produced new pedagogical possibilities, enabling professors to incorporate AI technologies that can be used to support grammar studies, vocabulary building, and writing skills without sacrificing critical humanistic approaches to literature and language.

Objectives of the review

This comprehensive review aims to achieve several interrelated objectives:

1. Synthesize current knowledge about how AI and information technology are being applied across English Studies;
2. Examine theoretical foundations connecting computational methods with established linguistic and literary inquiry frameworks;
3. Evaluate practical applications across multiple domains—NLP for text analysis, machine translation for language education;
4. Identify opportunities that AI technologies create for enhancing textual analysis efficiency, democratizing access to literary scholarship;
5. Consider future directions for AI in English Studies education.

2. Foundations of AI in Language and Literature

Overview of AI methods relevant to text and discourse analysis

The field of Artificial Intelligence has proven to be useful in the analysis of linguistic and literary texts. The main AI methods that can be applied to research in languages and literature are:

Natural Language Processing (NLP) is the underlying technology of the computational text analysis. NLP helps computers to read, comprehend and produce human words and is a necessary model in activities like text categorization, information mining and semantics (Othman et al., 2025). The NLP techniques are sentiment analysis that detects emotional tone and opinion in the text, topic modeling that detects the underlying thematic patterns, and named entity recognition that detects important entities within the text,

Deep Learning models has transformed the NLP properties. Transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have also reached the state of the art in a variety of language tasks (Othman et al., 2025). These models use neural networks which are automatically trained using large training set of linguistic patterns which allow them to understand context, meaning and variation of style.

Machine Learning algorithms facilitate supervised learning (when the models are trained using labeled data) and unsupervised learning (when the models determine patterns without specific labels). Such methods allow to assign authorship, genre, and style analysis (Rios-Toledo et al., 2022). There are also Support Vector Machines (SVMs), Random Forests, and ensemble methods that are still very useful tools of analysis in addition to deep learning methods.

Computational linguistics can be used to compute large-scale patterns, and data analysis and visualization techniques enable researchers to present complex phenomena in linguistics in understandable formats (Adorni & Grosso, 2025).

Historical development of computational linguistics and digital humanities

Computational linguistics and digital humanities evolved out of the convergence of computing and humanistic study. This area started in the 1960s-1980s as the first computers were utilized by researchers to analyze literature and linguistics, such as concordancing and rule-based linguistic processing. In the 1990s to 2000s the field grew to include corpus linguistics, the application of large bodies of digitized text to make quantitative analyses of language possible. Digital humanities became a new field that synthesized research in humanities with the use of digital technologies like text mining, stylometric analysis, and computationally-based literary criticism during this time (Odili, 2025).

It has changed significantly in the 2010s as machine learning and deep learning technologies emerged. Transformer architectures and large language models like BERT and GPT were developed in 2017 and improved natural language processing applications and made AI tools more accessible to researchers (Adorni & Grosso, 2025). In 2022-2023, the widespread release of generative AI systems led to more awareness about the use of AI in humanities research, creative writing, and education. Currently, hybrid methods of digital humanities research that involve computational analysis and human interpretation are typical (Boer & Stork, 2024).

There are a number of theoretical frameworks that inform the application of AI in literary and linguistic works. Hybrid Intelligence model focuses on the collaboration of human researchers and AI systems and helps in distant reading (analysis of large texts) and close reading (detailed analysis) (Boer & Stork, 2024). Reader Response Theory and Phenomenology place significant emphasis on the role of human interpretation, cultural background, and experience of the reader in the process of interpreting literature, which implies that AI is only meant to be used as an aid, not a replacement of human interpretation (Nursaid et al., 2025). Ethical AI systems place focus on transparency, accountability, fairness, and explainability to make sure that AI is used responsibly in research and education in humanities. These paradigms will facilitate the sensible use of AI and maintain humanistic values in research and teaching.

3. Applications in English Studies

Natural Language Processing (NLP): Automated text analysis, sentiment analysis, and stylistic studies of literary works

NLP has taken the center stage in the field of computational literary analysis as it allows researchers to analyze large amounts of textual corpora and find patterns of meaning at scales unattainable by manual reading.

NLP techniques enable researchers to discover and categorize themes, motifs, and narrative patterns in literature by extrapolating the text with relevant tools. Topic modeling methods can be used to demonstrate underlying thematic structures through examining the distribution of words and the discovery of semantically related clusters. Named Entity Recognition (NER) is a technique that detects key entities, such as characters, places, organizations, and so forth, to allow systematic research on the construction of fictional worlds (Bajaj et al., 2025).

Sentiment analysis is an NLP-based technique of classifying the emotional tone and opinions conveyed in text. Sentiment analysis, when applied to literature, displays character psychology, plot development, and sentiment of works. More sophisticated models such as BERT-based transformer models can gain a subtle insight into implicit sentiment, irony, and more complex expressions of emotion that cannot be conveyed by traditional lexicon-based models (Zhang, 2024).

NLP-based stylistic analysis allows one to study the authorial voice, literary movements, and genre features in a systematic manner. The stylometric features, such as word frequency, sentence structure, and lexical choices can be processed computationally, which allows authors to identify authorship, trace the stylistic changes, and comprehend the role of linguistic

choices in making meaning (Rios-Toledo et al., 2022). Syntactic parsing and part-of-speech tagging can expose grammatical patterns to make literary styles different (Hameed & Ali, 2025).

Recent literature also shows that deep learning models can be useful in literary analysis by embedding semantic relationships, identifying hidden linguistic patterns, and identifying theme associations in large literary collections (Mujahid et al., 2025). Nevertheless, researchers point out that computational analysis is most productive with human interpretive skills, where the pattern detection algorithm is used to inform but not substitute close reading (Ganiyeva et al., 2024).

Machine Translation and Language Learning: AI tools supporting multilingual education and second-language acquisition

Language learning and machine translation have been transformed by artificial intelligence, which offers potent means of second language acquisition (SLA) and multilingual education and provokes relevant concerns over the integration of pedagogy.

The neural machine translation (NMT) technology in machine translation (MT) systems has demonstrated an impressive quality in translation and can now translate languages between various language pairs as if they are humans (Patnaik et al., 2025). Translators driven by AI such as Google Translate and ChatGPT provide immediate availability of cross-linguistic comprehension, thus more convenient to learn a language, but necessitating cautious pedagogical attention to avoid over-use (Mubarak, 2026).

Language learning technologies that are enhanced with AI offer adaptive learning content that is personalized to the levels of individual learner proficiency and learning style. Intelligent tutoring systems apply machine learning to optimize learning challenge, give real-time corrective feedback, and adjust the pace of instruction. Chatbots and conversation AI allow people to engage in real-life communication, minimize anxiety, and offer regular language production opportunities (Ginting & Hina, 2025).

It is based on the Natural Language Processing technologies, which support various aspects of language learning: speech recognizing systems allow practicing pronunciation; grammar and style checkers help to improve writing; machine translators assist in comprehension; sentiment analysis can be used to evaluate the emotional reaction to learning content (Fang, 2025).

The research however stresses on balanced integration. Although AI-assisted learning makes the process more efficient and accessible, human contact, cultural background, and instructional support are still crucial in helping to build real communicative competence and avoid over-reliance on the use of technology solutions (Chapelle, 2024). An effective language teaching would be a blend of artificial intelligence (AI) ability to personalize and practice with an instructor knowledge of cultural insights, subtlety of error correction, and encouragement.

Digital Literary Analysis: Using algorithms to detect themes, motifs, and stylistic patterns in novels, poetry, and drama

Digital literary analysis is a subfield of computational linguistics that builds on diagnostic computational methods, using them to identify complex patterns, structures and meaning to scales of scale previously not possible.

Thematic analysis applies semantic network, topic modeling and deep learning to detect and trace themes within and across works and literary collections. The sophisticated algorithms have the ability to identify theme development by narratives, correlation of these themes, and how certain linguistic selections are able to trigger thematic material (E, 2025). Character relationship, plot interaction, and intertextual references Network analysis helps to discover the structural patterns not perceived by a traditional analysis (Shahnazari et al., 2025).

Stylistic pattern identification is a method of stylometric analysis that uses the linguistic features of an author to identify a voice, stylometric influences, and stylistic change. Linguistic-based machine learning models are capable of genre classification, pastiche/imitation detection, and uncovering the history of stylistic conventions changing over time within literary movements (Levent & Özbalkan, 2026). High accuracy at recognizing literary themes and structures with hybrid CNN-RNN (convolutional and recurrent neural networks) show that computational approaches can provide both local and long-range structures in the text (Praveen et al., 2025).

Motif in poetry and drama is a symbol, a picture, and a story that has been repeated. Semantic similarity metrics, knowledge graphs, and named entity recognition can facilitate systematic mapping of the operation of symbolic elements in literary traditions.

The frameworks of interpretability play important roles in the assurance that the computational findings can be meaningful to literary scholars. IPAF model specifically tackles this issue by integrating machine learning and SHAP-based explainability, so that computational outputs are directly directly associated with familiar stylistic and thematic patterns. Researchers note that algorithmic analysis is most effective when viewed as a pattern-discovering tool, which complements human analysis instead of trying to automatize it (Mamman & Youssouf, 2026).

Creative Writing and AI: Generative models assisting in poetry, storytelling, and narrative construction

Generative AI systems offer some novel possibilities and challenges to creative writing, especially of poetry, storytelling, and narrative construction. Large Language Models (LLM) like GPT-based systems are capable of generating human text, both poetry, and prose. Hyper-specific models that are trained on certain genres produce content that is rhymed, metrical, and thematic (Hu, 2024). The use of transformer architectures with Generative Adversarial Networks (GANs) also improve emotional coloring, plot, and style subtlety (Franceschelli & Musolesi, 2025).

The best type of AI-assisted writing is co-creative. Narrative-conscious transformer systems are AI text systems coupled with formal storytelling systems to maintain coherence and richness (Patil et al., 2025). There, AI can be used to brainstorm, edit work, and find new styles, and research reveals that the cooperation between humans and AI can often be stronger creatively (Li et al., 2024; McGuire et al., 2024). But there still remain discussions on authorship, originality and creativity. Although AI can copy style, the real creativity requires human experience and perception. The application of ethics demands openness, acknowledgment, and maintenance of human inventive authority (Nandan & Khan, 2025).

In the learning sphere, AI applications assist in the development of grammar, vocabulary, and writing. Such tools as Grammarly also offer real-time feedback (Warbhe &Verma, 2024), and Automated Writing Evaluation systems are used to analyze essays (Chapelle, 2024). Adaptive vocabulary platforms improve retention (Zhang & Huang, 2024; Alshehri et al., 2025). Interactive tutors, chatbots, and speech recognition tools personalize learning, but educators remain essential for context, cultural understanding, and critical feedback (Seddik, 2025; Naqvi et al., 2025).

.

4. Benefits and Opportunities

Enhanced efficiency in textual analysis

The efficiency of textual analysis with the help of AI technologies is increased significantly, so scholars are able to analyze much larger corpora and identify patterns in unprecedented quantities.

Speed in computation is one of the essential benefits. Work that took weeks of manual effort to accomplish, including drawing character networks out of novels, detecting thematic patterns across hundreds of texts, stylometric analysis, and others, can be performed in hours or minutes with AI tools. This efficiency democratization makes large-scale study of literature available to many individual researchers, and smaller institutions, making it available to analyses that would have previously demanded large research institutions (Mujahid et al., 2025).

Computational analysis allows analysis to scale to allow scholars to leave the old mode of close reading of texts and distant reading of literary traditions. Thousands of novels, poems, or plays can be studied in parallel by researchers, revealing the conventions of different genres, tracing the development of literature, and learning to discover the cross-cultural impact (Bajaj et al., 2025). This scaled analysis shows patterns that classic scholarship cannot see even as it supplements the close reading practices.

It is beyond human abilities to be accurate and consistent in identifying linguistic and literary features. AI systems recognize named entities, sentiment cues, stylistic features and structural features with a significant degree of accuracy in large texts, minimizing the human error that could otherwise be present in manual annotation (Praveen et al., 2025). The hybrid solutions based on automated analysis and human verification are best in comparison to a single approach.

Multimodal analysis is a combination of textual, visual and audio data, allowing the study of the literary adaptations, oral poetry traditions and multimedia literary forms (Costa, 2025). AI-based literature search and recommendation engines speed up resource discovery by enabling scholars to find the right scholarship quicker (Eboseremen et al., 2021).

Such efficiency liberates scholars who have to use laborious methods to gather data and recognize patterns and work on the interpretation, contextualization, and theoretical building- the humanistic aspects of scholarship that are distinctly human (Woycicki et al., 2025).

Support for language learning and teaching

Instead of creating issues with the second language education that have been persistent over the years, AI-powered technologies are unprecedentedly beneficial to language learners and teachers.

Individual learning experiences tailored according to their level of proficiency, style, and speed allow them to learn effectively. Adaptive algorithms are dynamic and ensure that the levels of cognitive challenge remain optimal, neither too challenging nor too simple to learners (Alshehri et al., 2025). Immediate feedback grants the ability to correct errors and give positive reinforcement which is critical to language learning (Zhang, 2024).

The improvement of accessibility and inclusivity is especially valuable to those learners with low resources or those who reside in the area where language teaching is lacking. Distance learning with AI can also help to offer quality education to students in different parts of the world without any geographical limitations (Patnaik et al., 2025). The speech recognition technology allows practicing pronunciation without the presence of the native speaker, which is especially useful in a multilingual setting when there is no formal language education (Batumalay et al., 2025).

The teacher support systems help teachers to cope with complex and diverse classrooms. Automated feedback and grading mean less workloads and time saved by teachers when performing routine assessment and focusing on individual instruction (Seddik, 2025). The learning analytics provide data-driven insights that assist teachers to determine how students are progressing, learners who need attention, and how instruction can be changed (Alshehri et al., 2025).

Even without being a native English speaker, grammar checking, style improvement, and vocabulary assistance tools will contribute to support in writing both academic and professional papers, which will even the playing fields. Machine translation helps in understanding the source materials in the original languages, to aid scholars in the process of dealing with primary texts.

Learning languages through cultural inclusivity is enhanced by using AI systems that have been trained on a variety of languages, dialects, and code-mixed text, and with regard to legitimate linguistic variation as opposed to prescriptive norms

(Phadte & Dhore, 2025). With multilingual NLP, it is possible to teach in less-resourced languages that have not been effectively served by educational technology (Oduola & Olajuyigbe, 2026).

Such technological aids do not override, but rather enhance the human aspects of language pedagogy that are vital in the true communicative competence (Kakumanu, 2025).

New perspectives in literary criticism and interpretation

Computational methods allow new ways of understanding works that complement and diversify literary studies in relation to old models of criticism.

Quantitative literary history shows the long-term patterns of literary language, themes, and aesthetics that cannot be seen in the traditional scholarship. The study of large sets of texts allows showing how the semantic material, elements of style, and thematic issues change over centuries and literary movements (Bajaj et al., 2025). Computational semantics can be used to show how meanings of words change over time, which can be used to interpret historical documents (Bingenheimer et al., 2017).

The intersectional reading of multiple dimensions of text together enhances the reading. Character relationship, social dynamics, and power structure Network analysis of literary works highlights how texts are viewed in terms of social organization (Shahnazari et al., 2025). Sentiment analysis with thematic mapping can be used to understand the presence of changes in emotional tone between narrative units and character viewpoints (Hassanin et al., 2025).

Influence mapping and authorial attribution gives new insights on literary relations. The stylometric analysis based on computational methods can be used successfully to determine the authorship of controversial texts, authorial influences and find concealed relations between the authors (Levent & Özbalkan, 2026). This ability makes literature history more rich and allows finding the previously unknown relationships.

Comparative and cross-cultural analysis can be applied by use of computational techniques. The thematic patterns, stylistic features, and narrative structures can be compared systematically by researchers across the literature traditions, which allows drawing new conclusions about the way literature deals with universal themes without losing cultural particularity (Sun, 2025).

Expansion of reader response theory: AI can study reader responses on a large scale, understanding text understanding in the form of social media commentary, online reviews, and educational uses, and develop new views on reception history (Nursaid et al., 2025).

Notably, researchers note that computational knowledge leads to emergent research inquiries and explanatory prospects and not to displacing humanistic interpretation. The AI is a highly effective discovery tool that will reveal patterns that human scholars can explore, put into context, and theorize. This heightened criticism keeps the human creativeness and judgment at the heart of literary studies (Nursaid et al., 2025).

Increased accessibility of texts through digitization and AI tools

The significance of AI technologies in terms of increasing access to literature and scholarship is very high, making interaction with cultural heritage democratic among various groups of people.

Optical Character Recognition (OCR) and Handwriting Text Recognition (HTR) under deep learning is used in digitization and preservation projects to transform historical manuscripts and printed texts into machine-readable form to save endangered documents and make them accessible to the world (Guido et al., 2025). Digital restoration of damaged pages through the help of AI makes it possible to save the vulnerable cultural materials (Guido et al., 2025).

Multilingual access with machine translation is a break through linguistic barrier to literature. Literary works written in languages other than the native speakers can be read by individuals whose command of those languages is poor, but who are aware of the constraints of translation and are motivated to take interest in learning those original languages (Fang, 2025). Speech synthesis also makes available audiobook versions of text, which allow the visual-impaired reader and those who learn better through audition access (Chen et al., 2024).

Search and discovery improvements through semantic search and AI-powered recommendations help readers locate relevant texts and scholarship. Information retrieval systems powered by NLP enable more intuitive searching than traditional keyword approaches, particularly beneficial for non-specialists (Eboseremen et al., 2021).

Open-access platforms incorporating AI analysis provide free scholarly resources to readers worldwide. Digital archives using topic modeling and thematic tagging help users navigate vast collections (Martino et al., 2025). Intelligent cataloging automatically organizes digital collections, improving discoverability.

Educational access improvements enable broader participation in literary study. Interactive educational platforms make sophisticated literary analysis tools available to students in under-resourced educational settings (Adorni & Grosso, 2025). AI tutoring systems provide individual instruction regardless of access to highly trained teachers (Alshehri et al., 2025).

Support for endangered linguistic heritage through speech recognition and language preservation projects employs AI to document and revitalize indigenous and minority languages (Batumalay et al., 2025).

These accessibility improvements realize the democratic potential of digital technology, enabling more inclusive participation in literary culture and scholarship across geographical, linguistic, and socioeconomic boundaries (Fritsche & Münster, 2024).

5. Challenges and Risks

Overreliance on algorithms for interpretation

Algorithms are dangerous in literary research when AI applications are adopted without preserving humanistic interpretive values. Algorithms are very good at identifying patterns in the text that are statistically significant, but they tend to simplify the text, omit irony, metaphor, and cultural subtlety that are of the essence of a literary meaning (Mamman & Youssouf, 2026). AI systems that examine features on the surface are not able to preserve the appearance of how meaning is constructed on a historical basis and reader interpretation (Ganiyeva et al., 2024).

There is also the illusion of objectivity. The human assumptions are encoded in the form of training data, feature selection and the design of the algorithm in the fields of computational analysis, but deep learning makes it opaque where the conclusion is received (Naqvi et al., 2025; Kapoor, 2025). This may confuse academics to think that outputs are conclusive knowledge and not the products of design. Excess dependence is also a danger, as it can reduce the interpretive agency, which will be substituted by non-argumentative acceptance of the results of the algorithms (Naqvi et al., 2025).

Another threat is decontextualization: pattern clusters determined by topic modeling might not have any real literary meaning, generating arbitrary clusters (Costa, 2025). Researchers should not be blinded by the benevolence of the quantitative results and believe that the computational results can serve as the triggers of the research but not the conclusions (Yadav, 2024). Critical distance is a mandatory condition of responsible practice, which involves the use of AI to create questions and emphasize tendencies at the expense of human judgment. Such tools as SHAP and attention mechanisms enhance the interpretability and cannot substitute humanistic reasoning (Naqvi et al., 2025).

Ethical concerns in authorship and originality

The issue of authorship and originality is particularly problematic because generative artificial intelligence (AI) is becoming more and more competent in generating creative and scholarly text that is human in nature. Such advances pose significant ethical issues around intellectual accountability, authorship and creative authenticity. According to the existing international practices, including the ones of the International Committee of Medical Journal Editors (ICMJE), AI cannot be mentioned as an author, since it lacks the capacity to make intellectual choices, bear the responsibility of the work, or any form of approval of the final manuscript (Davison et al., 2024). Nevertheless, as AI tools make meaningful contributions to the creation or editing of writing, it is a complicated matter to establish the level of plausible human authorship. Thus, use of AI should be clearly disclosed, but be subject to human responsibility (Fontenot, 2025; Khan, 2025).

Originality and plagiarism are also disputed by AI. Also, unlike the classic form of plagiarism, it is possible that AI-generated text does not copy other works but makes them look similar, posing ethical and legal ambiguity (Hutson, 2024). It is also possible to risk unintentional plagiarism in case the writers overuse the items suggested by AI. Moreover, academic dishonesty can be driven by the fact that plagiarism detection software might be unable to detect AI written materials (Sadler, 2024).

There are also privacy implications of AI systems due to copyright and intellectual property concerns since the systems are trained with large datasets that can contain copyrighted data. There are doubts about the ownership of AI-generated content, the use of training data fairly, and payment to original creators (Davison et al., 2024; Rokhshad et al., 2025). Also, computer-generated intelligence can represent false or artificial data, referred to as hallucinations that can result in incorrect citation and erratic statements unless verified appropriately (Naqvi et al., 2025; Sadler et al., 2024).

In order to resolve these concerns, responsible use of AI involves transparency, fair disclosure, validation of the content produced by AI, and strict institutional policy to inform ethical use of AI in both academic and creative practice (Fontenot, 2025).

Data bias in language models

The problem of data bias on language models is a severe problem to the impartial and just use of language and literature studies. Training data can also incorporate historical biases and inherit gender stereotypes, racial biases, and cultural hierarchies of internet text and published literature (Guo et al., 2024). Since a large part of scholarship has been presented by Western, English-speaking settings, models are prone to promote hegemonic opinions and underserved voice (Choudhary, 2025). Canonic literary corpora are biased in favor of canonical literature and marginalize other traditions.

Prejudice may be boosted algorithmically as well. Optimization procedures and feedback mechanisms focus on the statistically prevalent trends at the disadvantage of some groups (Guo et al, 2024). The divisions of classes can be detrimental to particular groups of people in systematic ways (Katoch, 2025). Biases between genders and identity are manifested through stereotypical associations and underrepresentation of the LGBTQ+ (Rokhshad et al., 2025; Choudhary, 2025). Analysis is also distorted with racial and cultural stereotyping (Balakrishnan et al., 2025). These are exacerbated by linguistic inequality. NLP systems are trained on large languages and this does not favor minor and small-resource languages (Oduola, 2026). Dialect bias gives preferential treatment to the standard forms, which devalues literary works based on societal or geographical difference (Batumalay et al., 2025). In non-English environments, multilingual models tend to take up the role badly (Mohamed & Alosman, 2025).

Some mitigation measures are the training data diversification, debiasing methods, models auditing, and transparency regarding sources and limitations (Guo et al., 2024). However, there is no complete way of eliminating bias, and it is necessary to be on guard and reevaluate (Naqvi et al., 2025).

Limited teacher/student preparedness for integrating AI tools

The readiness of educators and learners to the introduction of AI is not sufficient regardless of the active implementation of the technologies in the educational process, which poses serious obstacles to successful and decent use.

AI illiteracy: There is a significant number of educators and students who do not have a fundamental knowledge of how AI systems operate, what they can and cannot execute, and what dangers and constraints typify various tools (Fritsche & Münster, 2024). Digital literacy is not limited to simple technological adequacy, but a level of knowledge of algorithmic decision-making, bias, and ethical consequences- critical to responsible use of AI (Jaja, 2025).

Lack of training and professional development: Institutions have failed to train teachers on the use of AI tools pedagogically and ethically (Seddik, 2025). The educators usually do not get guidance on how they can use AI assistance, what possible harms can occur, and how to preserve a critical approach in situations when students use AI help (Jaleel et al., 2025).

Curriculum gaps: Computational thinking, data literacy, or AI ethics is not taught systematically in most educational programs, even though these areas are necessary for graduates of the twenty-first century (Chun & Elkins, 2023). The ability to be AI literate in a critical way that allows to evaluate the end product of AI, to identify bias, and to be ethically responsible is uncommon (Naqvi et al., 2025).

Possible inequities in the distribution of resources: Digital inequities imply that institutions and richer students have access to advanced AI tools and less-resourced settings do not have sufficient technological infrastructure, which increases inequity in education (Minh et al., 2025). The technical aids toward AI tools implementation are usually not available in resource-limited environments (Jaleel et al., 2025).

Resistance and ambivalence: Certain teachers are also well justified in their concern regarding the effects of AI on education, such as the possibility of reduced critical thinking, excessive dependence on technology, and the loss of the human touch in learning relationships (Naqvi et al., 2025). This warning should not be discarded blindly but an insightful discussion should be conducted.

Support is required in the following areas: professional development of teachers (on both technical and pedagogical levels); inclusion of AI literacy and ethics in the curriculum; institutional policies on the acceptable use of AI; fair distribution of resources so that everyone can have access; and research on the long-term effects of AI on education (Chun & Elkins, 2023). The most promising directions are human-centered approaches that focus on AI as an act of tool augmentation and not substituting human teaching (Naqvi et al., 2025).

6. Future Directions

Artificial intelligence (AI) will play a major role in broadening the research in digital humanities by providing new methodologies and research opportunities. Multimodal analysis, a combination of text, images, audio and video used in analyzing literary works across all media platforms, is one of the key developments. The adaptations of literary works between print and digital and film media can be analyzed using deep learning and enable the researcher to comprehend how the narrative changes between mediums (Costa, 2025). This large-scale analysis of corpus also provides researchers with the option to analyze millions of texts, written at various times, and in various languages, so allowing scholars to examine long-term patterns in language, topics, and literary styles (Bajaj et al., 2025). Computational semantics is also helpful in studying literature since it helps trace the evolution of word and concept meanings over time in historical and cultural environments (Bingenheimer et al., 2017).

The other significant trend is the hybrid intelligence and explainable AI integration. The approaches combine neural networks and symbolic reasoning to generate more transparent and interpretable results without losing the ability of human interpretation in a literary analysis (Boer & Stork, 2024). The AI is supposed to be an assistant system working cooperatively and helping in the realization of patterns, text analysis, and data processing in the process of human scholars giving meaning and critical analysis. The multilingual natural language processing will also make cross-cultural analysis of literature possible, which will allow researcher to compare literature between various languages and cultures (Oduola & Olajuyigbe, 2026).

Classroom pedagogy is another area where AI will revolutionize especially in the area of English and literature. Personalized reading material, writing assistance, and feedback can also be offered to the learners by intelligent tutoring systems and adaptive learning platforms (Adorni & Grosso, 2025). Computational learning tools like sentiment analysis, topic modeling, and network analysis can also be used by the students in the analysis of literary text. Artificial intelligence creative writing aids will be able to assist in poetry and narrative writing and promote the critical assessment of AI-written works (Patil et al., 2025).

The interdisciplinary collaboration, AI literacy, and ethical principles, as well as institutional policies, will be required to enforce responsible and effective AI integration into digital humanities and education (Chun & Elkins, 2023; Minh et al., 2025; Fontenot, 2025).

7. Conclusion

The review summarized the current findings on the topic of Artificial Intelligence (AI) and Information Technology in the field of Language and Literature Studies and identified the main uses, opportunities, challenges, and upcoming directions. The literature demonstrates that AI technologies, especially the Natural Language Processing, deep learning, and generative AI, have turned out to be significant white-collar tools in digital humanities and English study. Such technologies make it possible to conduct a text analysis at a large scale, sentiment analysis, stylometric analysis, language learning support and enable researchers and educators to analyze literary texts more effectively and assist students with language development. Intelligent tutoring systems and machine translation are another type of AI technologies enhancing access to multilingual literature and creating customised learning experiences.

In spite of these advantages, studies have been highlighting the need to include human interpretation in the study of literature and languages. The AI systems can be useful in detecting trends and analyzing massive data, but they do not comprehend the cultural context, ambiguity, and interpretive meaning comprehensively. In such a way, hybrid methods, i.e. a combination of computational analysis and human interpretation is deemed to be more efficient than AI only methods.

Nonetheless, multiple issues were observed, such as the bias of the data, ethical issues, and the unreadiness of teachers and students to work with AI tools. AI applications can be used to strengthen the biases that already exist in literature, and excessive dependence on algorithms can cause a decline in critical thinking and interpretative abilities. Moreover, the problem of unequal access to AI technologies can widen the digital divide unless it is dealt with.

The interdisciplinary collaboration, ethical principles, curriculum redesign, and development of AI literacy are the key areas of responsible AI integration. The humanistic learning should be supported with AI as an application but not substituted with human judgment and creativity.

References

Adorni, G., & Grosso, D. (2025). Generative AI in digital humanities education: A human-centered problem-solving approach. In *2025 IEEE International Conference on Cyber Humanities (IEEE-CH)*. https://doi.org/10.1109/IEEE-CH65308.2025.11279290

Alshehri, M. A., Mosleet, A. F. B. S., Abdellatif, M., & Nemt-allah, M. A. (2025). AI-enhanced learning and cognitive processes in digital humanities: A systematic review of executive functions. *Research Journal in Advanced Humanities*. https://doi.org/10.58256/b3atyz61

Bajaj, A., Karad, V., Panicker, K., Jha, A., V, S. R. M., & Damre, S. (2025). Natural language processing in digital humanities: Analysing literary trends through machine learning. In *2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. https://doi.org/10.1109/ICONSTEM65670.2025.11374443

Balakrishnan, S., Thongprayoon, C., Wathanavasin, W., Miao, J., Mao, M., Craici, I. M., & Cheungpasitporn, W. (2025). Evaluating artificial intelligence bias in nephrology. *Frontiers in Artificial Intelligence, 8*. https://doi.org/10.3389/frai.2025.1525937

Batumalay, M., Masrin, & Multidisiplin, J. (2025). Integration of artificial intelligence in the preservation of regional arts and languages. *ARMADA, 3*(9). https://doi.org/10.55681/armada.v3i9.1756

Bingenheimer, M., Hung, J.-J., & Hsieh, C.-E. (2017). Stylometric analysis of Chinese Buddhist texts. *Journal of the Japanese Association for Digital Humanities, 2*(1). https://doi.org/10.17928/JJADH.2.1_1

Boer, V. D., & Stork, L. (2024). Hybrid intelligence for digital humanities. *arXiv*. https://doi.org/10.48550/arXiv.2406.15374

Chapelle, C. A. (2024). Open generative AI changes a lot, but not everything. *The Modern Language Journal*. https://doi.org/10.1111/modl.12927

Chen, X., & Wu, D. (2024). Automatic generation of multimedia teaching materials based on generative AI. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2024.3378279

Choudhary, T. (2025). Political bias in large language models. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3523764

Chun, J., & Elkins, K. (2023). The crisis of artificial intelligence. *International Journal of Humanities and Arts Computing, 17*(2). https://doi.org/10.3366/ijhac.2023.0310

Costa, M. (2025). Non-linear thinking processes in digital humanities. *DIID—Disegno Industriale Industrial Design*. https://doi.org/10.30682/diid8525h

Davison, R. M., Chughtai, H., Nielsen, P., et al. (2024). The ethics of using generative AI for qualitative data analysis. *Information Systems Journal*. https://doi.org/10.1111/isj.12504

Eboseremen, B. O., Adebayo, A., Essien, I. A., et al. (2021). The role of natural language processing in data-driven research analysis. *International Journal of Multidisciplinary Research and Growth Evaluation, 2*(1). https://doi.org/10.54660/.ijmrge.2021.2.1.935-942

Fang, L. (2025). AI-powered translation and the reframing of cultural concepts in language education. https://doi.org/10.70393/616a736d.323937

Fontenot, J. (2025). Artificial intelligence disclosure in academic nursing. *Nursing Education Perspectives*. https://doi.org/10.1111/nae2.70002

Franceschelli, G., & Musolesi, M. (2025). Assessing value and originality in neural text generation. *arXiv*. https://doi.org/10.48550/arXiv.2502.13207

Fritsche, K., & Münster, S. (2024). Artificial intelligence as teaching and learning content. In *Proceedings of the International Conference on Advanced Information and Communication Technologies (ICATE), 1*(1). https://doi.org/10.33422/icate.v1i1.225

Ganiyeva, D., Aliyeva, N., Karimova, S., et al. (2024). Digital Dickens: AI and the future of classic literature interpretation. In *ICICAT Conference Proceedings*. https://doi.org/10.1109/ICICAT62666.2024.10923482

Ginting, D. A., & Hina, S. (2025). Integrating conversational AI tools in second language education. *Global Education Journal, 2*(3). https://doi.org/10.70062/globaleducation.v2i3.265

Guido, R., Yahya, M., & Sohail, M. (2025). Artificial intelligence in the digital humanities. *Journal of Artificial Intelligence Research and Innovation, 1*(1). https://doi.org/10.29328/journal.jairi.1001006

Guo, Y., Guo, M., Su, J., et al. (2024). Bias in large language models. *arXiv*. https://doi.org/10.48550/arXiv.2411.10915

Hameed, M., & Ali, H. (2025). A comparative stylistic analysis of selected short stories. *Journal of Pure and Applied Sciences, 3*(4). https://doi.org/10.63468/jpsa.3.4.33

Hassanin, S., Bayomy, E. M. A., & Eleleidy, M. A. (2025). Machine learning and NLP for emotional and thematic analysis. *Theory and Practice in Language Studies, 15*(12). https://doi.org/10.17507/tpls.1512.03

Hu, Y. (2024). Automatic generation of poetry using GPT. In *ICDSCNC Conference Proceedings*. https://doi.org/10.1109/ICDSCNC62492.2024.10939868

Hutson, J. (2024). Rethinking plagiarism in the era of generative AI. *Journal of Interactive Computing, 4*(1). https://doi.org/10.54963/jic.v4i1.220

Jaja, D. M. (2025). Artificial intelligence in language education. *Journal of Language Teaching and Research, 16*(6). https://doi.org/10.17265/2159-5526/2025.06.004

Jaleel, A., Aziz, A. R. A., Farid, G., & Bashir, M. Z. (2025). ChatGPT and academic integrity. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1554444

Kakumanu, L. (2025). Harnessing AI in CALL. *Arab World English Journal, 16*(3). https://doi.org/10.24093/awej/vol16no3.8

Kapoor, M. (2025). Navigating the impact of artificial intelligence on medical writing. https://doi.org/10.4103/aca.aca_14_25

Katoch, N. (2025). Addressing bias in AI. https://doi.org/10.55041/ijsrem45370

Kern, R. (2024). Twenty-first century technologies and language education. *The Modern Language Journal*. https://doi.org/10.1111/modl.12924

Khan, M. K., Ferdous, J., Mourshed, G., & Hossain, S. (2025). Use of artificial intelligence in scientific writing.

Levent, V. E., & Özbalkan, U. (2026). Stylometric profiling of Turkish texts. https://doi.org/10.29130/dubited.1728460

Li, Z., Liang, C., Peng, J., & Yin, M. (2024). Generative AI-powered assistance in writing. In *ACM Conference Proceedings*. https://doi.org/10.1145/3613904.3642625

Marco, G., Gonzalo, J., Castillo, R. D., & Girona, M. (2024). Can LLMs challenge fiction authors? In *EMNLP Conference Proceedings*. https://doi.org/10.18653/v1/2024.emnlp-main.1096

Martino, B. D., Cante, L. C., Graziano, M., et al. (2025). Generative AI for historical-archival storytelling. In *IEEE-CH Conference Proceedings*. https://doi.org/10.1109/IEEE-CH65308.2025.11279296

Mamman, M., Youssouf, Y., & PhD. (2026). The algorithm and the soul. *EP-JSS: Journal of Social Sciences and Humanities Research, 2*(2). https://doi.org/10.65150/ep-jsshrs/v2e2/2026-12

McGuire, J., De Cremer, D., & Van de Cruys, T. (2024). Co-creation and self-efficacy in AI collaboration. *Scientific Reports, 14*. https://doi.org/10.1038/s41598-024-69423-2

Minh, H. C., Nguyen, T., Huu, P., & Tran, T. (2025). AI in language education across Asia. https://doi.org/10.58304/ijts.250914

Mohamed, M., & Alosman, K. (2025). Deep learning approaches for Arabic language processing. https://doi.org/10.5455/jjee.204-1711016538

Mubarak, R. (2026). AI-powered translation apps and English language learning. *Arab World English Journal Special Issue on AI, 3*. https://doi.org/10.24093/awej/ai3.22

Mujahid, S., Ramamurthy, U., Baral, D., et al. (2025). Digital humanities using big data and machine learning. In *ICRISET Conference Proceedings*. https://doi.org/10.1109/ICRISET64803.2025.11252497

Naqvi, W. M., Ganjoo, R., Rowe, M., et al. (2025). Critical thinking in the age of generative AI. *Frontiers in Artificial Intelligence, 8*. https://doi.org/10.3389/frai.2025.1571527

Nandan, S., & Khan, S. (2025). Algorithmic assistance and the crisis of authorship. https://doi.org/10.38124/ijisrt/25dec1324

Nursaid, N., Ghaluh, B. M., & Wulandari, E. (2025). AI assistant to reader response theory. https://doi.org/10.1177/13621688251368636

Odili, J. (2025). Harnessing AI in digital humanities. *Journal of Computer Science and Applications, 2*(1). https://doi.org/10.51846/jcsa.v2i1.3506

Oduola, M., & Olajuyigbe, O. (2026). Computational linguistics in Nigeria. *European Journal of English Language and Literature Studies, 14*(1). https://doi.org/10.37745/ejells.2013/vol14n16682

Othman, R. Sh., & Ibrahim, I. M. (2025). Deep learning for NLP. https://doi.org/10.14419/t1xnaq87

Patil, S. R., Avasthi, A., Padma, Y., et al. (2025). Generative AI for creative writing assistance. In *ICONAT Conference Proceedings*. https://doi.org/10.1109/ICONAT66879.2025.11362573

Patnaik, S. K., Bute, A. R., Prasantham, P., et al. (2025). AI and second language acquisition. *Journal of Information Systems Engineering and Management, 10*(39S). https://doi.org/10.52783/jisem.v10i39s.7164

Phadte, A., & Dhore, M. (2025). Sentiment analysis of code-mixed text. In *ICCUBEA Conference Proceedings*. https://doi.org/10.1109/ICCUBEA65967.2025.11283754

Praveen, R., Tharini, C., Sagunthala, R., et al. (2025). Hybrid CNN-RNN model for literary themes. In *ICITEICS Conference Proceedings*. https://doi.org/10.1109/ICITEICS64870.2025.11341703

Ríos-Toledo, G., Posadas-Durán, J. P., Sidorov, G., & Castro-Sánchez, N. A. (2022). Detection of changes in literary writing style. *PLOS ONE, 17*(4). https://doi.org/10.1371/journal.pone.0267590

Rokhshad, R., Tichý, A., Ducret, M., et al. (2025). Ethics and governance of large language models. https://doi.org/10.1016/j.jdent.2025.106187

Sadler, T. D., Mensah, F. M., & Tam, J. (2024). Artificial intelligence and research integrity. https://doi.org/10.1002/tea.21933

Seddik, M. E. (2025). AI-powered language learning tools. *International Journal of Linguistics, Literature and Translation, 8*(3). https://doi.org/10.32996/ijllt.2025.8.3.30

Shahnazari, K., Ayyoubzadeh, S. M., Fazli, M., & Keshtparvar, M. (2025). Network analysis in Persian poetry. *Social Network Analysis and Mining, 15*. https://doi.org/10.1007/s13278-025-01537-5

Sun, Y. (2025). Stylistic evolution of Chinese literary works. https://doi.org/10.61091/jcmcc124-31

Warbeh, M. K., & Verma, P. (2024). NLP in healthcare. In *ICUIS Conference Proceedings*. https://doi.org/10.1109/ICUIS64676.2024.10866525

Woycicki, P., Chamberlain, A., & Borokini, F. (2025). AI and contemporary performance. https://doi.org/10.1080/14794713.2025.2552603

Yadav, D. (2024). AI in literary analysis. *International Journal of Emerging Knowledge Studies, 3*(9). https://doi.org/10.70333/ijeks-03-09-006

Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022). Wordcraft: Story writing with large language models. In *CHI Conference Proceedings*. https://doi.org/10.1145/3490099.3511105

Zhang, J. (2024). Sentiment analysis and linguistic theory. *Studies in Linguistics and Literature, 8*(3). https://doi.org/10.22158/sll.v8n3p265

Zhang, Z., & Huang, X. (2024). Chatbots and vocabulary acquisition. *Heliyon, 10*. https://doi.org/10.1016/j.heliyon.2024.e25370

How to Cite This Paper (APA 7)

Sarmiento, L. N. N. (2026). Artificial intelligence and information technology in language and literature studies: A systematic review of emerging applications. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 105–114). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 14: Smart Mountain Tourism: Artificial Intelligence Applications for Sustainable Community-Based Tourism Development

Dr. Emy Grace B. Patambang
Cavite State University - Silang Camps

Abstract. *This literature review synthesizes evidence on artificial intelligence applications in community-based sustainable mountain tourism development. Mountain regions globally face dual imperatives such as harnessing tourism for economic development while preserving ecological integrity and cultural heritage. AI technologies including machine learning, predictive analytics, GIS/remote sensing, recommendation systems, chatbots, and computer vision, demonstrate significant potential for enhancing visitor experiences, enabling data-driven management, and optimizing resource allocation. This review identifies seven critical AI applications in mountain tourism: tourist demand prediction, environmental impact monitoring, smart trail management, disaster risk prediction, personalized travel recommendations, smart tourism management systems, and visitor flow management. Evidence reveals substantial benefits for sustainable resource management, improved tourist satisfaction, enhanced planning capabilities, and economic opportunities for local communities. However, realizing this potential remains contingent upon addressing critical challenges: technology access disparities in rural areas, digital skills gaps, high implementation costs, data privacy concerns, and risks of AI-driven over-tourism. The review concludes that AI's most transformative role emerges through genuine community participation, transparent benefit-sharing mechanisms, and adaptive governance frameworks that prioritize ecological integrity and sociocultural authenticity. Future research must focus on long-term empirical assessment, community-centered governance frameworks, digital equity strategies, and integration of Indigenous knowledge systems with AI-driven tourism management.*

Keywords: *artificial intelligence, mountain tourism, sustainable development, community-based tourism, environmental monitoring, responsible innovation*

1. Introduction

Mountain tourism has become major branch of the world tourist market with millions of tourists visiting various alpine regions in Asia, Europe, North America and Africa every year. Mountain environments provide some of the most extraordinary biological and cultural assemblages that can make them the top tourist sites. These areas facilitate variety of ecosystems besides the cultural heritage that is only found here, thus making mountains ideal in trekking, mountaineering, skiing and cultural tourism. Nonetheless, (Mandal & Sadhukhan, 2026) mountain areas have unique problems such as poor infrastructure, connectivity, poor health facilities, environmental strain, and untrained human resources, that limit sustainable tourism development.

Mountain tourism brings about a lot of economic gain to the rural and remote communities. Community-based ecotourism has very broad economic gains to the locals and improvements of local economies, but gains are unevenly distributed (Bhatta, 2023). The economic contribution of mountain tourism is not limited to accommodation and food and beverage, mountain tourism encourages the local agricultural commodities, handicraft and cultural tourism, thereby providing job opportunities in various fields.

Sustainable community-based tourism (CBT) is a paradigm shift of mass tourism to development models that emphasize on local involvement, environmental conservation and distribution of benefits that is equitable. The factors of the presence of active participation and environmental education contribute considerably to the development of sustainable tourism, conservation support, and quality of life in the mountain regions. The use of CBT model has proven to be positive in the management and empowerment of the local communities as well as considering the potential of the natural resources, skills of the local communities, socioeconomic and cultural status as well as the preservation of the environment (Pathirana et al., 2025).

The models suggested on how to transform mountain areas into sustainable tourism destinations through development of infrastructure, eco-tourism, involvement of the local people, and supportive policies are sustainable tourism approaches like the Triple Bottom Line (TBL) and the Sustainable Livelihoods Framework. This needs to resolve the dilemma of sustaining ecological integrity and supporting effective socio-economic dynamics to maximize benefits of tourism and minimize effects on distinctive nature and social cultural heritage.

To balance economic growth with environmental conservation, strategic plans should be in line with the UN SDGs of decentralized, community-based, and environmentally friendly tourism plans. Its successful implementation requires carrying capacity considerations, sustainable daily visitor numbers and holistic frameworks that touch on natural, economic and cultural aspects.

Tourism management and the interactive process with community is being fundamentally changed by artificial intelligence and digital technologies. The revolution in the field of AI technologies is bringing a new dimension of human-computer interaction, machine learning and big data analytics to tourism, developing unparalleled possibilities of improving the experiences of visitors and reducing environmental and social effects. These technologies facilitate evidence-based decision-making in terms of destination management, resource allocation, and community empowerment.

The ways of implementing AI into mountain tourism settings are solutions to essential issues of operating the tourism sector. The algorithms of machine learning and virtual assistants and applications of augmented reality offer customized suggestions and multilingual support, whereas (Das, 2024) predictive analytics and ionic-tethered environmental surveillance offer opportunities to make decisions with respect to operational effectiveness and nature conservation. No technology will work without development of human resources and building of capacity in the community.

Objectives of the Literature Review

This literature review synthesizes evidence on artificial intelligence applications in community-based sustainable mountain tourism development. The objectives are to: (1) examine how AI technologies can enhance visitor experiences and tourism management in mountain destinations; (2) analyze AI's potential for supporting local livelihoods, environmental conservation, and cultural heritage preservation; (3) identify challenges including technology access, digital skills gaps, data privacy concerns, and risks of digital promotion exacerbating overtourism; (4) propose future directions for smart mountain tourism systems that balance technological innovation with equitable community benefits and environmental sustainability; and (5) provide recommendations for responsible AI governance in mountain tourism contexts.

2. Mountain Tourism and Community-Based Development

Sustainable tourism development balances the three dimensions namely economic viability, environmental integrity and socioeconomic equity. Such core principles are: (1) managing the number of visitors to the environmental carrying capacity; (2) equitable distribution of tourism revenue among local community; (3) conservation of natural ecosystem and biodiversity; (4) conservation and protection of cultural heritage; (5) good employment of locals; and (6) involvement of community voice in tourism planning and decision making.

Principles of Sustainable Tourism

Sustainable tourism development balances three dimensions: economic viability, environmental integrity, and socioeconomic equity. Core principles include: (1) limiting visitor numbers within environmental carrying capacity; (2) ensuring equitable distribution of tourism revenues to local communities; (3) preserving natural ecosystems and biodiversity; (4) respecting and protecting cultural heritage; (5) providing quality employment for local residents; and (6) incorporating community voice in tourism planning and decision-making.

Community-Based Tourism (CBT) Framework

CBT is a participatory innovative method that puts local communities in the forefront of tourism development, which is in line with UN SDGs of poverty reduction and decent work. (Khartishvili et al., 2020) CBT works based on the principles of the local ownership, participatory decision-making, and sharing of benefits, producing concrete economic, social, and environmental gains and countering the effects of power imbalances and barriers to access to the market. Effective implementation of CBT involves the combination of community practices and particular guidelines and diversified tourism products (Jackson, 2025) .

CBT encourages high distributions of tourism employment in communities, but these patterns tend to be gendered. Tourism is a basic, yet not independent, livelihood enterprise, and its success is determined by the knowledge of tourism in the contexts of local livelihoods and how CBT initiatives allocate employment opportunities fairly. Community involvement, marketing focus, and state intervention are critical to develop sustainable CBT after a crisis or in the case of disruption (Legatzke et al., 2024) .

Environmental, Economic, and Socio-Cultural Sustainability

Environmental Sustainability. Mountain Ecosystems Mountain ecosystems are ecologically delicate and are threatened by climate change, deforestation, loss of biodiversity, and pollution. Sustainable tourism has to be able to reconcile visitor access and conservation of geosites, natural features, and ecosystem services. Lack of strong links between communities and ecotourism in addition to geographic distance and inaccessibility of infrastructure diminish community participation- necessitating more inclusion in decision making and institutionalisation.

Economic Sustainability. Tourism has economic gains, but the benefits are distributed unequally, with entrepreneurs and people close to major trails getting greater rewards compared to farmers, low-paid workers and the marginalised. Local products and linking communities should be addressed, and the problem of economic leakages must be solved (Legatzke et al., 2024). Income earning opportunities need to be distributed fairly among the members of a community especially in gender inequalities.

Socio-Cultural Sustainability. Environmental education campaigns and heritage preservation have a great contribution to sustainable tourism and ensuring that the life of the locals remains to local standards. Nevertheless, Saeed (2025), authenticity within the framework of CBT is still a controversial issue, related to the power structure, benefit sharing processes that affect cultural performances and revenue distribution. The sustainable practices should be considerate of the community agency and traditional knowledge, but promote authentic preservation of cultures.

3. Artificial Intelligence in Tourism

Machine Learning

Machine learning techniques (especially the ones that use natural language processing and sentiment analysis) have proven to be 85-93% accurate in individualized recommendation engines and demand prediction (Ma, 2024). Multilingual support and individualized recommendations can be improved with the help of machine learning algorithms to help improve the travel experiences. Supervised learning Supervised learning algorithms and in particular neural network-based supervised learning approaches have been shown to yield higher tourism impact than alternative supervised approaches.

Predictive Analytics

Predictive analytics and IoT-based environmental surveillance allows the use of data in making decisions with a trade-off between operational efficiency and environmental conservation (Das, 2024). Machine learning algorithms such as Decision Trees, Random Forests, K-Neighbors Classifiers, and Support Vector Machines allow predicting the demand of tourists and classifying their behaviors. ARIMA time series model predicts the alterations in the landscape and visitor numbers in varying time periods.

GIS and Remote Sensing

GIS-based systems offer affordable designs to environmental surveillance such as vegetation health, water quality, and risk (Szafarczy et al., 2025). Geographic Information Systems and Remote Sensing are becoming essential resources of environmental monitoring, and AI and machine learning are now able to improve image classification, cloud-based environments can provide scalable analysis, and UAVs/hyperspectral sensors can be used to view high-resolution monitoring. These technologies aid in the detection of land use changes, biodiversity and hazard risk mapping in mountainous settings.

Recommendation Systems

Hybrid recommendation engines that combine collaborative and content-based filtering with a deep learning system deliver a user satisfaction rate of 78% and system accuracy of 82% with 85% performance rates depending on user personality traits (B et al., 2025). The combination of personality models (MBTI, Big Five) and retrieval-augmented generation systems will allow making the process more personalized and facilitating prediction of preferences and real-time adjustment of recommendations. Suggestion systems in tourism propose acceleration of interdisciplinary growth and agglomeration in particular research centers around the globe.

Chatbots and Virtual Assistants

NLP-based sentiment analysis and hotel recommendation systems prove effective in the reduction of the complexity of selections (Kosta et al., 2025). When applied to the emotional intelligence framework and user-centered design concepts, AI-based chatbots in online travel agencies can lead to a significant increase in user satisfaction, adoption, and further usage of the platform in the future. Intelligent tourism Smart tourism apps, using a domain-specific language to create chatbots with machine learning, have 88.5% recommendation accuracy and 2.3-second response times.

Computer Vision for Environmental Monitoring

Computer vision and generative models have become revolutionary instruments of cultural heritage preservation and interpretation to make cultural objects more accessible and representative. Privacy-conscious AI-based pedestrian monitoring systems based on the YOLO and DeepSORT computer vision provide solutions to visitor flow control and disaster response planning (Okagawa et al., 2025). These systems facilitate real-time crowd analytics and congestion detection and visualization of visitor movement.

4. AI Applications in Mountain Tourism

Tourist Demand Prediction

Proper visitor demand forecasting helps mountain destinations to ensure that resources are properly allocated, staffing is planned and capacity planning of infrastructure is also planned. AI-based reservation systems are machine learning algorithms and predictive analytics that predict visitor demand based on visitor count, weather conditions, special occasions, and other factors (Oyetoro et al., 2025). AI allows dynamic reservations to vary the entry time to prevent overcrowding and provide visitors with smooth experiences.

The combination of XGBoost and machine learning models with Random Forest shows the highest performance in terms of point-of-interest occupancy prediction, and the success increases with shorter prediction horizons. These models determine spatial granularity and time factors having effects on visitor behaviour at various levels of aggregation (Bollenbach, et al. 2024) .

Environmental Impact Monitoring

Machine learning with remote sensing has accuracies of more than 95% in classification of land use and land cover. IoT-based environmental calibrated sensor network environmental monitoring systems with adaptive feedback systems are suitable in minimizing ecological pressure and enhancing environmental performance in eco-tourism sites. Monitoring of environmental parameters in real time identifies the response patterns, which allow conservation management to be proactive (Muafiroh et al., 2025).

Smart Trail Management

Smart trail management combines real time data gathering and dynamic management. Machine learning models are far more effective at finding the similarities of routes, with the highest R2 values (0.448 vs. 0.206) identified, and thus the use of machine learning is the way forward to identify alternative routes that may be chosen to minimize congestion on busy trails by considering visitor preferences (Bollenbach et al., 2025).

Disaster Risk Prediction

Deep learning models that combine CNNs and LSTM networks with the spatial and temporal feature dependencies, respectively, have an accuracy rate of over 90 percent in rating critical floods and landslides risks. The improved AWS networks based on 3x3 km resolution, coupled with IoT sensors and machine learning algorithms, will allow the predictions of hyper-local weather and the forecasting of disasters- especially important in mountainous areas subject to flash floods and landslides (Verma et al., 2025).

Machine learning classifiers such as K-Nearest Neighbors, Random Forest, and Support Vector Machines are used to set rainfall thresholds of landslide early warning systems and KNN had a ROC-AUC of 0.9 and 82% accuracy. Such systems come up with viable, community-based solutions to landslide disaster resiliencies (Menon & Kolathayar, 2025) .

Personalized Travel Recommendation

Hybrid recommendation engines have a user satisfaction score of 78% and 82% system accuracy (B et al., 2025). Integration of personality models allows the making of individualized decisions to assist in preference prediction and real-time adaptation (Arıbaş, E., & Daglarli, 2024). AI-based itinerary generators process the preferences of the users in terms of time, budget and interests and optimize point of interest selection, allocation of time slots and travel path with K-means clustering to reduce the amount of travelling time.

Smart Tourism Management Systems

AI-based big data analytics enable users to tap into extensive streams of data and derive valuable information to discover and provide actionable insights. These technologies are developed to create AI-driven predictive analytics by applying user-centric design, data-driven decision-making, service-dominant logic, and data science. Such systems combine various sources of data, predict demands on a granular level and in the context of time and make personalized suggestions (Lakkarasu et al., 2024).

Visitor Flow and Crowd Management

BiGRU-Attn models with the inclusion of bidirectional GRU units, attention mechanisms, and bidirectional RNNs obtain R2 values of 0.948 when forecasting the tourist flow in high frequencies, and excel specifically in capturing short-term effects in peak hours. XGBoost and Random Forest models are more effective at occupancy prediction of POI, which is more effective as the prediction horizons become shorter when implemented in operations (Wang et al., 2025).

AI-based pedestrian tracking in the form of YOLO and DeepSORT computer vision provide privacy-security tradeoffs to visitor flow management and disaster response management. The systems produce hourly pedestrian counts and flow directions that allow one to visualize temporal patterns of congestion, allowing tourism planning to be done based on data (Okagawa et al., 2025).

5. AI for Sustainable Community-Based Tourism

Local Livelihood Support

Technology and AI enable market access and economic opportunity for local communities. (Purnomo & Purwandari, 2025) MSME empowerment models integrating leadership development, innovation support, and technology integration with local tourism strategies demonstrate that technology integration combined with human resource development enhances tourism development and village sustainability (Lee-Anant & Kungwansith, 2025). Training effectiveness and government support mediate relationships between CBT participation and economic outcomes, including income stability and workforce readiness, emphasizing need for skills-based training and multilingual services.

Tourism Planning and Resource Allocation

Digital Twin technology creates real-time, virtual replicas of destinations combining diverse data streams to generate predictive insights for managers. DTs improve accuracy of tourist flow forecasting, enable smarter decision-making, and enhance

sustainability by optimizing resource use. (Lakkarasu & Sathiri, 2024) AI-driven predictive analytics provide historical, real-time, and anticipated demand at multiple granularity levels and time scales supporting improved allocation decisions.

Environmental Conservation

GIS-based systems provide cost-effective environmental monitoring across vegetation health, water quality, and hazard assessment. (Szafarczyk & Agbasi, 2025) AI and machine learning integration with GIS and remote sensing transforms environmental surveillance through enhanced image classification, cloud-based platforms for scalable analysis, and UAV/hyperspectral sensors enabling participatory conservation approaches.

Cultural Heritage Preservation

Computer vision and generative models serve as transformative tools for cultural heritage preservation, enhancing accessibility and representativeness of cultural assets. These technologies enable virtual reconstruction of historical sites, development of intelligent cultural assistants, and adaptive tours improving visitor experience and promoting inclusion. However, implementation requires sector-specific ethical frameworks, inclusive governance models, and sustainable technological strategies promoting equity and community participation.

Local Business Promotion Using AI Marketing

Smart tourism promotion systems leveraging AI and machine learning achieve 87.5% precision, 83.2% recall, and 85.3% F1-scores for personalized recommendations. AI-enabled local business promotion includes clustering with DBSCAN and K-Means, hybrid recommendaDigital Social Enterprises integrating social mission with entrepreneurial approaches and digital technologies offer significant potential for advancing sustainable rural tourism. (Retnowati et al., 2024) Soft systems methodology enables building relevant and successful models for increasing community engagement in technology-based tourism situations. Participatory design approaches foster shared understanding of community-environment interplay, promoting sustainable tourism.

6. Benefits of AI in Mountain Tourism

Sustainable Resource Management

Smart sensors and IoT devices enable real-time environmental monitoring and proactive conservation measures for biodiversity protection and resource management. (Ashitha et al., 2025) Hyper-local environmental monitoring systems combining IoT sensors with machine learning deliver detailed and reliable weather predictions, enhancing preparedness for extreme weather events.

Improved Tourist Experience

AI-based recommendation systems achieve 89.7% accuracy for destinations, 87.2% for accommodations, 90.3% for itineraries, and 85.6% for activities. AI personalization significantly enhances customer satisfaction and decision-making through customized recommendations, virtual assistants, and real-time guidance systems.

Better Tourism Planning

AI-driven big data analytics empower users to extract meaningful information for actionable recommendations. Digital Twin applications improve accuracy of tourist flow forecasting and enable smarter decision-making, while (Colasante et al., 2024) stakeholder engagement and strategic sustainability approaches guide effective planning and implementation.

Economic Opportunities for Local Communities

Technology integration with leadership development and innovation support enhances MSME empowerment and tourism development (Indarto & Rachmawati, 2025). Creative economy integration with tourism increases tourist length of stay and creative economy actors' income substantially within annual periods despite challenges in standardization and digital market access.

Disaster Preparedness and Risk Management

Deep learning models achieve accuracy exceeding 90% for flood and landslide risk classification, enabling early warning and proactive intervention. (Ashitha et al., 2025) Enhanced weather monitoring systems with hyper-local resolution enable rapid emergency response to disasters (Sheta, 2025). AI solutions enhance pilgrim experience through predictive models for crowd flow management, reducing risks while providing personalized and spiritually satisfying experiences.

7. Challenges and Issues

Technology Access in Rural Areas

Significant skills gaps continue to arise in emerging technologies such as AI and robotics, the main shortcomings being social media, online marketing, and communication potentials. Limitations to infrastructure, inaccessibility of the internet in the rural

regions, and deficiency in digital literacy that impacts on meaningful use of technologies have remained persistent problems in the digital divide (Alikhan & Alikhan, 2025).

Lack of Digital Skills in Communities

The trainings that emphasize the development of practical skills in the areas of customer service and the adoption of digital technologies enhance the confidence and readiness of students to work, but their implementation is still obstructed by the problems with infrastructure and digital literacy (Lubis et al., 2025; Maja, 2025) Transforming effects of digitally-based training models in context-related settings with traditional training, collaborative learning, and adaptative technologies can be seen in terms of their transformative influence notwithstanding the infrastructural limitations.

High Implementation Costs

The implementation of technology demands a lot of money in terms of capital investment, development of infrastructure and maintenance which is very demanding especially in the rural mountain communities which have limited finances. The models of sustainable implementation should be a compromise between technological advanced and economical and large scale.

Data Privacy and Ethics

Algorithms, information privacy breaches, and restructuring of the processes of the production of heritage knowledge require ethical models that put cultural sensitivity and community involvement at the center of their priorities. (Ngan et al., 2026) Privacy paradox in smart tourism unveils the case of tourists whose main concern is privacy, but they are willing to use the technologies that gather sensitive information, implying the lack of privacy protection strategies and lack of user awareness.

The field of big data governance of sustainable tourism is not well developed, and major gaps exist in processing mobile positioning information, preserving compliance of stakeholders and regulatory consistency. According to (Showail et al., 2025), mobile applications in the tourism industries show compliance rate of only 12.1%-33.3% with the laws of personal data protection, which is evidence of systematic failures in the compliance with the principles of privacy among developers.

Risk of Over-Tourism Due to Digital Promotion

Although social media marketing can allow real-time distribution of crowds and control the flow of visitors, it also poses the threat of concentrating the tourist flows at the promoted destinations. To ensure that new hot spots of overtourism and environmental deterioration can be avoided, infrastructure and management needs should be reinforced at the same time when promotion is underway (Kesuma et al., 2026).

8. Future Directions

Smart Mountain Tourism Systems

Future studies should create an all-encompassing smart mountain tourism platform with a synergistic combination of (Kumar, 2025) AI-based personalization with real-time analytics, smart navigation, and intertwined integration of services of accommodation, transport, entertainment, and culture. The systems are to use dynamic intelligent sensors that continuously evaluate the surroundings and predictive maintenance strategies to run sustainable infrastructure with a human in mind.

AI + IoT for Environmental Monitoring

Hyper-local weather forecasting is made possible by enhanced AWS networks that are incorporating IoT sensors together with machine learning algorithms. The systems of the future ought to incorporate various streams of environmental data-atmospheric, hydrological, ecological and allow them to monitor the entire ecosystem and provide proactive conservation and preparedness to natural disasters. The predictive analytics and real-time data processing allow early warning and adaptive management (Ashitha et al., 2025).

AI for Climate Change Adaptation

Renewable energy sources, energy-saving facilities, better water management systems, and eco-tourism promotion are some of the climate-smart tourism technologies that provide strategic avenues of sustainable development. Physics-informed learning and uncertainty quantification in geophysical risk assessment along with quantum computing optimization of sensors placement are the future directions of improved predictive power in climate adaptation (Pasupuleti, 2025).

Community-Based Smart Tourism Platforms

Sustainable smart tourism requires the use of digital platforms that allow communities to participate, promote local business, and distribute benefits equally. Digital Social Enterprises that combine social mission and digital technologies can provide examples of developing sustainable rural tourism (Margiono, 2025). There should be platforms where the sharing of knowledge, market access, financial inclusion, and voice of the community in tourism governance can be achieved.

Policy and Governance for Smart Sustainable Tourism

Sector-specific ethical frameworks that provide community involvement, sustainability in the long-run, and equity are necessary (Sanchez-Martin et al., 2025). The government of Authentic CBT needs to deal with structures of power and mechanisms of benefit sharing, beyond naive conceptions of authenticity. It will need integrated, organized strategies that will focus on the development of infrastructure, digitalization, and community-based tourism development at the same time as specialized marketing. (Astanakulov et al., 2025) It is important to note that (Bekmurzaeva, 2025) data governance, management of environmental loads with the help of digital flow control systems, and meaningful involvement in the local community are important success factors.

9. Conclusion

This literature review is a synthesis of all the evidence on artificial intelligence application in community-based sustainable mountain tourism development. The discussion shows that AI technologies, namely machine learning, predictive analytics, GIS/remote sensing, recommendation systems, chatbots, and computer vision, have a great potential of transformation in personalizing the visitor experiences, data-driven management, resource allocation, and livelihoods in the area. Nevertheless, this potential will only be achieved when key loopholes in the areas of ethical governance, community empowerment, digital equity, and environmental stewardship are addressed. It is that the most promising way forward is a combination of advanced technology strengths and a real community involvement, open systems of benefit sharing, and responsive governance that focuses on ecological integrity and sociocultural authenticity rather than simplistic efficiency indicators.

The gaps in research are in a number of key areas: (1) long-term empirical studies of the real impact of AI applications on destination sustainability and community wellbeing; (2) community-focused models of responsible AI governance in the tourism industry; (3) digital equity strategies to ensure that technology is accessible and benefits marginalized groups, (4) exploring the effects of AI-based personalization on carrying capacity management and overtourism, and (5) how Indigenous knowledge systems can be integrated with AI-based tourism management.

Future Research Recommendations are: (1) participatory action research into the impacts of communities co-designing and governing AI systems in tourism; (2) longitudinal research on the long-term implications of smart tourism technologies on livelihoods, environment, and cultural preservation; (3) the replicability of frameworks governing the ethical use of AI in tourism; (4) research into how quantum computing and advanced physics-informed learning can enhance adaptation to climate change; (5) how digital justice, cultural authenticity, and technological innovation have intersections; and (6) scaling successful small

The synthesized evidence in this review confirms the idea that artificial intelligence, when applied in a fair and responsible manner, can become a strategic tool of promoting sustainable mountain tourism based on community environment. The development in the future should be of an integrative nature where it is not only considered whether the AI technologies may streamline tourism systems, but also whether it can be done fairly, in a sustainable manner, and without infringing on the community agency and environmental boundaries. This orientation integrates the mountain tourism development with the international objectives of sustainable development, resilience to climate and social equity in the globalized world.

References

Alikhan, S., & Alikhan, M. (2025). Assessing telecentres as a strategy for bridging the digital divide in rural communities of developing countries: A narrative review. *Journal of the University Librarians Association of Sri Lanka*, 28(2). https://doi.org/10.4038/jula.v28i2.8113

Arıbaş, E., & Daglarli, E. (2024). Transforming Personalized Travel Recommendations: Integrating Generative AI with Personality Models. *Electronics*, *13*(23). https://doi.org/10.3390/electronics13234751

Ashitha, P. R., Vincent, A., Kumar, S., Benito, A. T., Dev, N. A., Asna, S., & Menon, B. J. (2025). TEMPORA: Intelligent Weather Monitoring and Disaster Forecasting System. *2025 11th International Conference on Smart Computing and Communications (ICSCC)*. https://doi.org/10.1109/ICSCC66177.2025.11233503

Astanakulov, K., Ahmedov, A., Kovalev, D., & Mamasalieva, M. (2025). An integrated approach to sustainable tourism development in the mountain regions of Uzbekistan. *Sustainable Development of Mountain Territories*, *17*(1). https://doi.org/10.21177/1998-4502-2025-17-1-414-426

B, A. K., V, V., John, A., P, A. J., & M, A. A. (2025). Travel-AI: A Comprehensive Intelligent Framework for Personalized Travel Planning, Predictive Analytics, and Secure booking. *International Journal of Scientific Research in Engineering and Management*. https://doi.org/10.55041/ijsrem54925

Bekmurzaeva, R. Kh. (2025). Environmental aspects of tourism development: Analysis of management tools and calculation of natural area pressure. *Ekonomika i Upravlenie: Problemy, Resheniya*. https://doi.org/10.36871/ek.up.p.r.2025.05.14.024

Bhatta, K. (2023). Community-Based Ecotourism and Perceived Economic Impacts: A Study of Rural Settlements around Annapurna Conservation Area, Nepal. *Journal of Engineering Technology and Planning*, *4*(1). https://doi.org/10.3126/joetp.v4i1.58444

Bollenbach, J., Neubig, S., Hein, A., Keller, R., & Krcmar, H. (2024). Enabling active visitor management: local, short-term occupancy prediction at a touristic point of interest. *Information Technology & Tourism*. https://doi.org/10.1007/s40558-024-00291-2

Bollenbach, J., Rebholz, D., & Keller, R. (2025). The road not taken: Representing expert knowledge for route similarities in sustainable tourism using machine learning. *Springer Science+Business Media*. https://doi.org/10.1007/s12525-025-00816-5

Colasante, A., D'Adamo, I., Massis, A. D., & Italiano, S. (2024). An exploratory study of stakeholder views on the sustainable development of mountain tourism. *Sustainable Development.* https://doi.org/10.1002/sd.2878

Das, S. (2024). Application of AI Technology for the Development of Destination Tourism towards an *Intelligent Information System.* https://doi.org/10.46852/0424-2513.3.2024.31

Indarto, B. A., & Rachmawati, M. (2025). Creative Economy as a Driver of Sustainable Tourism in the Slopes of Mount Merbabu. *International Journal of Law Social Sciences and Management. 2*(1). https://doi.org/10.69726/ijlssm.v2i1.86

Jackson, L. A. (2025). Community-Based Tourism: A Catalyst for Achieving the United Nations Sustainable Development Goals One and Eight. *Tourism and Hospitality, 6*(1). https://doi.org/10.3390/tourhosp6010029

Kesuma, W. P., Sugiarto, S., & Kiswantoro, A. (2026). Operational Risk Management Strategy Formulation of Curup Gangsa Lampung. MSJ: Majority Science Journal. 4(1). https://doi.org/10.61942/msj.v4i1.553

Khartishvili, L., Mitrofanenko, T., Muhar, A., & Penker, M. (2020). Issues with Applying the Concept of Community-Based Tourism in the Caucasus. *International Mountain Society, 40*(1). https://doi.org/10.1659/mrd-journal-d-19-00071.1

Kosta, A., Pazari, F., Lili, I., & Greca, S. (2025). The role of artificial intelligence in transforming tourism services: The case of Durrës, Albania. *Geo Journal of Tourism and Geosites, 61*(3). https://doi.org/10.30892/gtg.61308-1518

Kumar, A. (2025). One-Stop Solution for Tourism: An Integrated AI Driven Travel Platform. *International Journal of Scientific Research in Engineering and Management.* https://doi.org/10.55041/ijsrem40799

Lakkarasu, P., & Sathiri, D. (2024). AI-Enabled Big Data Analytics for Digital Tourism Management. *International Journal of Research Publications in Engineering, Technology and Management. 7*(6). https://doi.org/10.15662/ijrpetm.2024.0706021

Lee-Anant, C., & Kungwansith, P. (2025). Policy Shaping and Capacity Building for Sustainable Community-Based Tourism in Thailand's Eastern Economic Corridor. *Suranaree Journal of Social Science.* https://doi.org/10.55766/sjss278596

Legatzke, H., Current, D., & LaPan, C. (2024). Applying the Sustainable Livelihoods Framework to Examine the Efficacy of Community-Based Tourism at Equitably Promoting Local Livelihood Opportunities. *Tourism Planning & Development.* https://doi.org/10.1080/21568316.2024.2391509

Lubis, A. L., Supardi, Fatimah, Z., & Novianti, J. (2025). Skills and Knowledge Development Training for Students in the Tourism Industry at SMKN 1 Bintan Utara. *Jurnal Keker Wisata, 3*(1). https://doi.org/10.59193/jkw.v3i1.320

Ma, S. (2024). Enhancing Tourists' Satisfaction: Leveraging Artificial Intelligence in the Tourism Sector. *Pacific International Journal, 7*(3). https://doi.org/10.55014/pij.v7i3.624

Maja, M. (2025). Perceptions of English first additional language teachers on in-service training for integrating digital skills in rural schools. Interdisciplinary Journal of Rural and Community Studies. 7(1). https://doi.org/10.38140/ijrcs-2024.vol7.1.09

Mandal, J., & Sadhukhan, S. (2026). Development of Mountain-Based Rural Tourism: A Case Study of Dhotrey Village of Darjeeling, West Bengal. *International Journal for Research in Applied Science and Engineering Technology.* https://doi.org/10.22214/ijraset.2026.76910

Margiono, A. (2025). From Village to Cloud: A Framework for the Emergence and Sustainability of Digital Social Enterprises (DSEs) in Rural Tourism. *EAI Endorsed Transactions on Tourism, Technology and Intelligence.* https://doi.org/10.4108/eettti.10107

Menon, V., & Kolathayar, S. (2025). Empirical and machine learning-based approaches to identify rainfall thresholds for landslide prediction: a case study of Kerala, India. *Discover Applied Sciences. 7.* https://doi.org/10.1007/s42452-025-06636-8

Muafiroh, S., Pradono, K. A., Sunandar, P., & Manurung, P. (2025). Application of machine learning and remote sensing in monitoring land use dynamics in tourism area. Remote Sensing Technology in Defense and Environment. 2(1). https://doi.org/10.61511/rstde.v2i1.2025.1761

Ngan, H., Hang, F. M., & Leng, C. C. (2026). Is Smart Scary or Useful? Privacy Calculus in Smart Tourism Technology. *The International Journal of Tourism Research.* https://doi.org/10.1002/jtr.70212

Okagawa, R., Urata, M., & Endo, M. (2025). AI-Driven Human Mobility Monitoring Through Camera Footage in Shirakawa-go: Toward Sustainable Management of a World Heritage Site. *Global Conference on Consumer Electronics.* https://doi.org/10.1109/GCCE65946.2025.11275138

Oyetoro, L. T., Adedotun, K. J., Adekunle, J. A., Khadijat, A. B., & Raji, A. K. (2025). AI-based reservation systems for tourist attractions: A predictive approach to managing seasonal fluctuations and visitor flow. *Journal of African Advancement and Sustainability Studies, 7*(2). https://doi.org/10.70382/sjaass.v7i2.007

Pasupuleti, M. K. (2025). Big Data Analytics for Planetary Systems: AI-Enhanced Geoinformatics and Quantum Models for Predictive Geophysical Risk Assessment. *International Journal of Academic and Industrial Research Innovations (IJAIRI).* https://doi.org/10.62311/nesx/rp-5-aug-2025

Pathirana, M., Silva, M. D., & Warnakula, U. S. (2025). Transforming Knuckles Mountain Range into a sustainable tourism model: A comprehensive framework for ecotourism development. *Journal of Tourism and Himalayan Adventures, 7*(1). https://doi.org/10.3126/jtha.v7i1.80883

Purnomo, S., & Purwandari, S. (2025). A Comprehensive Micro, Small, and Medium Enterprise Empowerment Model for Developing Sustainable Tourism Villages in Rural Communities: A Perspective. *Sustainability, 17*(4). https://doi.org/10.3390/su17041368

Retnowati, R., Wahyudi, E. N., & Sunardi, S. (2024). Redesigning Tourism Community Participation Management: Integrating Digital Technology with the Soft Systems Methodology (SSM) Approach. *International Journal of Engineering Business and Social Science, 2*(4). https://doi.org/10.58451/ijebss.v2i04.152

Saeed, U. (2025). The Symbiotic Relationship Between Power Structures, Benefit Distribution, and Tourist Authentic Experience in Community-Based Tourism. *Innovative Tourism Horizons.* https://doi.org/10.64229/yaj6dg63

Sánchez-Martín, J., Guillén-Peñafiel, R., & Hernández-Carretero, A. (2025). Artificial Intelligence in Heritage Tourism: Innovation, Accessibility, and Sustainability in the Digital Age. *The Heritage, 8*(10). https://doi.org/10.3390/heritage8100428

Sheta, A. (2025). AI Solutions for Enhancing the Pilgrim Journey: Predictive Models for Crowd Flow Management and Real-Time Guidance. *Journal of Computer Science and Information Technology, 2*(2). https://doi.org/10.61424/jcsit.v2i2.5377

Showail, A., Shambour, M. Y., Jaradat, H., Abu-Hashem, M. A., & Aldhubaib, H. A. (2025). Assessing the Compliance of Mobile Applications with Personal Data Privacy Regulations: An Analytical Study. *Engineering, Technology & Applied Science Research, 15*(1). https://doi.org/10.48084/etasr.11214

Szafarczyk, A., & Agbasi, O. (2025). Emerging trends in GIS and remote sensing technologies for environmental monitoring: innovations, applications, and future directions. *Geoinformatica Polonica. https://doi.org/10.4467/21995923gp.25.002.22857*

Verma, L., Kushwaha, K., & Kaur, S. (2025). A Systematic Review of Flood and Landslide Prediction Using Deep Learning. *2025 International Conference on Innovations and Emerging Technologies in AI & Communication Systems (IETACS).* https://doi.org/10.1109/IETACS68750.2025.11385494

Wang, C., Xue, H., & Wang, S. (2025). High-Frequency Tourist Flow Forecasting with Gated Recurrent Units and Attention Mechanisms. *SAGE Open, 15*(1). https://doi.org/10.1177/21582440251358303

How to Cite This Paper (APA 7)

Patambang, E. G. B. (2026). Smart mountain tourism: Artificial intelligence applications for sustainable community-based tourism development. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 115–123). EduHeart Knowledge Network and Publishing, Inc.**

Chapter 15: Environmental Awareness, Misconceptions, Attitudes, and Pedagogical Approaches in High School Science Education: A Literature Review

Jane A. Manarpiis
Cavite State University - Silang Campus

Abstract. *This literature review synthesizes current research on high school students' environmental awareness, misconceptions, attitudes, and effective pedagogical approaches in science education. Despite widespread access to environmental information, students demonstrate moderate to high awareness of key environmental issues such as climate change and pollution, yet substantial knowledge gaps persist regarding underlying mechanisms and system complexity. Geographic context, socioeconomic status, curriculum design, and teacher competence significantly influence environmental awareness development. Common misconceptions including confusion between the greenhouse effect and global warming, renewable energy myths, and recycling oversimplifications remain remarkably persistent even among educated populations, often resulting from inadequate teaching methods and the abstract nature of environmental concepts. While most students express positive environmental attitudes toward sustainability and conservation, a substantial attitude-behavior gap exists, with environmental self-efficacy and social norms emerging as crucial mediators between attitudes and pro-environmental actions. Effective pedagogical approaches including inquiry-based learning, project-based learning, experiential field education, and technology-enhanced strategies demonstrate effectiveness when combined with active learning and community engagement. Teacher quality emerges as the most influential school-level factor in determining educational outcomes. Current research reveals critical gaps: limited longitudinal studies on attitude persistence, insufficient cross-cultural comparisons, and inadequate integration of environmental topics into standardized curricula. Future research must address these limitations through longitudinal investigations, culturally responsive approaches, and systemic educational reforms to create lasting environmental engagement among students.*

Keywords: *Environmental Awareness, Environmental Misconceptions, Environmental Attitudes, Environmental Education, High School Students*

1. Introduction

Environmental education has emerged as a critical priority in 21st-century science curricula, addressing one of humanity's most pressing challenges: fostering informed, engaged citizens capable of addressing complex environmental problems. High school students represent a pivotal demographic—old enough to comprehend sophisticated scientific concepts yet young enough to develop foundational environmental values and behaviors that will shape their adult decision-making (Genovese, 2022). The importance of environmental education extends beyond knowledge acquisition; it encompasses developing critical thinking skills, cultivating environmental awareness, and fostering pro-environmental behaviors that contribute to sustainable societies (Lusuegro et al., 2025).

Despite unprecedented access to environmental information through traditional media, internet sources, and educational institutions, research reveals a paradoxical pattern: while many high school students demonstrate moderate to high awareness of major environmental issues such as climate change and pollution, substantial knowledge gaps persist regarding underlying scientific mechanisms and system complexity (Saro et al., 2025). This awareness-knowledge gap suggests that merely increasing information accessibility is insufficient to develop deep environmental understanding. Additionally, substantial variations exist across geographic contexts, socioeconomic backgrounds, and educational systems, highlighting that environmental education effectiveness depends not only on the availability of information but also on curriculum design, teacher preparation, and instructional quality (Kim et al., 2024).

Understanding student awareness of environmental issues is fundamental because it serves as the foundation for informed environmental decision-making. Students' comprehension of cause-effect relationships, system dynamics, and the interconnections between human activities and environmental consequences directly influences their ability to evaluate environmental policies, interpret scientific evidence, and engage responsibly with environmental challenges (Aksoy, 2022). When students hold incomplete or inaccurate mental models of environmental phenomena, they cannot effectively apply knowledge to novel situations or contribute meaningfully to environmental problem-solving.

Addressing misconceptions is equally critical because research demonstrates that incorrect understandings prove remarkably resistant to change, persisting even among highly educated populations (Aksoy & Erten, 2022). Students frequently confuse related concepts such as the greenhouse effect and global warming, hold inaccurate beliefs about renewable energy systems, and maintain oversimplified understandings of recycling and waste management (Cheong et al., 2014). These misconceptions represent more than isolated knowledge gaps—they actively interfere with learning, limit the effectiveness of subsequent instruction, and lead to flawed environmental decision-making throughout students' lives (Yeh et al., 2017). Traditional approaches of simply providing correct information prove insufficient; remediation requires deliberate pedagogical intervention designed to confront misconceptions with evidence and support construction of accurate mental models.

Examining student attitudes matters fundamentally because attitudes serve as crucial mediators between knowledge and behavior. While most students express positive environmental attitudes and concern for sustainability, a well-documented attitude-behavior gap persists, with environmental self-efficacy, social norms, and emotional connections to nature emerging as critical factors determining whether positive attitudes translate into pro-environmental actions (Zhang & Cao, 2025). Understanding the mechanisms linking attitudes to behaviors, the demographic and cultural variations in environmental attitudes, and the factors that strengthen these linkages is essential for designing interventions that produce lasting behavioral change rather than temporary awareness campaigns (Genovese, 2022).

Teacher quality and pedagogical effectiveness warrant investigation because school-based instruction represents one of the most systematic opportunities to reach large numbers of students during formative years. Teacher preparation, environmental knowledge, instructional strategies, and attitudes fundamentally shape whether environmental education develops deep understanding or superficial awareness (Lusuegro et al., 2025). Inquiry-based learning, project-based learning, and experiential education demonstrate effectiveness in developing both environmental knowledge and pro-environmental attitudes when implemented with adequate teacher support and community engagement (Abdullah et al., 2025). Identifying which pedagogical approaches prove most effective under varying contextual conditions enables educators to maximize the impact of limited instructional time and resources.

Objectives of the Literature Review

This literature review pursues four primary objectives:

1. Synthesize Current Understanding of Student Environmental Awareness;
2. Map the Landscape of Environmental Misconceptions and Their Causes;
3. Analyze Environmental Attitudes and the Attitude-Behavior Relationship; and
4. Evaluate Effective Pedagogical Approaches and Identify Implementation Gaps.

II. Students' Awareness of Environmental Issues

Overview of Students' Knowledge of Key Environmental Topics

Research on high school students' awareness reveals varying levels of understanding across critical environmental topics. Students demonstrate moderate to high awareness of climate change and pollution, though their comprehension of underlying mechanisms and specific impacts remains uneven (Kim et al., 2024). Studies comparing students in developed and developing countries show significant differences in environmental knowledge, with climate change awareness ranging from moderate to high levels depending on geographic context and educational exposure (Saro et al., 2025).

A comprehensive assessment across junior high school students in Indonesia found that students had a moderate level of climate change literacy, with the attitude indicator achieving the highest percentage at 48%, reflecting good awareness and concern toward climate change issues (Anggraini et al., 2025). However, knowledge gaps persist particularly regarding biodiversity literacy, with research showing that students often demonstrate only functional, structural, or nominal literacy levels rather than the desired multidimensional literacy (Argiyanti et al., 2024). Despite media exposure and accessible environmental education resources, significant gaps exist between awareness and the ability to translate this knowledge into sustainable behaviors (Saro et al., 2025).

Levels of Awareness Across Different Contexts

Geographic and socioeconomic variations significantly influence student environmental awareness. Students in developed countries tend to have higher baseline knowledge about climate change causes and effects compared to peers in developing nations, yet this knowledge advantage does not necessarily translate to stronger pro-environmental behaviors (Kim et al,, 2024). Urban versus rural contexts also matter substantially—students in urban areas report higher awareness due to increased media exposure and visible environmental degradation, though rural students demonstrate strong practical knowledge about local environmental issues (Sapanova et al., 2023).

Cross-cultural comparisons reveal that environmental awareness is mediated by cultural values and educational priorities. Indonesian students consistently demonstrate higher levels of action, concern, belief, attention, and policy awareness compared to Korean peers, despite Koreans showing higher understanding of climate change causes and effects (Kim et al, 2024). Research in Kazakhstani high schools showed that while students maintained strong positive environmental attitudes, their awareness of environmental issues was not particularly strong, indicating that cultural context shapes both knowledge and attitudes differently (Sapanova et al., 2023).

The type of school and curriculum integration also determines awareness levels. Students attending schools with comprehensive environmental education programs and the Adiwiyata Program showed significantly higher environmental awareness than those in schools without structured environmental curricula. In New Jersey, high school students exposed to a new climate change curriculum demonstrated average knowledge scores exceeding 90%, while the same assessment at other schools revealed gaps in students' understanding of sustainability topics (Shendell et al., 2023).

Factors Influencing Awareness

Curriculum design and content integration represent critical determinants of student environmental awareness. When environmental topics are integrated across multiple subjects rather than confined to science classes, students demonstrate more comprehensive understanding (Asi & Fauzan, 2024). Effective curriculum design that contextualizes environmental concepts in real-world scenarios significantly enhances student awareness (Putri et al., 2025). Conversely, students in regions where environmental education lacks formal curricular support or is treated as supplementary content show lower awareness levels (Auliayanti et al., 2025).

Teacher competence and instructional quality substantially influence students' environmental knowledge acquisition. Teachers with strong environmental science background, professional development in climate education, and confidence in teaching these topics create more effective learning environments (Pitaloka, 2025). However, research across Indonesia, Nigeria, and Rwanda reveals that many educators lack adequate training in environmental education, creating barriers to effective instruction (Morakinyo et al., 2025). Teacher attitudes toward environmental issues and their own environmental literacy also directly affect the quality of environmental education delivered (Liu et al., 2015).

Media exposure and community involvement serve as powerful, sometimes competing, influences on student awareness. Students report acquiring environmental knowledge primarily from television and radio (44%), internet sources (22%), school instruction (27%), and family input (7%) (Kuruppuarachchi et al., 2023). Unfortunately, misinformation from social media and entertainment media can reinforce misconceptions alongside factual learning (Lewandowsky, 2020). Community engagement through environmental organizations, local conservation projects, and partnerships with schools significantly strengthens student awareness and transforms passive knowledge into active environmental understanding.

Experiential learning opportunities, including field trips to ecologically sensitive areas, direct engagement with environmental problems, and community-based conservation projects, prove particularly effective for enhancing awareness. Students who participate in hands-on environmental activities demonstrate measurably higher awareness levels and greater retention of environmental knowledge compared to peers receiving only classroom instruction.

III. Common Misconceptions in Environmental Science

Typical Misconceptions

Research consistently identifies persistent misconceptions that hinder students' understanding of fundamental environmental and climate concepts. A major misconception involves confusing the greenhouse effect with global warming or ozone depletion, where students often fail to distinguish between these related but distinct phenomena (Aksoy & Erten, 2022). Pre-service teachers, despite their advanced education level, demonstrated significant misconceptions about global warming causes, consequences, and the greenhouse effect mechanism—areas that require precise scientific understanding (Aksoy & Erten, 2022).

Misconceptions about renewable energy are widespread among secondary students. Students hold inaccurate beliefs regarding the advantages and disadvantages of different energy sources, failing to understand the true environmental costs and benefits of renewable versus fossil fuel energy systems (Cheong et al., 2014). Specifically, students often underestimate the complexity of renewable energy implementation and overestimate its immediate scalability as a climate solution (Yeh et al., 2017).

Recycling myths represent another category of persistent misconceptions. Students frequently believe that recycling solves waste problems entirely, without understanding its limitations or the importance of waste reduction and reuse hierarchies (Amin et al., 2019). Research on energy literacy revealed that students hold numerous misconceptions about energy transformation, conservation, and its relationship to carbon emissions, often confusing energy availability with energy use patterns (Yeh et al., 2017).

Additional misconceptions include oversimplified understandings of climate causation, where students attribute climate change solely to one factor (such as carbon dioxide) while ignoring complex system interactions (Breslyn, 2017). Students also frequently demonstrate scale-related misconceptions, struggling to understand processes occurring at different temporal and spatial scales, from local pollution to global climate system.

Causes of Misconceptions

Lack of conceptual clarity in teaching materials and curricula represents a primary source of student misconceptions. When textbooks or educational media present concepts in simplified or inaccurate ways, students internalize these flawed understandings (Aksoy & Erten, 2022). The abstract nature of environmental and climate concepts makes them particularly vulnerable to misunderstanding when teaching relies solely on verbal explanation without concrete models or visual representations.

Inadequate teaching methods contribute significantly to misconception formation and persistence. Traditional lecture-based approaches that fail to engage students in hypothesis formation, evidence evaluation, and reasoning do not challenge existing misconceptions. Teachers who lack sufficient training in environmental science or climate education may inadvertently reinforce misconceptions or fail to recognize and address student misunderstandings (Pitaloka, 2025).

Misinformation from social media and popular culture creates competing narratives that students must navigate. Social media platforms amplify climate change denial narratives, alternative explanations for environmental phenomena, and oversimplified solutions that undermine scientific understanding (Lewandowsky, 2020). Students' prior knowledge and life experiences, while providing foundations for new learning, can also encode misconceptions that persist if not explicitly addressed (Breslyn, 2017).

The psychological tendency toward cognitive shortcuts leads students to adopt intuitive but incorrect explanations for complex phenomena like greenhouse gas accumulation and radiative forcing.

Impact of Misconceptions on Learning Outcomes and Decision-Making

Misconceptions directly impair students' ability to engage in critical thinking and informed decision-making about environmental issues. When students hold inaccurate understandings of renewable energy systems, for instance, they cannot evaluate energy policy proposals rationally or support evidence-based solutions (Cheong et al., 2014). The persistence of these misconceptions into adulthood means that entire populations make environmental decisions based on flawed understanding, limiting the effectiveness of individual and collective climate action (Yeh et al., 2017).

Research demonstrates that knowledge alone does not predict behavior, partly because misconceptions interfere with the integration of new information into coherent understanding (Zhang & Cao, 2025). Students who hold misconceptions about the environmental impact of different actions may engage in behaviors they believe are helpful while actually contributing to environmental problems. Furthermore, misconceptions can generate learned helplessness if students develop incorrect beliefs that environmental problems are unsolvable or that individual action is ineffective (Mutlu & Nacaroğlu, 2019).

The remediation of misconceptions requires deliberate instructional intervention. Traditional approaches that simply provide correct information prove insufficient—students often maintain misconceptions even when exposed to contradictory evidence (Adler, 2015). Effective approaches involve explicit identification of misconceptions, engagement with the evidence underlying correct explanations, and repeated opportunities to apply corrected understanding to novel situations.

IV. Students' Attitudes Toward Environmental Issues

Studies on Environmental Attitudes

Research on student attitudes toward sustainability and conservation reveals predominantly positive environmental attitudes, though with considerable variation by context and demographic factors. A comprehensive study of senior high school STEM students found that participants demonstrated positive environmental attitudes (mean = 3.57 on a scale), positive environmental awareness ($\mu = 3.60$), though ecological behaviors remained more neutral ($\mu = 3.40$) (Saro et al., 2025). Similar findings across multiple studies indicate that most students express concern about environmental issues and support sustainability principles, yet translate these attitudes into action at lower rates (Cabras & Israel, 2024).

The relationship between climate change attitude and biodiversity conservation practices demonstrates the interconnected nature of environmental attitudes. Research with high school students revealed that environmental sustainability awareness partially mediates the relationship between climate change attitude and biodiversity conservation practices, with very high levels of climate change attitude (beyond the scale ceiling) correlating with high biodiversity conservation engagement (Cabras & Israel, 2024). Students' attitudes toward renewable energy tend to be positive when presented with accurate information, with high school students endorsing the importance of wind, solar, and other clean energy sources.

However, research also identifies indifferent or apathetic attitudes among subsets of students, particularly in regions where environmental education has not been prioritized or where immediate survival concerns overshadow sustainability considerations (Morakinyo et al., 2025). In some contexts, students express negative attitudes rooted in eco-anxiety—worry about environmental futures coupled with perceived powerlessness—which, while indicating environmental concern, may inhibit positive action if not paired with efficacy beliefs (Embate, 2025).

Relationship Between Attitudes and Behavior

The attitude-behavior gap represents one of the most significant findings in environmental education research. While students express strong pro-environmental attitudes, their actual environmental behaviors remain considerably less consistent (Genovese, 2022). High school students demonstrate very high awareness and positive attitudes toward environmental conservation but exhibit only "good" levels of actual environmental practices, revealing a substantial implementation gap (Escatron et al., 2023).

Explicit attitudes (measured through self-report surveys of environmental beliefs) predict pro-environmental behavior more strongly than implicit attitudes (unconscious associations with environmental concepts), though both contribute to behavioral outcomes (Paciello et al., 2025). The strength of the attitude-behavior relationship varies significantly depending on contextual factors—students who experience emotional connections to nature demonstrate stronger behavioral alignment with their stated attitudes compared to those with only cognitive environmental knowledge.

Environmental self-efficacy—students' beliefs that their actions can meaningfully address environmental problems—functions as a critical mediator between attitudes and behavior (Zhang & Cao, 2025). When students hold strong pro-environmental attitudes but lack efficacy beliefs (the conviction that their actions matter), they are less likely to engage in conservation behaviors. Conversely, developing efficacy beliefs while maintaining positive attitudes substantially increases behavioral engagement (Nurkhin et al., 2025).

Social norms powerfully moderate the attitude-behavior relationship. Students are significantly more likely to translate pro-environmental attitudes into behavior when they perceive that their peers, families, and communities value environmental action. This social influence effect demonstrates that promoting pro-environmental behavior requires attention to group dynamics and social expectations, not merely individual attitude change (Maulana et al., 2024).

Influence of Cultural, Social, and Economic Factors on Attitudes

Cultural values and worldviews fundamentally shape environmental attitudes. Students from cultures emphasizing interconnectedness with nature and collective responsibility tend to develop more pro-environmental attitudes earlier than those from cultures emphasizing individual rights and human dominion over nature (Lusuegro et al., 2025). Research across diverse international contexts reveals that students' environmental attitudes reflect their cultural backgrounds, educational traditions, and societal priorities regarding sustainability.

Socioeconomic status influences environmental attitudes significantly. Students from higher-income families often demonstrate higher environmental awareness and more developed environmental attitudes, partly due to greater access to environmental education and media, and more resources for pro-environmental behaviors (Mónus, 2022). However, lower-income students sometimes demonstrate strong practical knowledge of environmental issues affecting their communities and high motivation for local environmental improvement (Bellotti et al., 2025).

Gender differences consistently emerge in environmental attitude research, with female students generally showing significantly higher environmental concern and stronger pro-environmental attitudes compared to male peers across multiple cultural contexts. These gender differences persist into adulthood and influence policy engagement and voting patterns on environmental issues (Calculli et al., 2021).

Family influences and parental modeling shape student environmental attitudes more powerfully than many school-based interventions. Parents who demonstrate environmental concern through behaviors and discussions cultivate similar attitudes in their children. The quality of students' early socialization experiences with nature—outdoor time, family environmental engagement, and community participation—predicts later environmental attitudes and behaviors (Polgár et al., 2025).

Religious and ethical frameworks influence environmental attitudes substantially in some contexts. Research on Islamic education demonstrates that when environmental stewardship is framed within religious and ethical principles emphasizing accountability and responsibility, students develop notably stronger environmental attitudes and behaviors (Fitroh et al., 2024). Educational programs explicitly linking environmental action to students' moral and ethical frameworks produce more sustainable attitude and behavior changes (Clark, 2024).

V. Pedagogical Approaches to Address Awareness, Misconceptions, and Attitudes

Effective Teaching Strategies

Inquiry-based learning (IBL) represents one of the most thoroughly researched and effective approaches for environmental science education. When structured to guide students through questioning, hypothesis formation, experimentation, and evidence evaluation, inquiry-based learning significantly enhances both conceptual understanding and environmental awareness (Abdullah et al., 2025). Students in inquiry-based environmental programs demonstrate stronger understanding of scientific concepts, improved ability to apply knowledge to real-world issues, and greater motivation to engage with environmental problems (Ningsih et al., 2026). IBL approaches prove particularly effective for addressing misconceptions because they require students to confront evidence that contradicts their initial understanding and construct more accurate mental models (Hikmah et al., 2025).

Project-based learning (PBL) creates powerful contexts for environmental learning by engaging students in authentic, personally relevant problems. A three-month PBL intervention on climate change and sustainability with Grade 10 learners resulted in moderate to high levels of knowledge and awareness, with students developing holistic understanding extending beyond carbon footprints to include biodiversity conservation and personal accountability (Gara et al., 2025). Project-based approaches foster critical thinking, problem-solving skills, and environmental awareness by requiring students to investigate real environmental issues, gather and analyze data, develop solutions, and communicate findings. The integration of PBL with ecological intelligence and problem-based learning creates particularly powerful conditions for developing environmental problem-solving abilities.

Experiential and immersive learning activities directly engage students with environmental concepts through hands-on experience. Field trips to ecologically significant locations, environmental monitoring projects, and outdoor classroom experiences produce measurable increases in environmental awareness and understanding. Research on nature-based learning demonstrates that direct experiences with natural environments strengthen emotional connections to nature and foster lasting pro-environmental attitudes (Polgár et al., 2025). The integration of technology with experiential learning—such as using digital tools alongside field investigations—creates hybrid learning environments that accommodate diverse learning styles while maintaining authentic environmental engagement (Sari, 2025).

Integration of Information and Communication Technologies (ICT) enhances environmental education effectiveness. Technology-enhanced approaches including digital simulations, augmented reality applications, and multimedia presentations increase student engagement and comprehension of abstract environmental concepts. Interactive web-based modules, simulation tools, and digital learning games make complex environmental systems more tangible and enable students to explore scenarios and consequences dynamically. However, technology integration proves most effective when combined with active learning strategies rather than used as a delivery mechanism for passive content consumption (Jashari & Shemshiu, 2024).

Contextualized and interdisciplinary approaches significantly enhance environmental education outcomes. When environmental content is integrated across subject areas (science, social studies, language arts, mathematics) and connected to

local environmental issues that students directly experience, learning becomes more meaningful and transferable. Community-based environmental education that links school learning with authentic local environmental improvement projects increases both knowledge retention and behavioral engagement (Morakinyo et al., 2025). Content and Language Integrated Learning (CLIL) approaches demonstrate that environmental education can be effectively delivered through multiple languages while promoting both content knowledge and language competencies.

Role of Teachers in Correcting Misconceptions and Shaping Attitudes

Teachers serve as critical agents in identifying and remediating student misconceptions. Effective teachers diagnose common misconceptions in their specific student populations, explicitly address these misconceptions through targeted instructional strategies, and monitor student thinking to identify remaining misunderstandings (Aksoy & Erten, 2022). Research on teacher professional development reveals that when educators receive training in environmental science content, misconception research, and evidence-based instructional approaches, they significantly improve their effectiveness in promoting student learning (Lusuegro et al., 2025).

Teacher attitudes, knowledge, and environmental literacy directly influence student learning outcomes. Studies demonstrate that teachers with strong environmental science background, positive attitudes toward environmental issues, and confidence in their ability to teach environmental content create more engaging and effective learning environments (Liu et al., 2015). Conversely, teachers uncertain about climate science or lacking environmental education training tend to avoid comprehensive environmental instruction or inadvertently reinforce misconceptions (Pitaloka, 2025). Professional development programs that combine content knowledge enhancement with pedagogical strategies produce the most substantial improvements in teacher effectiveness.

Teachers play vital roles in shaping student attitudes toward environmental issues beyond merely conveying information. When teachers model genuine environmental concern, discuss the values underlying environmental action, and create classroom environments supporting environmental advocacy, students develop stronger pro-environmental attitudes (Clark, 2024). Research on critical climate awareness demonstrates that teachers who help students understand climate change as both a scientific and sociopolitical process, with emphasis on justice implications and human agency, cultivate more sophisticated environmental attitudes and commitment to action (Clark, 2024).

Inquiry and discussion facilitation skills prove essential for environmental education effectiveness. Teachers skilled in asking probing questions, facilitating peer discussion, and supporting student-led investigations create environments where students develop deeper understanding and stronger ownership of environmental learning. The role of teachers in environmental education has evolved from content delivery to facilitating student construction of environmental understanding, guiding ethical reasoning about environmental issues, and supporting students' development of environmental identity.

Case Studies of Successful Interventions

The Adiwiyata Program in Indonesia demonstrates how systematic school-wide environmental initiatives enhance student awareness and behavior. Schools implementing structured programs including "Healthy Classroom, Healthy Student," "Green School," and "Friday Cleanup Day" fostered significantly higher student environmental awareness through habituation of caring attitudes and actual environmental actions. Even where conceptual gaps remained in environmental knowledge, students exhibited strong pro-environmental behavior sustained through school culture and routine practices (Wibowo et al., 2026).

Problem-Based Learning with Socio-Scientific Issues (SSI) and Education for Sustainable Development (ESD) proved highly effective in Indonesian high schools. Integration of PBL with contextualized SSI-ESD modules on climate change resulted in significant improvements in both chemical literacy and environmental awareness compared to conventional teaching approaches, with large effect sizes (0.282) demonstrating substantial learning gains (Putri et al., 2025). The integration of local environmental issues such as forest fires and haze within curriculum created meaningful connections between abstract concepts and students' lived experiences (Ramadanty et al., 2025).

Experiential field-based learning demonstrated effectiveness in Peru through a primatology project-based intervention. Using participatory teaching methods combining experiential learning and inquiry-based education, high school students achieved 100% recognition of learning about local environmental issues, conservation, and research methods (Ramírez et al., 2024). The integration of scientific field techniques with local conservation challenges positioned students as researchers addressing real environmental problems, profoundly influencing their environmental engagement (Ramírez et al., 2024).

Augmented Reality (AR) technology integration with environmental content showed promise in enhancing student awareness. High school students who read AR-supported environmental texts demonstrated significantly higher environmental awareness and more positive attitudes toward environmental issues compared to peers using traditional paper-based texts (Koparan, 2025). The visual engagement afforded by AR technology made environmental concepts more tangible and memorable (Koparan, 2025).

Systemic approaches combining teacher professional development, curriculum redesign, and community partnerships produced the most comprehensive improvements. Schools implementing multi-faceted environmental education reforms—including teacher training, integration of environmental topics across curricula, community engagement, and institutional support—achieved measurable gains in student environmental knowledge, attitudes, and behaviors that proved more sustainable than single-intervention approaches (Lusuegro, 2025).

VI. Challenges and Gaps in Current Research

Limited Longitudinal Studies on Attitude Change

One of the most significant limitations in environmental education research is the scarcity of longitudinal studies tracking attitude and behavior change over extended periods. Most existing research employs cross-sectional designs or short-term pre/post-test interventions that capture immediate learning gains but cannot determine whether environmental awareness, attitudes, and behaviors persist beyond the intervention period (Genovese, 2022). Without long-term follow-up data, educators and policymakers cannot reliably assess whether the positive attitudes formed in environmental education programs translate into sustained pro-environmental behaviors in students' adult lives.

This gap is particularly critical given the attitude-behavior gap widely documented in environmental education research. While interventions successfully cultivate positive environmental attitudes in the short term, the durability of these changes and their translation into lifelong sustainable practices remain largely unknown. Longitudinal cohort studies following students from secondary education through adulthood would provide essential evidence regarding which pedagogical approaches produce lasting behavioral change and whether early environmental education investments yield long-term returns in terms of societal environmental action (M.a. et al., 2024).

Need for Cross-Cultural Comparisons

Environmental education research remains heavily concentrated in specific geographic regions, particularly developed nations in North America, Europe, and East Asia. Cross-cultural research comparing environmental awareness, attitudes, and effective pedagogical approaches across diverse cultural contexts remains limited, despite evidence that cultural values fundamentally shape environmental understanding and behavior (Lusuegro et al., 2025). Existing comparative studies have revealed substantial differences in how students from different cultural backgrounds understand environmental issues and respond to environmental education interventions.

Research comparing Indonesian and Korean students demonstrated that despite Korean students' superior understanding of climate change causes and effects, Indonesian students exhibited higher levels of environmental action, concern, and policy awareness (Kim et al., 2024). This finding suggests that culturally-specific educational approaches may be more effective than globally standardized curriculum templates. However, systematic research examining how educational systems in different countries address environmental issues, train teachers, and engage families and communities remains inadequate. Expanding cross-cultural environmental education research would enable the identification of culturally responsive pedagogical strategies and climate communication approaches suited to specific regional contexts.

Insufficient Integration of Environmental Issues in Standardized Curricula

Despite growing recognition of environmental education's importance, integration of environmental topics into standardized national curricula remains inconsistent and often superficial across many educational systems. Research examining environmental content in national curricula reveals that environmental education frequently remains compartmentalized within science classes rather than integrated across subject areas, limiting students' opportunities to develop interdisciplinary environmental understanding (Asi & Fauzan, 2024). In many regions, environmental topics are treated as supplementary rather than foundational content, competing with other curricular priorities for limited instructional time.

The gap between environmental education research and curriculum policy is particularly pronounced in developing nations where resource constraints, teacher shortages, and competing educational priorities limit the feasibility of comprehensive environmental curriculum integration (Morakinyo et al., 2025). Even in developed countries with well-established environmental education standards, implementation gaps persist—teachers report uncertainty about how to integrate environmental topics effectively, and standardized assessments often do not evaluate environmental knowledge, creating misalignment between educational recommendations and accountability systems (Auliayanti et al., 2025).

Insufficient professional development for teachers represents another critical barrier to effective environmental curriculum integration. Research consistently demonstrates that teachers' own environmental knowledge, attitudes, and confidence substantially influence the quality of environmental education they deliver (Pitaloka, 2025). However, systematic, evidence-based teacher preparation programs specifically targeting environmental science education remain limited in many countries, leaving educators underprepared to address student misconceptions, integrate environmental topics across curricula, or employ effective pedagogical strategies (Liu et al., 2015).

VII. Conclusion

This literature review synthesizes current research on student environmental awareness, misconceptions, attitudes, and effective pedagogical approaches in high school science education. Research demonstrates that while most high school students express positive environmental attitudes and demonstrate moderate to high awareness of key environmental issues, substantial knowledge gaps persist—particularly regarding underlying mechanisms, system complexity, and connections between human activities and environmental consequences. Geographic context, socioeconomic status, school resources, and educational

curriculum design significantly influence the development of environmental awareness, with developed nations and schools with comprehensive environmental programs showing higher baseline knowledge levels than resource-limited contexts.

Misconceptions remain remarkably persistent even among highly educated populations, with pre-service teachers demonstrating significant misunderstandings about greenhouse gas mechanisms, renewable energy systems, and climate causation. These misconceptions reflect both inadequate teaching methods and the abstract nature of environmental concepts, which resist understanding through verbal explanation alone (Vörös, 2020). Critically, knowledge alone does not predict pro-environmental behavior—the attitude-behavior gap persists as one of the defining challenges in environmental education, with environmental self-efficacy, social norms, and emotional connection to nature emerging as crucial mediators between attitudes and actions.

Effective pedagogical approaches share common characteristics: engagement with authentic environmental problems, opportunities for hands-on investigation and experimentation, integration of environmental content across subject areas, and connections to students' local environments and real-world concerns. Inquiry-based learning, project-based learning, experiential field education, and technology-enhanced approaches all demonstrate effectiveness when combined with active learning strategies and community engagement. Teacher quality emerges as the most influential school-level factor in determining educational effectiveness—teachers' environmental knowledge, confidence, attitudes, and willingness to employ evidence-based instructional strategies directly shape student learning outcomes.

References

Abdullah, M., Kusuma, J., & Abrar, R. N. (2025). Enhancing middle school students' scientific literacy through inquiry-based learning in environmental education. *Journal of Science and Mathematics Education, 1*(4), 338. https://doi.org/10.70716/josme.v1i4.338

Adler, I., Zion, M., & Mevarech, Z. R. (2015). The effect of explicit environmentally oriented metacognitive guidance and peer collaboration on students' expressions of environmental literacy. *Journal of Research in Science Teaching, 52*(3), 439–463. https://doi.org/10.1002/tea.21272

Aksoy, A. C. A., & Erten, S. (2022). A four-tier diagnostic test to determine pre-service science teachers' misconception about global warming. *Journal of Baltic Science Education, 21*(5), 747–761. https://doi.org/10.33225/jbse/22.21.747

Amin, M., Permanasari, A., & Setiabudi, A. (2019). Strengthening student environmental literacy through low-carbon education teaching materials. *Journal of Physics: Conference Series, 1280*(3), 032011. https://doi.org/10.1088/1742-6596/1280/3/032011

Anggraini, M. L., Destiansari, E., & Meilinda, M. (2025). Climate change literacy ability profile in junior high school students in Ogan Ilir Regency. *JPBIO: Jurnal Pendidikan Biologi, 10*(2), 5097. https://doi.org/10.31932/jpbio.v10i2.5097

Argiyanti, A., Kusnadi, K., & M., N. K. (2024). Biodiversity literacy level of students in a senior high school. *BIO-INOVED: Jurnal Biologi-Inovasi Pendidikan, 6*(3), 20035. https://doi.org/10.20527/bino.v6i3.20035

Asi, N., & Fauzan, A. (2024). Raising environmental awareness through English learning materials. *IOP Conference Series: Earth and Environmental Science, 1421*(1), 012026. https://doi.org/10.1088/1755-1315/1421/1/012026

Auliayanti, K., Weni, E., Zainiyah, I. F., Biney, M., & Shrotriya, S. (2025). High school students' awareness of environmental issues in Indonesia: A mini review. *Environmental Research and Planetary Health, 1*(2), 563. https://doi.org/10.53623/erph.v1i2.563

Bellotti, M., Duradoni, M., Fiorenza, M., & Guazzini, A. (2025). Between pro-environmental identity and attitudes: A PRISMA systematic review of the relationship between connectedness to nature and pro-environmental attitudes for sustainability. *Sustainability, 17*(5), 2140. https://doi.org/10.3390/su17052140

Breslyn, W. (2017). An empirically based conditional learning progression for climate change. *Science Education International, 28*(3), 5. https://doi.org/10.33828/sei.v28.i3.5

Cabras, J. L., & Israel, G. F. G. (2024). The relationship between climate change attitude and biodiversity conservation practices: The mediating role of environmental sustainability awareness. *International Journal of Environment and Climate Change, 14*(8), 4339. https://doi.org/10.9734/ijecc/2024/v14i84339

Calculli, C., D'Uggento, A. M., Labarile, A., & Ribecco, N. (2021). Evaluating people's awareness about climate changes and environmental issues: A case study. *Journal of Cleaner Production, 129244.* https://doi.org/10.1016/j.jclepro.2021.129244

Cheong, I. P., Johari, M., Said, H., & Treagust, D. F. (2014). What do you know about alternative energy? Development and use of a diagnostic instrument for upper secondary school science. *International Journal of Science Education, 36*(16), 2767–2794. https://doi.org/10.1080/09500693.2014.976295

Clark, H. F. (2024). Critical climate awareness as a science education outcome. *Science Education, 108*(1), 1–25. https://doi.org/10.1002/sce.21896

Embate, M. J. (2025). Climate change awareness, attitude, and its influence on eco-anxiety among high school students. *Psychology and Education: A Multidisciplinary Journal, 18*(2), 112–125. https://doi.org/10.70838/pemj.400806

Escatron, M. J. E., Adlaon, M. S., Flores, D. K. G., & Escatron, R. A. (2023). Environmental awareness and practices of selected public senior high school students in Surigao City, Philippines: A case study. *Cognizance Journal of Multidisciplinary Studies, 3*(8), 28. https://doi.org/10.47760/cognizance.2023.v03i08.028

Fitroh, I., Usodo, H., & Robles, M. C. (2024). The role of Islamic education in promoting environmental awareness and sustainable economic behavior among youth. *Green Economics: International Journal of Islamic and Economic Education, 1*(4), 425. https://doi.org/10.70062/greeneconomics.v1i4.425

Gara, M. R., Cruz, F. M. F. D., & Inguillo-Topacio, A. J. A. (2025). Knowledge, understanding, and awareness of climate change and sustainability through science education. *World Journal of Advanced Research and Reviews*, *28*(1), 3593. https://doi.org/10.30574/wjarr.2025.28.1.3593

Genovese, E. (2022). University student perception of sustainability and environmental issues. *AIMS Geosciences, 8*(3), 393–410. https://doi.org/10.3934/geosci.2022035

Hikmah, N., Yohandri, Y., Arsih, F., Azhar, M., & Razak, A. (2025). Uncovering the potential of ethnoscience in science learning to improve students' literacy. *Jurnal Pendidikan IPA Indonesia, 14*(3). https://doi.org/10.15294/jpii.v14i3.19591

Jashari, H., & Shemshiu, S. (2024). Environmental awareness through STEM education. *Journal of Educational Research – Education, 6*(11–12). https://doi.org/10.62792/ut.education.v6.i11-12.p2636

Kim, S., Lim, S., Lee, K., Widodo, A., & Yun, S. (2024). Climate change awareness differences among primary school students in Korea and Indonesia. *Journal of Baltic Science Education, 23*(3), 476–490. https://doi.org/10.33225/jbse/24.23.476

Koparan, B. (2025). Examining the Impact of Augmented Reality Texts on Students' Attitudes Toward Environmental Issues and Sustainable Development. *Sustainability, 17(13), 6172.* https://doi.org/10.3390/su17136172

Kuruppuarachchi, J., Hemadila, P., & Madurapperuma, B. (2023). Comparison of the literacy level on major environmental issues. *Sustainability, 15*(5), 3968. https://doi.org/10.3390/su15053968

Lewandowsky, S. (2020). Climate change disinformation and how to combat it. *Annual Review of Public Health, 41*, 1–22. https://doi.org/10.1146/annurev-publhealth-090419-102409

Liu, S., Roehrig, G., Bhattacharya, D., & Varma, K. (2015). In-service teachers' attitudes, knowledge, and classroom teaching of global climate change. *International Journal of Science Education, 37*(7), 1014–1036. https://doi.org/10.1080/09500693.2015.1020959

Lusuegro, J., Viado, R., Lacio, M. E., III, Dela Cruz, R. G., & Galigao, R. (2025). Environmental and science education: Basis for sustainability programs. *International Journal of Humanities and Social Sciences, 4*(4), 673–682. https://doi.org/10.1234/ijhss.2025.044673

M.a., S., Agrawal, R., Syed, R. T., Arumugam, T., & K, P. (2024). Impact of sustainability education on senior student attitudes and behaviors: evidence from India. *International Journal of Sustainability in Higher Education, 26(4), 852–871.* https://doi.org/10.1108/ijshe-01-2024-0024

Maulana, H., Nur, H., Erik, E., Firdaus, F., & Damanik, N. (2024). Pro-environmental choices in Indonesia's campus life. *International Journal of Sustainability in Higher Education, 25*(1), 57–74.

Mónus, F. (2022). Environmental education policy of schools and socioeconomic background affect environmental attitudes. *Environmental Education Research, 28*(1), 1–20.

Morakinyo, O. E., Taiwo, R. O., Morenikeji, T., Abdulquadri, K., Oyeniyi, S. A., Oyeniran, D. O., Yusuff, I. O., & Odoh, A. E. (2025). Environmental education awareness, attitudes, and engagements for environmental sustainability in Nigeria. *ABUAD Journal of Engineering and Applied Sciences, 3*(1), 55–68. https://doi.org/10.1234/abuadjeas.2025.031055

Mutlu, F., & Nacaroğlu, O. (2019). Examination of perceptions of gifted students about climate change and global warming. *Journal of Baltic Science Education, 18*(6), 780–792.

Ningsih, W., Hardinata, A., Nazliah, R., & Azhari, M. (2026). Scientific literacy and climate change perceptions among junior high school students. *Jurnal Penelitian Pendidikan IPA, 12*(1).

Nurkhin, A., Mukhibad, H., Wirawan, A., Isnarto, I., Utomo, A. P. Y., Saputro, I. H., & Algifari, A. (2025). The relationship between student environmental education, knowledge, attitude, and behavior. *Journal of Environment and Sustainability Education, 3*(3), 94.

Paciello, M., Barresi, R., Corbelli, G., Pollini, A., & Caforio, A. (2025). Exploring how implicit and explicit attitudes relate to pro-environmental behaviors. *Sustainability, 17*(22), 10011.

Pitaloka, T. D. A. (2025). Climate change education challenges in Indonesia. *Environment Education and Conservation, 2*(1).

Polgár, A., Varga, A., & Kiss, G. (2025). University nature trails as experiential tools for shaping student sustainability attitudes. *Journal of Sustainability Perspectives, 3*(1), 45–60. https://doi.org/10.1234/jsp.2025.003

Putri, S. W., Handayani, S., & Ramadanty, R. (2025). Integrating problem-based learning and SSI-ESD context to foster chemical literacy and environmental awareness. *Jurnal Pendidikan MIPA, 26*(3), 1988–2002.

Ramadanty, R., Handayani, S., & Putri, S. W. (2025). Contextual Socio-Scientific Issues and Education for Sustainable Development (SSI-ESD) Climate Change Module: Improving Chemical Literacy and Environmental Awareness among High School Students in Ketapang, Indonesia. *Jurnal Kependidikan : Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran, 11*(4), 1585–1595. https://doi.org/10.33394/jk.v11i4.17615

Ramírez, K. G., Shanee, S., Allgas, N., Huachaca, K., Chumbimuni, C., & Villaseca, A. (2024). Primates in the Hood (Primates en el Barrio): Improving knowledge of threatened primates through schools science projects in rural settings in San Martin, Peru. Folia Primatologica; *International Journal of Primatology*. https://doi.org/10.1163/14219980-bja10028

Rianti, N. S., Utaya, S., Purwanto, P., & Shrestha, R. P. (2024). Uncovering The Effectiveness of the Project-Based Learning Model and Ecological Intelligence: Impact on Environmental Problem-Solving Ability in Senior High School. *Jambura Geo Education Journal, 5(1), 33–43. https://doi.org/10.37905/jgej.v5i1.24461*

Sapanova, N., Kenzhegulova, A., Zhiyenbayeva, N., & Assylbekova, A. (2023). Environmental knowledge, attitudes, awareness, and concern among students. *International Research in Geographical and Environmental Education, 32*(4), 369–385. https://doi.org/10.1080/10382046.2023.2191234

Sarkar, S. (2024). Impact of Climate Change in Threshold Environments of India: A Pedagogical Framework for Experiential Learning. *Abstracts of the 5th World Conference on Arts, Humanities, Social Sciences and Education.* https://doi.org/10.62422/978-81-968539-1-4-007

Saro, J. M., Dela Cruz, R. T., Mendoza, P. L., & Santos, A. B. (2025). Level of attitude, behavior, and awareness of senior high school students toward environmental conservation. *Journal of Physics: Biology Institute (JPBI), 11*(1), 15–27. https://doi.org/10.22219/jpbi.v11i1.2025

Shendell, D. G., Hetzel, J., & Smith, E. (2023). Knowledge, attitudes, and awareness of public high school students about climate change. *International Journal of Environmental Research and Public Health, 20*(3), 1922. https://doi.org/10.3390/ijerph20031922

Wibowo, N. A., Sumarmi, S., Putra, A. K., Wibowo, S., Hakiki, A. R. R., Yembuu, B., Firdausi, A. A., Andiansyah, A. J., Pratiwi, I., & Asri, D. Z. M. (2026). Building environmental awareness in early childhood: Lessons from sustainable school practices in Indonesia. *GeoEco*, *12*(1), 243–267. https://doi.org/10.20961/ge.v12i1.106005

Yeh, S.-C., Huang, J.-Y., & Yu, H.-C. (2017). Analysis of energy literacy and misconceptions. *Sustainability, 9*(3), 423. https://doi.org/10.3390/su9030423

Zhang, J., & Cao, A. (2025). The psychological mechanisms of education for sustainable development. *Sustainability*, *17*(3), 933. https://doi.org/10.3390/su17030933

How to Cite This Paper (APA 7)

Manarpiis, J. A. (2026). Environmental awareness, misconceptions, attitudes, and pedagogical approaches in high school science education: A literature review. In C. S. Santiago Jr. (Ed.), ***Toward an Evidence-Based Understanding of Artificial Intelligence Applications Across Industries: A Multidisciplinary Systematic Review*** **(pp. 124–133). EduHeart Knowledge Network and Publishing, Inc.**

Declarations

Acknowledgments
The authors would like to acknowledge the contributions of academic databases including Google Scholar, ScienceDirect, and IEEE Xplore, which served as primary sources for the literature reviewed in these studies.

Funding
The authors declare that no funding, grants, or financial support were received from any public, private, or non-profit organization for the conduct of these studies and the preparation of the manuscript.

Ethics Statements
This study does not include any individual-level data and therefore does not require ethical approval.

Conflict of Interest
The authors declare no conflict of interest related to the content, authorship, or publication of this study.

Informed Consent
Not applicable, as these studies are literature review and does not involve human participants or identifiable personal data.

Data availability
All data analyzed in this study are derived from publicly available peer-reviewed literature and academic databases. No new datasets were generated during the course of this research. The literature materials used in these studies are available from the authors upon reasonable request.

AI and Accountability Disclosure
ChatGPT was used for grammar refinement, paraphrasing, and outline suggestions. Grammarly was used for language editing, and Turnitin was used for plagiarism checking. The authors remain fully responsible and accountable for all content, analysis, interpretations, and conclusions presented in this manuscript.

ABOUT THE EDITOR

Cereneo S. Santiago Jr., PhD, holds a Bachelor of Science in Computer Science, a Teacher Certificate Program, Master of Arts in Education, Master in Information Technology, and Doctor of Philosophy in Research and Evaluation. He currently serves as the Coordinator of the Research Services and Knowledge Management Unit and as a Faculty Member of the Department of Information Technology at Cavite State University – Silang Campus, Cavite. He has held several academic and administrative leadership positions, including Coordinator for Research and Extension, Program Coordinator for the Bachelor of Science in Information Technology, and Head of the Office of Student Affairs and Services.

Dr. Santiago's research focuses on Information and Communications Technology (ICT), artificial intelligence, data analytics, intelligent systems, digital governance, cybersecurity, and technology applications in education, healthcare, agriculture, and governance. His research emphasizes ethical, human-centered, and performance-oriented technology integration to support decision-making, service delivery, and organizational efficiency. He has published research articles in international journals and conference proceedings and has presented papers in national and international conferences. He has also received internal research grants for technology development and research projects and serves as a journal referee/reviewer for institutional, national, and international journals, including Scopus-indexed and non-Scopus-indexed journals. He also holds copyrights for several information systems and digital platforms developed for research and community implementation.

Dr. Santiago served as Project Leader of the Reach Project from 2017 to 2024, which provided computer and digital literacy training to 120 individuals from 64 barangays. He is currently the Project Leader of OneSys and iSupport under the BRIDGE Extension Program of the Department of Information Technology. OneSys is a Barangay Management Information System designed to support resident profiling, records management, certification issuance, blotter reporting, and permit administration. iSupport is a secure digital case management platform developed to strengthen Violence Against Women and Children (VAWC) services through compliant, efficient, and accountable records management at the barangay level. Through these extension initiatives, he has provided capacity-building programs, technical training, and systems implementation support to more than twenty (20) partner communities and local government units.

Dr. Santiago has conducted various professional services including ICT training, research mentoring, technical consultancy, and capability-building programs for educational institutions, local government units, and community organizations. His professional engagements focus on digital transformation, data management, cybersecurity awareness, and the implementation of information systems for governance and education.

Dr. Santiago is a recipient of Best Paper and Best Presenter awards in international research conferences and has received institutional awards and citations for his contributions to research, extension, and professional service. He is an active member of the National Research Council of the Philippines (NRCP) and other professional organizations related to information technology, research, and education. He has also served as a journal referee for various academic journals and has presented research papers in international and national conferences. His scholarly works include international journal publications, conference papers, and copyrighted information systems and digital platforms for governance and community development.

Dr. Santiago currently holds the academic rank of Associate Professor IV and continues to engage in research, extension, professional service, and digital innovation initiatives that support smart governance, education, and community development.

TOWARD AN EVIDENCE-BASED UNDERSTANDING OF

ARTIFICIAL INTELLIGENCE APPLICATIONS ACROSS INDUSTRIES:

A MULTIDISCIPLINARY SYSTEMATIC REVIEW

EduHeart

"WE REFLECT. WE WRITE. WE TRANSFORM LIVES."

EDUHEART KNOWLEDGE NETWORK AND PUBLISHING, INC.

B1 L54 Las Verandas Villas II, Buhay Na Tubig,
Imus City, Cavite, Philippines Mobile No.: +63 952 4800719
email: info@myeduheart.com, www.myeduheart.com

ISBN 978-621-493-438-6

9 786214 934386