Journal of
Digital Science

CONTENTS

Generative AI: Concepts, Challenges, and Research Opportunities

Anh Tran[0009-0005-9471-8375], Ojo Folake, Karthik Srinivasan[0000-0002-1608-6190]

University of Kansas School of Business, Lawrence, Kansas

Abstract This paper examines the landscape of generative artificial intelligence (GenAI), a rapidly evolving segment of technology characterized by its capacity to produce various forms of original content, such as text, images, and music, through deep neural network models. The rising popularity of tools like ChatGPT, which reached over 100 million users within months of its release in November 2022, signifies a paradigm shift in enterprise AI adoption, provoking organizations to realign their strategic objectives with digital transformation. While the immense potential of GenAI promises enhancements in productivity and personalized services across industries, it simultaneously presents critical challenges, including ethical concerns surrounding bias, accountability, and human-centric decision-making. This study highlights the need for effective testing frameworks that ensure GenAI systems enhance, rather than detract from, human cognition. This paper contributes to the discourse on GenAI by outlining its foundational mechanisms, implications for various sectors, and the interplay of opportunities and risks posed by its widespread implementation.

Keywords: Generative AI (GenAI), Foundation Models, Large Language Models (LLMs), Generative Pre-trained Transformers (GPT), Ethics in AI, Explainability and Interpretability

Introduction

Generative artificial intelligence, often referred to as GenAI or Generative AI, is the branch of artificial intelligence (AI) that focuses on generating original content such as texts and visual compositions like videos or music using advanced machine learning models (Kumar et al., 2025). GenAI achieves its advances in interpretability and predictability through learning from extensive datasets. This learning process involves a system that connects the underlying structure and discerns patterns from the dataset to produce creative and coherent outputs that reflect or expand upon the input data. While traditional AI is designed for repetitive tasks and classification, GenAI uses techniques like neural networks and transformer-based architecture to generate original and quality-relevant responses (Vaswani et al., 2017). These architectures enable GenAI to process vast amounts of information in parallel, capturing long-range dependencies in language and thereby enhancing its ability to understand context and nuance.

This article explores the foundation of GenAI and its structural integration within information systems, focusing on how GenAI generates optimal outputs based on human-written prompts (Staff, 2024). We delve into the key mechanisms behind GenAI's ability to produce diverse and contextually relevant content across various

domains. Additionally, we address some of the critical challenges associated with GenAI, such as bias, interpretability, and ethical concerns, highlighting the importance of frameworks like GenAIX to ensure the transparency and ethical use of AI-driven systems. These considerations are particularly relevant for researchers and practitioners within the information systems discipline, as they influence both the design and implementation of AI systems in real-world contexts. Our aim in this paper is to provide a comprehensive understanding of GenAI's foundations, challenges, and its business applications. This knowledge is essential for leveraging GenAI's full potential while ensuring its responsible and ethical adoption in the evolving landscape of information systems. By examining how GenAI can be thoughtfully integrated into business strategies, this article contributes to the ongoing dialogue on its transformative potential in business and technology.

Artificial Intelligence (AI) began to receive mainstream attention during 2022 when simple-to-operate applications emerged that could produce images that had highly convincing levels of 'humanness'. These AI systems demonstrated the ability to generate images, videos, and even text that appeared indistinguishable from human-produced content, sparking widespread public interest in their potential applications. Although ChatGPT is not the first text-generating AI, it represented a significant breakthrough in generative AI (Gen-AI) due to its user-friendly chat-like interface (Bender, 2024). The versatility of GenAI extends across various industries. In healthcare, for instance, GenAI aids in analysing medical images, facilitating early disease detection, and personalized treatment planning. In the financial sector, it enhances fraud detection by identifying unusual patterns in transaction data. The entertainment industry leverages GenAI for content creation, such as generating scripts or composing music, showcasing its creative potential. In the realm of education, GenAI is transforming learning experiences by providing personalized tutoring and generating educational content tailored to individual student needs. This adaptability facilitates more effective learning outcomes and supports educators in curriculum development. Additionally, GenAI's ability to generate realistic simulations and scenarios serves as a valuable tool in training and skill development across various fields. Despite its impressive capabilities, GenAI also presents challenges, particularly concerning ethical considerations. The potential for generating misleading or biased content necessitates the implementation of robust guidelines and oversight. Ensuring transparency in how these models operate and make decisions is crucial to maintain trust and accountability in their applications. As we look toward the future of GenAI, it is important to remember that its rapid advancements today are built upon decades of foundational research. The journey of Generative AI has not been instantaneous; rather, it has been shaped by years of experimentation and breakthroughs that have gradually brought us to our current state. Understanding the historical context of GenAI's development helps us appreciate how far the field has come and the challenges that have been overcome to reach this point. To trace the roots of GenAI, we must look back to the early days of artificial intelligence research, which laid the groundwork for the sophisticated models being developed today.

Although GenAI systems may seem recently developed and widely applied, it has their roots in early AI research, with the foundational artificial intelligence concept researched by Alan Turing in the 1950s. Turing's pioneering work on machine intelligence, notably through the concept of the 'Turing Test,' laid the groundwork for future exploration in machine learning and AI systems. His ideas about the ability of

machines to mimic human cognition and behaviour remain a cornerstone in the development of intelligent systems. The vision that Turing set out, though conceptual at the time, inspired many subsequent innovations in AI, including the modern field of generative AI. Early manifestations of generative AI include the development of rule-based chatbots, such as the ELIZA chatbot in 1961, which marked a milestone in enabling language models to learn predefined patterns from training data (Lawlow & Chang, 2024). While ELIZA was quite simple by today's standards, it was groundbreaking at the time as it demonstrated how computers could simulate human-like conversation within specific contexts. Over time, AI systems evolved to incorporate more sophisticated algorithms, and by the 1980s and 1990s, the rise of neural networks and early forms of machine learning paved the way for more complex generative models. Over the past decade, a significant rise in neural networks and deep learning techniques has laid the foundation for modern AI systems to learn complex patterns from data. These networks, inspired by the human brain, allowed machines to perform more advanced tasks, including recognition, classification, and prediction. The introduction of deep learning, in particular, which involves training multiple layers of neural networks, has enhanced the capability of AI systems to process vast amounts of data and extract meaningful insights, further fuelling the development of GenAI.

The journey toward developing GenAI began in 2014 when Ian Goodfellow unveiled Generative Adversarial Networks (GANs). This innovation marked a significant advancement, enabling AI to create content using unsupervised machine learning techniques (Lawlow & Chang, 2024). GANs revolutionized generative models by using a two-part system generators and discriminators, to produce realistic images, text, and other content. This innovation opened up new possibilities for AI to not just analyse data, but to create new and original content that mimics real-world attributes. GANs are still widely used today in fields like art, design, and entertainment for creating everything from photorealistic images to music and video. Another type of model that played a significant role in GenAI development is the Transformer architecture model, presented in the 2017 paper "Attention is All You Need," which has become the foundation for modern GenAI (Vaswani et al., 2017). This architecture enables models to process entire sequences of data simultaneously, rather than sequentially, allowing for more efficient training and the handling of long-range dependencies in data. The self-attention mechanism within transformers allows models to weigh the importance of different words in a sentence, enhancing the generation of coherent and contextually relevant content. The Transformer architecture led to the creation of large language models such as BERT and GPT.

In parallel with our exploration of GenAI's role in business, it's also important to consider the impact of GenAI on educational fields, particularly in management and English education. For instance, the adoption of GenAI in management curricula presents a unique opportunity to enhance student learning while mitigating risks, such as overreliance. This approach, as discussed by (Hyde et al., 2024) suggests that rather than resisting GenAI, educators should embrace its potential to create engaging, experiential learning opportunities aligned with andragogical principles. This can help foster autonomy, competence, and real-world application among students, enriching their learning experience and improving academic outcomes.

Similarly, in English education, the integration of GenAI, such as ChatGPT, presents both transformative potential and challenges. The essay by Bender (2024)

emphasizes the importance of understanding GenAI as an essential component of digital literacy. Rather than focusing solely on writing skills, teachers can leverage GenAI to enrich lessons in reading, viewing, and interpretation. While ethical concerns, such as plagiarism and equity, remain critical, the paper argues for the productive use of GenAI to augment traditional learning methods, offering students an opportunity to develop critical thinking skills in the digital age.

The long-term societal implications of widespread GenAI adoption are profound. While GenAI has the potential to democratize access to knowledge and creativity, it also raises questions about its impact on employment, income inequality, and the redistribution of labour. As automation continues to transform various industries, there may be significant shifts in the labour market, requiring new forms of education and training to prepare individuals for emerging job roles. Furthermore, ethical governance frameworks must be developed to address the unintended consequences of GenAI, ensuring that its benefits are equitably distributed while minimizing harm to vulnerable groups. As the integration of GenAI continues to unfold, the ongoing conversation about its impact will evolve. Researchers, educators, and industry leaders must collaborate to navigate these challenges, ensuring that the use of GenAI promotes positive outcomes for individuals, organizations, and society at large. This dynamic landscape of technological evolution offers exciting opportunities for innovation, but it also necessitates a thoughtful approach to ensure that GenAI contributes to a future that is both equitable and sustainable.

2. Methodology

GenAI evolution

At the heart of artificial intelligence lies a simple but powerful goal: helping machines understand and respond to human input meaningfully. Whether it is answering a question, writing a story, or assisting with tasks, AI must first understand the intent and meaning behind what we say. This requires the bridge between human communication and machines, which gave rise to Natural Language Processing (NLP). Natural Language Processing is one of the foundations of modern GenAI that focuses on allowing machines to understand, generate, and interact using human language. The introduction of NLP allowed the development of LLMs, which are trained on a massive amount of text data to better understand and produce language (Sandhu, 2024). Two key innovations, word embedding and attention mechanisms, have been critical in teaching AI to process language more like humans do.

Word embedding is the foundational block of LLMs and GenAI, which are dense vector representations of words designed to capture their semantic meanings and contextual relationships (Mikolov et al., 2013). When a user inputs a prompt into the system, the models will break the raw text into smaller units called tokens. Each token is then transformed into a dimensional numerical vector, and the vectors exist to embed meaning. Similar words like "doctor" and "nurse" are closer in the vector space. This helps models understand not just individual words, but the deeper context behind them.

Building on this, attention mechanisms revolutionized how models handle language in LLMs. Instead of processing words one at a time, attention allows the model to weigh the importance of all words in a sentence when interpreting each one

(Vaswani et al., 2017). Consider this sentence: "The cat sat on the floor, and it purred." The attention mechanism helps the model interpret "it", which refers to "cat", by assigning a higher weight to "cat" when processing it. This mechanism allows models to overcome the limitations of processing sequences step-by-step, as it can directly access any part of the input and weigh its significance. This ability to focus on relevant information dramatically improved how machines understand complex sentences.

Transformer architectures are a type of AI model designed to predict and generate human language text with significant accuracy by analysing large amounts of text data combined with word embeddings and attention (Vaswani et al., 2017). This technology powers tools like ChatGPT and Gemini, which respond to user prompts in a conversational way by learning from vast sources of information. Transformers often use advanced techniques like supervised and reinforcement learning to improve their responses over time (Iyengar et al., 2023).

Figure 1 illustrates the model structure of an LLM. The foundation models we know today, such as GPT, Gemini, and others, grew from transformer architecture (Toner, 2023). A foundation model is an AI system with general abilities that can be adapted to do specific tasks. Unlike traditional AI models designed for one specific purpose, foundation models serve as a base or "foundation" that can be fine-tuned for specialized applications. Early on, AI research focused on tasks like image classification and reinforcement learning, but in the late 2010s, researchers shifted towards language and content generation. GPT models played a key role in this shift, starting with GPT-2 (Toner, 2023). These models were trained on vast datasets, using increased computational power and parameters, allowing them to understand and create human-like text.

Today, these advancements have created a wave of innovation across the AI landscape. Google DeepMind's Gemini (Shahine, 2024), OpenAI's ChatGPT (Siles, 2024), DeepSeek's reasoning-focused models (DeepSeek-AI, 2025), Anthropic's Claude family (Harper, 2025), Meta's LLAMA models (Touvron et al., 2023), Microsoft's Phi series (Abdin et al., 2024), and xAI's Grok (xAI, 2025) are implementations that use advanced AI models that are more sophisticated, efficient, and socially aware for users.

Google GenAI represents a rapidly evolving and highly competitive research area that drives significant technological advancements, particularly among large technology companies. Gemini, developed by Google DeepMind, is a series of multimodal LLMs that integrate and process multiple types of data, including texts, audio, images, and more (Shahine, 2024). The latest Gemini 2.5 Pro, debuting in March 2025, further refines multimodal processing and overall performance with an expanded context window and adaptive learning features, outperforming its predecessors and proving to be a top competitor in this landscape (Young, 2025).
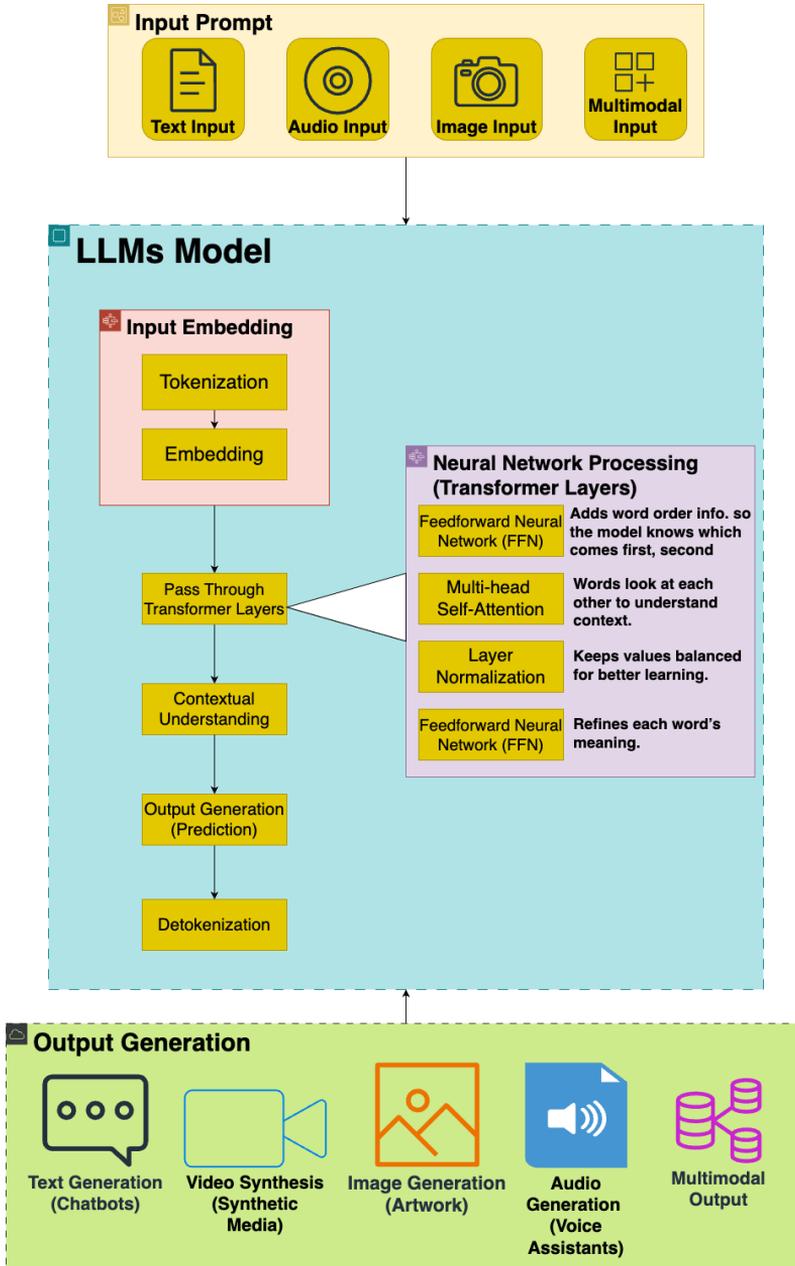
Fig. 1. Basic illustration of an LLM structure

ChatGPT, developed by OpenAI, is a widely recognized and conversational AI system powered by the GPT series of LLMs (most recently GPT 3.5 and GPT 4.0). GPT-3, the famous game changer in 2020, marked an advancement of LLMs with 175 billion parameters and became the backbone of many AI applications (Siles, 2024). Using transformer architecture at its core, GPT models' self-attention mechanism allows it to understand input words and generate accurate human-like responses that excel in various tasks (Bowman, 2023).

DeepSeek is a company based in China, and their DeepSeek AI is an advanced open-source LLM designed to excel in reasoning and problem-solving tasks (Szlezak, 2025). Building on its predecessor model, DeepSeek-R1, DeepSeek AI exhibits chain-of-thought capabilities that enable it to solve complex challenges across mathematics, coding, and logical reasoning (DeepSeek-AI, 2025). Similar to how GPT models leverage transformer architectures, DeepSeek AI uses its unique RL-driven training process in its LLMs (including DeepSeek Coder, DeepSeek-LLM, DeepSeek-V2, and DeepSeek-R1) (DeepSeek-AI, 2025).

Claude, developed by Anthropic, is a family of LLMs that focuses on flexibility and human-like precision across a wide range of tasks. Introduced in early 2023 and refined through successors such as Claude 2, Claude 2.1, Claude 3, and Claude 3.5, it has continually enhanced its reasoning abilities and contextual understanding (Harper, 2025). Anthropic's Claude AI has been examined using circuit tracing techniques that reveal its internal computational pathways (Ameisen et al., 2025). This circuit tracking uncovers the underlying mechanism of LLMs and provides interpretability of its "thought process" and explainability (Ameisen et al., 2025). With these capabilities, Claude AI positions itself as a key competitor in the evolving landscape of LLMs.

LLAMA (Large Language Model Meta AI), created by Meta AI, is a collection of open and efficient foundation language models ranging from 7 billion to 65 billion parameters (Touvron et al., 2023). Notably, LLaMA-13B outperforms GPT-3 (which has 175 billion parameters) on a variety of benchmarks, while LLaMA-65B competes with state-of-the-art models like Chinchilla-70B and PaLM-540 B (Touvron et al., 2023). Llama 4 uses a mixture-of-experts (MoE) architecture that activates only a fraction of its parameters per token, reduces computational overhead, and significantly speeds up inference while maintaining high-quality outputs (Horsey, 2025).

Microsoft's Phi series of small language models (including Phi-1, Phi-1.5, Phi-2, Phi-3, and Phi-4) demonstrates that high performance in reasoning, mathematics, and code tasks can be achieved without the massive model sizes traditionally expected (Abdin et al., 2024). In fact, the Phi3mini model is optimized to run locally on devices such as smartphones, even allowing for quantization to 4 bits and real-time inference directly on a modern phone (Abdin et al., 2024). Phi-4 specializes in complex reasoning challenges, particularly in mathematics, often outperforming much larger models on specialized tasks (Abdin et al., 2024).

Grok, developed by xAI under Elon Musk's leadership, is an AI that delivers real-time, context-aware responses by directly updating live data from X (formerly Twitter) (Mahale, 2025). Grok AI's design also focuses on flexibility and efficiency. In parallel with the full-featured Grok 3 Beta, the Grok 3 mini (Think) variant is optimized for cost-efficient performance on STEM tasks without compromising quality (xAI, 2025). These capabilities empower Grok AI to deliver context-rich interactions while continuously developing through real-world feedback.

Table 1. List of commonly known LLMs in 2025

Model Name	Developer	Architecture	Key Features/Strengths	Training Data/ Process (Briefly)	Market Differentiation
Gemini 2.5 Pro	Google DeepMind	Multimodal LLM	Handles text, images, audio, code; large context window; adaptive learning	Trained on diverse multimodal datasets with advanced context handling	Flagship model competing at top-tier levels with GPT-4.5 and Claude 3
Gemini Flash	Google DeepMind	Lightweight Multimodal LLM	Faster, cheaper inference; optimized for latency	Trained with efficiency-focused objectives for deployment	Performance-focused lightweight companion to Gemini Pro
ChatGPT (GPT-4.5)	OpenAI	Transformer	Enhanced reasoning, expanded memory, multimodal input	Built on GPT-4, fine-tuned with human feedback and tool use	Enterprise-grade model for coding, reasoning, and pro user tasks
GPT-3.5 / GPT-3.0 mini	OpenAI	Transformer	Conversational, fast, less resource-intensive	Trained on broad internet-scale data, with earlier transformer tech	Ideal for general chat, educational tools, and cost-effective apps
Claude 3 Opus	Anthropic	Transformer + interpretability tools	Top-tier reasoning, multilingual, complex task handling	Uses circuit-tracing for interpretability; refined human feedback loop	Known for transparency and safe AI alignment
Claude 3 Haiku	Anthropic	Lightweight Transformer	Smaller, faster Claude; optimized for speed and cost	Distilled and fine-tuned for efficiency	Real-time Claude assistant for mobile/embedded use cases
DeepSeek AI (V2, R1)	DeepSeek	Transformer + Reinforcement Learning	Chain-of-thought reasoning, excels in STEM	RL-driven training with structured task focus	Strong open-source option with emphasis on math and logic
DeepSeek Mini	DeepSeek	Lightweight Transformer	Compact model optimized for coding and reasoning	Trained on focused technical datasets; open-source	Reasoning ability in small-size form, great for local or offline tasks
LLaMA 4	Meta AI	Transformer (Mixture-of-Experts)	Scalable, efficient; activates only part of model per query	MoE architecture reduces cost while maintaining quality	Competitive open model rivalling GPT-4 with smaller active footprint

Model Name	Developer	Architecture	Key Features/Strengths	Training Data/ Process (Briefly)	Market Differentiation
LLaMA 3 8B/13B	Meta AI	Transformer	Lightweight models with strong performance	Trained on diverse multilingual data; highly efficient	Compact alternatives with strong performance for their size
Phi-4	Microsoft	Small Transformer LLM	High accuracy in math and code, even at small scale	Trained on synthetic high-quality data; reasoning-focused	Outperforms larger models on logic/math; designed for local inference
Phi-3 Mini	Microsoft	Quantized Transformer	4-bit model for mobile, offline, and low-power devices	Optimized for phones/tablets, quantized for edge	Real-time local inference on mobile—no cloud needed
Grok 3 (Beta)	xAI (Elon Musk)	Transformer + Live Data Layer	Live updates via X, social context integration	Continuously updated with live social data from X.com	Real-time, context-rich AI with unique access to X ecosystem
Grok 3 Mini (Think)	xAI	Lightweight Transformer	Optimized for cost-performance on technical tasks	Trained for STEM performance, updated efficiently	A lean variant designed for targeted STEM-heavy workloads

GenAI has undergone a long evolutionary journey (Dwivedi et al., 2023). Now, we have achieved significant progress in AI with more and more models created like GPT-4, Claude 3.0, or Gemini 2.5 that are widely used in daily tasks and projects. These newest models represent AI technologies that are widespread and carry the potential to be even more advanced in the future (Dwivedi et al., 2023). First, the use of larger-scale training data has significantly improved contextual understanding. Models now leverage vast and diverse datasets, often spanning multiple languages and modalities, to generate more nuanced and accurate outputs (Prasad & Choudhary, 2021). Second, increased parameter efficiency has allowed models to achieve higher performance with reduced computational demand. Techniques like mixture-of-experts (MoE) architectures and sparse training allow AI systems to recognize complex relationships without proportional increases in resource demands (Gan et al., 2025). Third, optimized transformer architectures, such as those used in models like GPT-4 and Gemini, have enhanced inference speeds while maintaining output quality. Innovations like flash attention, quantization, and distillation enable these models to run efficiently even on local devices, broadening their applicability (Dwivedi et al., 2023). Finally, the SOTA prioritizes privacy, safety, and interpretability. Techniques like federated learning, differential privacy, and explainable AI (XAI) frameworks address ethical concerns while bringing user trust (Lopez-Ramos et al., 2024).

3. Discussion

Research opportunities

ChatGPT operates on a business model where a highly valuable, high-capacity tool is provided for free, and it has been a common pattern embraced by disruption technologies in established markets. It has led to the rapid spread of generative AI, paving the way for a future with numerous new products and services that incorporate AI characteristics. ChatGPT emerged as the fastest-growing software in history by January 2023, gaining more than 100 million users within just a few months. This rapid expansion has prompted the majority of firms to develop their own AI strategies, with expectations that AI will bring significant changes across various sectors (Hyde et al., 2024). This shift also opens rich research opportunities related to organizational readiness, strategic alignment, and digital maturity. For example, how do firms in different industries adjust their infrastructure and culture to effectively adopt GenAI? What organizational factors foster successful GenAI transformation, and what role does leadership play in shaping AI-driven innovation?

Researchers can investigate how early AI adopters differ from laggards in terms of competitive advantage and whether these disparities persist over time. Furthermore, there is growing interest in exploring how GenAI can affect internal power dynamics within firms, for instance, by reducing middle management roles through automation, or by changing the nature of knowledge work. Research could also focus on how GenAI redefines the boundaries between firms and customers, especially as AI becomes more embedded in customer service, product development, and marketing personalization. On a macro level, scholars may examine the broader economic implications of widespread GenAI use: its influence on productivity growth, labour market polarization, and the emergence of new business models. In particular, the gig economy and remote work ecosystems are expected to be profoundly impacted by intelligent task automation. Interdisciplinary research combining business, computer science, ethics, and policy can help develop responsible frameworks for AI governance, focusing on fairness, explainability, and accountability in algorithmic decision-making. As organizations respond to the sweeping changes across industries and society, many are turning to AI-driven digital transformation as a business strategy to utilize artificial intelligence and improve the way they operate, make decisions, and engage with customers. By integrating AI into their daily operations and systems, organizations can enhance human capabilities, automate routine processes, and extract insightful information from data to facilitate better-informed decision-making. It renders enterprises more efficient, their resources better utilized, and their services more personalized for consumers. The general sentiment is that the adoption of AI will lead to higher productivity, nimbleness, and innovativeness in organizations (Aldoseri et al., 2024).

While the widespread integration of GenAI tools demonstrates their immense potential to transform industries and improve organizational outcomes, it also raises a critical concern: how well do we understand these systems? As businesses increasingly rely on AI-driven decision-making, the need for interpretability and explainability becomes more urgent. It is not enough for AI to deliver results; stakeholders must also be able to understand *why* and *how* those results were produced. This is particularly important in sensitive areas like education, healthcare,

and finance, where trust, fairness, and accountability are essential. It is commonly perceived that the underlying mechanisms for decision-making of GenAI are not well-known (Mah, 2024). Although GenAI models, such as ChatGPT, can respond in ways that invite users, like students, to interact more intensely with tasks like reading, writing, and critical thinking, the issue of how such systems generate their responses remains to be addressed. This black-box nature of GenAI systems poses challenges for transparency and accountability. When users are unable to trace the logic behind a response, especially in high-stakes domains like healthcare, education, or finance, it becomes difficult to trust or validate AI outputs. Interpretability and explainability are terms that describe being able to comprehend the rationale for AI-created responses. So far, little is understood about how GenAI could help users, such as students, to better comprehend complex texts or media in clear explanations. Without a clear comprehension of how such models function, their results can at times be difficult to fully rely on or learn from. Moreover, as GenAI becomes integrated into learning management systems and adaptive learning platforms, ensuring that these systems are interpretable is crucial for fostering trust among both students and instructors. Explainability could serve not just as a safeguard but also as a learning scaffold, enhancing metacognitive skills by allowing users to reflect on how they arrived at certain conclusions with AI support (Schneider, 2024). Improved explainability of GenAI would enhance its role in education and beyond by making its process clearer and more transparent. This would allow individuals not only to obtain answers but also to understand the reasons behind the answers they are given. (Bender, 2024) Ultimately, increased focus on explainability supports ethical AI use, encourages more meaningful human-AI collaboration, and lays the foundation for GenAI systems that can be audited, improved, and tailored to diverse user needs.

Similarly, concerns about AI-driven decision-making being unfair or biased within organizational contexts are emerging. AI systems are largely still not designed to provide fair treatment in the face of inherent biases within their training data or design. Scholars and practitioners are being encouraged to develop more open and fair AI methods to combat this. Without a special focus being given to issues of ethics, privacy, and fairness, GenAI may exacerbate existing inequalities and erode trust in AI technologies (Nadeem et al., 2022).

Evaluating Generative AI (GenAI) involves more than a mere quantification of its ability to produce correct or relevant answers; it also requires an understanding of how AI systems interpret meaning and interpretation. Since GenAI does not entail genuine human-like comprehension, answers are based solely on patterns in the data and not on actual comprehension. This inevitably creates serious questions about how much we can rely on content produced by AI, especially in situations where critical thinking or nuanced interpretation is required. Effective GenAI testing would consider not only technical efficacy but also whether or not such systems are in line with human cognition and whether they support consequential, responsible use. As AI gets better, it's essential to test both what AI can and cannot do so that it supports, rather than undermines, human decision-making and communication (Nacheva & Jansone, 2023). Table 2 summarizes the research opportunities for business researchers in the context of GenAI.

Table 2. GenAI research opportunities for business researchers

Research Area	Description
Integration in Information Systems	Explore optimal methods for integrating generative AI into existing information systems, focusing on effectiveness, efficiency, and user experience.
Contextual Relevance and Content Diversity	Research mechanisms that enhance the ability of generative AI to produce diverse and contextually relevant outputs in various domains.
Interpretability and Explainability	Investigate frameworks that enhance the transparency of generative AI models, allowing users to understand how decisions are made.
Bias Mitigation	Develop strategies to identify and reduce biases in generative AI outputs, ensuring equitable representation across diverse demographics.
Ethical Guidelines	Formulate comprehensive ethical guidelines for the responsible use of generative AI, addressing concerns related to misinformation and content manipulation.
Short-term User Impact and Longitudinal Effects	Examine the short-term impacts and long-term implications of generative AI on work, learning, and creative processes across different fields.
Historical Context Analysis	Conduct research that traces the evolution of generative AI, identifying key milestones and how past developments influence current technologies.
User-Centric Design	Investigate user interface designs that maximize the accessibility and usability of generative AI tools, particularly for non-experts.

4. Conclusion

In conclusion, this paper has provided a comprehensive overview of the rapidly evolving field of generative artificial intelligence (GenAI) and its implications. It delves into the foundational mechanisms of GenAI, highlighting how advanced machine learning techniques, such as neural networks and transformer-based architecture, enable these systems to generate original and contextually relevant content. The ability to process vast datasets in parallel and capture long-range dependencies in language has been a crucial driver of GenAI's advancements. The paper then examines the widespread adoption of GenAI tools, exemplified by the rise of models like ChatGPT, which has prompted organizations to realign their strategic objectives with digital transformation. This shift underscores the immense potential of GenAI to enhance productivity, personalization, and innovation across various industries. Our paper also addresses the critical challenges posed by GenAI, including concerns surrounding bias, fairness, accountability, and the need for rigorous testing frameworks to ensure alignment with human cognition and decision-making.

Our study contributes to the ongoing discourse on GenAI by providing a balanced perspective, highlighting both the transformative opportunities and the essential need to address the ethical, interpretability, and human-centric design considerations associated with this rapidly evolving technology. The research

opportunities outlined in this work serve as a roadmap for future investigations, aiming to maximize the benefits of GenAI while addressing the complex challenges that arise from its widespread implementation.

REFERENCES

1.   Abdin, M., Aneja, J., Awadalla, H., Awadallah, A., Awan, A. A., Bach, N., Bahree, A., Bakhtiari, A., Bao, J., Behl, H., Benhaim, A., Bilenko, M., Bjorck, J., Bubeck, S., Cai, M., Cai, Q., Chaudhary, V., Chen, D., Chen, D., … Zhou, X. (2024). *Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone*. https://arxiv.org/pdf/2404.14219

2.   Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2024). Methodological Approach to Assessing the Current State of Organizations for AI-Based Digital Transformation. *Applied System Innovation*, *7*(1). https://doi.org/10.3390/asi7010014

3.   Ameisen, E., Lindsey, J., & Pearce, A. (2025, March 27). *Circuit Tracing: Revealing Computational Graphs in Language Models*. https://transformer-circuits.pub/2025/attribution-graphs/methods.html

4.   Bender, S. M. (2024). Awareness of Artificial Intelligence as an Essential Digital Literacy: ChatGPT and Gen-AI in the Classroom. *Changing English: Studies in Culture and Education*, *31*(2). https://doi.org/10.1080/1358684X.2024.2309995

5.   Bowman, S. R. (2023). Eight Things to Know about Large Language Models. *Arxiv*. http://arxiv.org/abs/2304.00612

6.   DeepSeek-AI. (2025). *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning MATH-500 MMLU SWE-bench Verified DeepSeek-R1 OpenAI-o1-1217 DeepSeek-R1-32B OpenAI-o1-mini DeepSeek-V3*.

7.   Dwivedi, Y. K., Sharma, A., Rana, N. P., Giannakis, M., Goel, P., & Dutot, V. (2023). Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, *192*, 122579. https://doi.org/10.1016/J.TECHFORE.2023.122579

8.   Harper, G. (2025, March 14). *Claude AI: Everything You Should Know - Topdevelopers.co*. https://www.topdevelopers.co/blog/claude-ai/

9.   Horsey, J. (2025, April 7). *Llama 4 AI Model: 10 Million Token Context Tested - Geeky Gadgets*. https://www.geeky-gadgets.com/llama-4-ai-model-long-context-window/

10.   Hyde, S. J., Busby, A., & Bonner, R. L. (2024). Tools or Fools: Are We Educating Managers or Creating Tool-Dependent Robots? *Journal of Management Education*, *48*(4). https://doi.org/10.1177/10525629241230357

11.   Iyengar, K. P., Yousef, M. M. A., Nune, A., Sharma, G. K., & Botchu, R. (2023). Perception of Chat Generative Pre-trained Transformer (Chat-GPT) AI tool amongst MSK clinicians. *Journal of Clinical Orthopaedics and Trauma*, *44*. https://doi.org/10.1016/j.jcot.2023.102253

12.   Kumar, A., Shankar, A., Hollebeek, L. D., Behl, A., & Lim, W. M. (2025). Generative artificial intelligence (GenAI) revolution: A deep dive into GenAI adoption. *Journal of Business Research*, *189*, 115160. https://doi.org/10.1016/J.JBUSRES.2024.115160

13.   Lawlow, P., & Chang, J. (2024, February 12). *The rise of generative AI: A timeline of breakthrough innovations | Qualcomm*. https://www.qualcomm.com/news/onq/2024/02/the-rise-of-generative-ai-timeline-of-breakthrough-innovations

14.   Lopez-Ramos, L. M., Leiser, F., Rastogi, A., Hicks, S., Strümke, I., Madai, V. I., Budig, T., Sunyaev, A., & Hilbert, A. (2024). *Interplay between Federated Learning and Explainable Artificial Intelligence: a Scoping Review*. https://arxiv.org/pdf/2411.05874v1

15.   Mah, P. (2024). *Nobody Knows Why Large Language Models Can Do the Things They Do | CDOTrends*. https://www.cdotrends.com/story/3856/nobody-knows-why-large-language-models-can-do-things-they-do?refresh=auto

16.   Mahale, N. (2025, February 21). *What is Grok 3? A Detailed Guide to the AI Model [+Examples]*. https://writesonic.com/blog/what-is-grok-3

17.   Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. http://arxiv.org/abs/1301.3781)

18.   Nacheva, R., & Jansone, A. (2023). Heuristic Evaluation of AI-Powered Web Accessibility Assistants. *Baltic Journal of Modern Computing*, *11*(4). https://doi.org/10.22364/bjmc.2023.11.4.02

19.   Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Gender bias in AI-based decision-making systems: a systematic literature review. *Australasian Journal of Information Systems*, *26*. https://doi.org/10.3127/AJIS.V26I0.3835

20. Prasad, R., & Choudhary, P. (2021). State-of-the-art of artificial intelligence. *Journal of Mobile Multimedia*, *17*(1–3), 427–454. https://doi.org/10.13052/JMM1550-4646.171322

21. Sandhu, J. (2024, January 25). *What are LLMs and generative AI? A beginner's guide to the technology turning heads — Schwartz Reisman Institute*. University of Toronto. https://srinstitute.utoronto.ca/news/gen-ai-llms-explainer

22. Schneider, J. (2024). *Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda*. http://arxiv.org/abs/2404.09554

23. Shahine, M. (2024, May 5). *Gemini AI: A Breakthrough in Multimodal AI | ProfileTree*. https://profiletree.com/gemini-ai-a-breakthrough-in-multimodal-ai/

24. Siles, A. (2024, October 15). *Exploring the evolution of ChatGPT: From GPT-3 to GPT-4 - Telefónica*. https://www.telefonica.com/en/communication-room/blog/exploring-evolution-chatgpt-gpt-3-gpt-4/

25. Szlezak, W. (2025, March). *DeepSeek and the Evolution of Large Language Models: Cheaper, Better, Faster? | KKR*. https://www.kkr.com/insights/deepseek-large-language-models

26. Toner, H. (2023, May). *What Are Generative AI, Large Language Models, and Foundation Models? | Center for Security and Emerging Technology*. CSET. https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/

27. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and Efficient Foundation Language Models*. https://arxiv.org/abs/2302.13971v1

28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009.

29. xAI. (2025, February 19). *Grok 3 Beta — The Age of Reasoning Agents | xAI*. https://x.ai/news/grok-3

30. Young, B. (2025, March 28). *Evaluating the new Gemini 2.5 Pro Experimental model | Generative-AI – Weights & Biases*. https://wandb.ai/byyoung3/Generative-AI/reports/Evaluating-the-new-Gemini-2-5-Pro-Experimental-model--VmlldzoxMjAyNDMyOA

Developing indicators to assess the quality of scientific research in the digital age

Omar Ali Ismael[1][0000-0002-8384-6532], Ahmed Hani Mohammed [2][0000-0001-7849-6657], Zakariya Y Algamal [3][0000-0002-0229-7958]

[1,2&3] University of Mosul, Mosul, Iraq

Received 22.07.2025/Revised 27.12.2025/Accepted 29.12.2025/Published 30.12.2025

https://doi.org/10.33847/2686-8296.7.2_2

Abstract. The purpose of this paper is to develop indicators to assess the quality of scientific research in the digital age at the University of Mosul. The research relied on the method of a comprehensive survey of all Academic staff at the University of Mosul, with everyone given the opportunity to answer and express an opinion based on the electronic questionnaire that was officially distributed to the official mail, 200 valid responds were obtained for analysis, this purpose poses a great challenge for institutions and organizations that rely on the evaluation of scientific research to make important decisions. The results showed the existence of significant indicators that affect the quality of scientific research in the digital age. While the adoption of technology and the use of technological tools in conducting research, ethical behavior, and the commitment of researchers to ethical standards in research. Promotes the importance of adopting new indicators to improve academic performance and direct investments in priority research areas. This research makes an original contribution by introducing a set of new indicators that are compatible with the requirements of the digital age, thereby enhancing the value of scientific research in light of the rapid changes in the digital environment. The result presents an integrated framework to enhance the quality of academic research and promote collaboration among researchers. At the same time, control variables relied on Scopus publications, H-Index, academic certificate index, and citation count to sort and test the eighteen indicators.

Keywords: Quality of scientific research; Research quality indicators; Digital age; University of Mosul

1. INTRODUCTION AND LITERATURE REVIEW

In the new world, quality is a fundamental factor in any field, including scientific research. The quality of scientific research is referred to as a set of institutional and individual factors that contribute to improving the ability of universities to produce high-quality scientific research [1]. Lee, Ali [2] emphasizes that data quality assurance is an important aspect of transparency and reproducibility of research data, as established data quality assurance processes promote data sharing and reuse. Yang, Lin [3] state that researchers have conducted numerous studies to assess scientific research quality. Researchers have conducted many studies to assess the quality of scientific research. As scientific research continues to develop, it is necessary to focus on developing new indicators to assess its quality in the digital age. Meyer and Schroeder [4] argue that the digital transformation of research has brought about significant changes in how research is produced, published, accessed, and used. So, the Digital transformation is reshaping the research process, leading to more collaborative, open, and interdisciplinary

approaches with diverse outputs. Then, alternative and advanced metrics also have limitations and challenges, such as how to interpret and ensure their reliability and validity and avoid fraud and manipulation [5]. Another opportunity that the digital transformation of search quality providers, as claimed by Lokesh, Shankar [6], is the promotion of open science. According to Robson, Baum [7], Open science can enhance the quality, impact, and value of research by increasing its clarity, repeatability, and reusability. Assessing research quality through quantitative indicators such as citation counts, impact factors, and H-Index poses significant challenges, primarily because of their misuse and misinterpretation. Moreover, these metrics, while used in academic evaluations, often inaccurately reflect the true quality of research. This is evidenced by studies showing weak correlations between these indicators and actual research quality measures, such as statistical rigor and replicability [8]. These metrics can have negative effects on the quality and relevance of research if used incorrectly or uncritically to judge and reward researchers, which was emphasized by Robinson-Garcia, Costas, Nane, & van Leeuwen. To meet this challenge, there is a need for a responsible approach to using metrics, based on the principles of robustness, transparency, diversity, and reflectivity, recognizing limitations, uncertainties, and biases in metrics. At the same time, Herwix, Haj-Bolouri [9] show that the digital transformation of research is an ongoing and evolving process that will continue to shape and challenge future research quality assessment. Therefore, researchers need to be aware of the metrics and factors that affect their field and discipline, which is what he called for [10]. Thus, the measures should be proactive and adaptable in the development and adoption of relevant and contemporary indicators to assess the quality of scientific research [11]. Furthermore, researchers need to be critical and open about the goals of evaluating research quality [12]. Consequently, there is a growing call for the adoption of more responsible and nuanced research evaluation metrics that encompass a broader range of quality indicators [13].

Quality indicators of scientific research

According to Kaplan [14], despite the importance of research quality, there is evidence that current methods for assessing the quality of scientific research may be inadequate. Agreeing with this aspect, Wang and Feng [15] claim that traditional indicators, such as the number of publications and citations, have been criticized for their limited ability to accurately reflect the quality and impact of research. In addition, there is evidence that current peer-review methods, which are often used to assess research quality, can be biased and unreliable. [16]. Given these concerns, it is necessary to explore alternative ways to assess the quality of scientific research, especially in the digital age. According to Zhu, Yang [17] and Prasad [18], one approach is to use a set of standardized criteria to assess the quality of research, such as those developed by the Cochrane Collaboration and GRADE Working Group (Classification of Recommendations, Assessment, Development and Evaluation). Another approach is to develop new metrics that more accurately reflect the quality and impact of research, such as Altmetrics, or quote rate [19, 20].

In research conducted at the University of Mosul, there is a need to evaluate the effectiveness of current methods to assess the quality of research and explore alternative approaches. This is particularly important given the challenges researchers

face at the University of Mosul, including limited access to funding and resources, political instability, and ongoing conflicts in the region.

By improving research quality assessment, the University of Mosul can better support the production of high-quality research that contributes to advancing knowledge and developing the region. Implementing alternative evaluation methods, such as those discussed earlier, can provide a more comprehensive and accurate assessment of the impact and relevance of research conducted at the university. Based on an extensive review of previous relevant studies, we consider the following indicators to measure the quality of scientific research:

1. Digital Presence: The importance of a researcher's digital presence has become an important requirement, as it is essential to enhance the visibility and impact of their work. Studies indicate that a digital presence, characterized by an active engagement on platforms such as Google Scholar, ResearchGate, and social media, significantly enhances research output and facilitates collaboration opportunities [21]. According to Wagner [22], researchers are encouraged to establish authentic connections through social media, which not only helps in disseminating findings but also helps in building a professional reputation in the competitive academic landscape. Furthermore, effective communication strategies on social media can enhance public engagement and trust, especially in combating misinformation, thus reinforcing the importance of a well-managed digital presence [23].

2. Data openness and transparency: Research has pointed to the value of data openness and transparency in evaluating research quality [24]. Therefore, researchers are increasingly encouraged to share their data and findings to enhance the credibility and reliability of their research, in line with open science principles. This behavior not only helps with reproducibility and the construction of new knowledge but also contributes to increased citation rates, as demonstrated by a study showing a 4.3% citation advantage for shared data [25]. In addition, the FAIR (Findable - Accessible - Interoperable - Reusable) principles play an important role in promoting the discoverability and legal reuse of data; however, they require comprehensive documentation to ensure true understanding and reuse [26]. Overall, incorporating these practices promotes a more reliable and trustworthy research environment across disciplines.

3. Global collaboration and research networks: Global collaboration and participation in research networks are crucial to improving research quality and encouraging innovation. Studies in this area show that international collaboration significantly enhances research productivity and increases citations [27]. On the other hand, the characteristics of co-authorship networks reveal that countries with strong collaboration frameworks, such as Germany and the United States, exhibit higher innovation capabilities, as shown by the Global Innovation Index [28]. Finally, these collaborations not only enhance individual research outcomes but also contribute to the global advancement of knowledge and innovation [29].

4. Technology Adoption and Innovation: Adoption of innovations and new technologies enhances research practices, as shown in various studies in different fields. For example, positive faculty attitudes toward the use of technology are associated with improved learning outcomes, although demographic factors influence adoption rates [30]. Furthermore, another study shows that community well-being is

improved through the adoption of technological innovations, highlighting the importance of local values and decision-making processes in the successful integration of technology [31]. Overall, effective management of technology and innovation is essential for the success of an organization, which necessitates a strategic approach to fostering a culture of innovation and alignment with organizational goals [32].

5.   Digital Impact and Response: The ways of measuring research impact have transformed into the digital age, moving beyond traditional citation metrics to include more modern alternative metrics that reflect online engagement. Studies suggest that sophisticated metrics, including social media mentions, downloads, and blog interactions, provide a broader understanding of research impact and community engagement. Verma, Tripathi [19] argue that studies in this area have shown weak correlations between levels of interest in alternative metrics and traditional citations, suggesting that citations remain important but do not fully reflect the immediate impact of research on the digital landscape. Furthermore, access to academic social networks such as ResearchGate and Academia.edu has been associated with increased online visibility and positive correlations with both bibliometric and alternative metrics, which improves researchers' societal impact [33]. Thus, the incorporation of alternative metrics into research evaluation reflects a shift toward recognizing diverse forms of engagement in the evaluation of scholarly work [34].

6.   Performance measurement and data analysis: Research has shown the importance of using digital tools and data analysis techniques to measure research performance and evaluate results [35, 36]. The integration of big data analytics and emerging technologies has led to the development of performance measurement systems, which have necessitated improvements to traditional evaluation methods to accommodate the complexities of data-intensive environments [37]. Moreover, data analysis is crucial to research, enabling researchers to verify the accuracy of data and choose appropriate analytical procedures, thus ensuring reliable conclusions [38]. The use of various analytical methods, such as data mining and visualization, helps in extracting useful information, which ultimately supports informed decision-making processes. Therefore, leveraging these digital tools and techniques is vital to achieving accurate and reliable research outcomes.

7.   Open Access and Community Engagement: Open publishing and community engagement play a critical role in assessing research quality [39]. Heredia and Viggiani [40] confirm that making research available in open-source formats and encouraging community participation and interaction promotes transparency, inclusiveness, and wider knowledge dissemination. Cundeva-Blajer [41] focuses in his study on the importance of promoting open access to support progress in science and research technology and encourage community and academic collaboration. Thus, making research and its results available to the public helps in completing more scientific studies and is thus considered an approach towards the continuous improvement of scientific research.

8.   Social Innovation and Social Impact: Research emphasizes the importance of social innovation in addressing societal issues and its impact on society. de Koning and van der Bijl-Brouwer [42] view social innovation as a sustainable strategy that solves societal issues and promotes social cohesion and environmental responsibility. Sadic, Demir [43] discuss it as a multidisciplinary approach and state that it is closely linked to the Sustainable Development Goals (SDGs), indicating its

potential to facilitate solutions to global challenges. Furthermore, the integration of digital technologies into social innovation has gained momentum, driven by factors such as digital infrastructure and organizational capacities, which can enhance its effectiveness across various sectors, including health and education [44]. Together, these ideas highlight the transformative potential of social innovation in enhancing the quality of scientific research.

9. Digital Literacy and Information Fluency: Studies have highlighted the importance of digital literacy and information fluency in assessing research quality [45, 46]. Researchers need to possess the skills for effective navigation and critical evaluation of digital information sources, thereby ensuring the credibility and reliability of their research.

10. Ethical behavior and research integrity: Ethical behavior and research integrity are key indicators of research quality [47, 48]. Ethical conduct and research integrity are indeed critical indicators of research quality, as they support the credibility and reliability of scientific endeavors. Research integrity is based on principles such as honesty, transparency, and accountability throughout the research process, from design to publication [49]. Ethical considerations are not viewed as mere external requirements but as an integral part of the research process, influencing how studies are conceived and conducted [50]. Furthermore, the role of institutional review boards (IRBs) and ethics committees is vital in maintaining these standards, ensuring that researchers adhere to ethical principles and best practices. Thus, ethical conduct and integrity form the foundation for producing high-quality research that benefits society.

11. Comprehensive and diverse research practices: Inclusive and diverse research practices are an important aspect of enhancing research quality across disciplines. Studies show that integrating diversity into research agendas within universities overcomes institutional barriers and ensures that diverse perspectives contribute to meaningful social change [51]. Additionally, promoting diversity, equity, inclusion, and accessibility (DEIA) in academic settings is associated with improved scientific quality, underscoring the importance of a genuine commitment to these values [52]. Finally, these practices emphasize enhancing research quality by contributing to broader societal benefits; thus, enriching the research landscape and addressing global challenges.

12. Research Impact Assessment: Research impact assessment is seen as a vital measure of research quality, reflecting the importance of assessing the social, economic, and political impacts of research outcomes. Sivertsen and Meijer [53] highlight that societal impact often arises from routine interactions between organizations that use advanced knowledge to achieve their goals, rather than from isolated breakthroughs. Wood and Wilner [54] also emphasize the multidimensional nature of research impact assessment, advocating for cost-effective evaluation models that can be adapted to different institutional contexts. Abudu and Oliver [55] study of cancer survivorship research also demonstrates the importance of understanding subsequent impacts to ensure that funding is translated into tangible benefits for patients. Overall, these studies emphasize the importance of systematic assessment in enhancing the societal relevance of research.

13. Data management and reproduction: Effective data management and reproducibility are essential to ensure the quality of research in various fields.

Researchers are encouraged to adopt robust data management practices that include documentation, organization, storage, and data sharing, which enhance transparency and facilitate the reproducibility of outputs [56]. Implementing the FAIR principles—findable, accessible, interoperable, and reusable—further supports these efforts by promoting organized data sharing and management [57]. In addition, proactive data governance is important, as it informs policies related to data reuse and retention, thus maximizing the usefulness of data throughout the research lifecycle. Thus, by prioritizing comprehensive data management strategies, researchers can ensure the integrity and credibility of their results, ultimately contributing to cumulative knowledge and scientific progress.

14. Interdisciplinary Research: The increased focus on interdisciplinary research and collaboration is another important indicator of the quality of research in the digital age. Researchers are encouraged to engage in interdisciplinary collaborations, leveraging the power of digital tools and platforms to facilitate knowledge sharing, foster innovation, and address complex research questions that require diverse expertise. [58, 59] .

15. Data Privacy and Security: In the age of digital search, ensuring data privacy and security is critical to maintaining search integrity. Researchers should prioritize protecting sensitive data, implementing robust security measures, and adhering to ethical guidelines and data protection regulations to protect participants' privacy and maintain stakeholder trust. [60, 61] .

16. Tools Search: Search tools, including publications, datasets, and applications, are essential to improving research quality and accelerating effective literature reviews. Mutlu, Ulku [62] suggest that the development of applications such as web scraping tools reduces the time required to search the literature, reducing working days to mere hours. Bendeck, Bromley [63] claim that tools such as Slope-Seeker improve data exploration by allowing users to query complex trends through natural language, thereby enhancing the interpretability of analytical data. Open scholarly content search tools, including Google Scholar and Semantic Scholar, also play a crucial role in improving information retrieval skills among researchers [64]. Furthermore, summarization tools that generate concise overviews of research articles help researchers speed up the evaluation of relevant literature, thereby simplifying the search process [65]. Overall, these tools emphasize the importance of well-documented and accessible research outputs in the scientific community.

17. Research Funding and Resource Allocation: Adequate research funding and efficient resource allocation are essential for high-quality research in the digital age [66]. Zhang, Zhang [67] claim that the development of digital finance improves the efficiency of resource allocation by mitigating information asymmetries and enhancing capital utilization, which is essential for sustainable economic growth. In addition, Ou, Zhao [68] discussed the impact of research funding on scientific production and societal vision, and claimed that increasing research funding increases the number of scientific publications and improves academic production in various disciplines. Moreover, systematic reviews of government R&D budget allocation reveal diverse approaches aimed at maximizing the impact of funding on innovation and entrepreneurship, emphasizing the need for advanced optimization techniques [69]. Taken together, these ideas underscore the need for robust funding and strategic resource management to promote high-quality research outcomes.

18. Research Ethics and Responsible Behavior: Adherence to research ethics and responsible conduct is an important criterion for ensuring research quality, as it enhances trust and integrity in scientific research. Researchers are therefore required to adhere to ethical principles, including informed consent, confidentiality, and responsible data management, to protect participants' property and rights and preserve their scientific contribution [70]. Consequently, common forms of research misconduct, such as plagiarism, fabrication, and falsification, undermine scientific credibility and require strong institutional policies to report and deter them [71]. Furthermore, studies in this area show variation in support of responsible research practices, highlighting the need for stronger commitments to ethical standards across institutions [72]. Ultimately, fostering a culture of research integrity requires collaboration among researchers, institutions, and the academic community to uphold ethical standards and enhance the quality of scientific research.

This literature review provides insight into the importance of each indicator for assessing research quality in the digital age. By considering these indicators in assessing research quality in the digital age, researchers can enhance the accuracy, impact, and credibility of their work. By addressing these aspects, researchers can contribute to the advancement of knowledge and the improvement of society in the rapidly evolving digital research landscape. The table presents the indicators studied in this research.

2. METHODOLOGY

2.1. The problem statement

There seems to be an urgent need to develop new indicators to assess the quality of scientific research in the digital age[73]. The field of scientific research has undergone significant changes in recent decades, with the adoption of digital technologies, the increase in the volume of data, and the availability of information at a faster rate [74-76]. Most current evaluation indicators of research performance rely on traditional evaluation models [77, 78], which assess research quality based on the number of articles published in scientific journals and the frequency of citations [79-81]. However, these indicators may suffer from various limitations [82, 83], including the omission of other aspects of research performance, such as scientific impact, innovation, and contribution to community development. Moreover, many current indicators only include quantitative assessments [84-86], while scientific research in the digital age can generate huge amounts of data and information related to research performance[87], which are not effectively evaluated using traditional indicators. With the advancement of digital technology and the immense availability of information, it has become necessary to develop new indicators that help assess the quality of research and are in line with modern requirements.

On the other hand, there is an urgent need to study this topic in the environment of the University of Mosul, as research faces field challenges represented by limited resources, funding, and political instability. These challenges are directly reflected in the university's ability to evaluate the quality of its research based on traditional models. However, previous studies did not focus on developing evaluation indicators that are in line with the requirements of the digital age and did not take

into account the field specificity of research institutions in areas with complex conditions, such as the University of Mosul. Accordingly, the research problem appears in the importance of developing comprehensive research quality evaluation indicators that are in line with the requirements of the digital age and meet the needs of research procedures related to field restrictions, including the University of Mosul.

2.2. Purpose and originality

The lack of clear indicators to assess the quality of scientific research in the digital age poses a major challenge for institutions and organizations that rely on the evaluation of scientific research to make important decisions. Therefore, the main objective of this research will be to develop new and effective indicators to assess the quality of scientific research in the digital age and to meet the needs of researchers and academic institutions in this field. This field is considered an emerging and vital aspect of scientific research, and therefore, research makes a significant contribution to expanding knowledge and bridging existing gaps in this field. The research is based on an advanced and comprehensive research methodology, including a comprehensive review of the relevant scientific literature and the analysis and evaluation of previous studies. The research is characterized by a rigorous and systematic analytical approach aimed at extracting basic ideas and identifying the strengths and weaknesses of the current literature. In addition, research is characterized by creativity in providing new solutions and proposals to enhance the quality of scientific research. The research offers a variety of indicators that reflect different aspects of quality, including methodological accuracy, accuracy, objectivity and innovation.

2.3. Method and measurements

The paper is designed using a checklist covering all eighteen indicators, which are referenced in the literature. Research data was collected during a set period between February to September 2024. On the other hand, the participants were sorted according to key demographic criteria, which were adopted as control variables to sort responses that give higher credibility and strength, as they include the scientific title (professor, assistant professor, lecturer, and assistant lecturer), academic certificate (PhD or master), the number of scientific citations, and the H-Index coefficient, as it depends on these criteria to ensure the accuracy and balance of answers. The comprehensive survey method of the field was used because all lecturers in the faculties of the University of Mosul were selected. The electronic questionnaire form was prepared and distributed by sending it officially from the Presidency of the University to the official e-mail of all approximately 4600 faculty members of the university. 340 valid questionnaires were retrieved for analysis, a figure appropriate for the size of the population studied using the Krejcie-Morgan table. Depending on the criteria of differentiation (Control variables), the respondents who were given a zero value in the publication in Scopus and those who did not receive any quotation were excluded, and after screening, the responses became 200 valid responses for analysis. The data was analyzed using descriptive statistics to explore the percentage of each criterion, and then the impact based on these criteria to ensure

the quality of the answers.

For this study, the term Research Quality Index was adopted to describe the eighteen indicators of the quality of scientific research in the digital age, while the term criterion was used to describe the four criteria for the demographic data of the study sample. On the other hand, the study adopts a measure of the degree of importance of the Index (starting from 1-10) to judge the quality of the Index. In addition, we consider a score of (1-7) as a weak score, while a score of (8-10) represents the threshold for accepting the index.

3. RESULTS

3.1. Demographic Data Analysis

The results of the analysis presented in Table (1) reflect the demographic data of the study sample. The researchers selected four demographic criteria. These criteria enhance the quality of answers. We refer to the scientific title, academic degree, number of citations, and H-Index coefficient as supporting criteria for the quality of answers. The results showed that the percentage of those with the rank of professor and assistant professor reached (46.5%), while the percentage of lecturers and assistant lecturers was higher. This percentage reflects the scientific experience of the study sample. On the other hand, the percentage of Ph.D. holders (64. 5%, which is the largest percentage among the study sample, reflects the best scientific degree to answer the indicators of the quality of scientific research. In addition, the researchers chose the criterion of the number of citations for the researcher; this criterion reflects the value and originality of the published research of the teaching staff of the University of Mosul. More than (50%) of the research sample received (9) citations or more. This percentage indicates a better quality of published research for university academic staff. We identified the H-Index as the last criterion to judge the quality of the answers. The percentage of (67%) or more of the study sample indicates academic staff who have more than (2). This criterion is important, as it reflects the number of published research papers of the researcher with the number of citations for their research. Finally, we rely on these criteria to judge the quality of responses to the research quality indicators in this study (Table 1).

Table 1. Demographic Data Analysis Results

H-Index				Quotations				Certification		Scientific title				Standard Demographic
7 - or more	5-6	2-4	Less than 2	21- and more	11-20	2-10	Less than 2	Master	PhD	Assistant Lecturer	teacher	Assistant Professor	professor	Categories
5	9.5	41	44.5	30	12	35	23	35.5	64.5	14.5	39	35	11.5	**Rate %**

3.2. Analysis of the answers of the study sample on the indicators of the quality of scientific research

Based on the results of the analysis in Table (2) and figure (1), view the results of the study of each of the indicators of the quality of scientific research based on the scale values from (1-10). The percentage values of the agreement were used as a value for accepting the scientific research quality index. Additionally, the study relies on respondents' demographic data to assess the quality of their answers. We check the importance of the indicator based on demographic data. Accordingly, each indicator will be analyzed as follows:
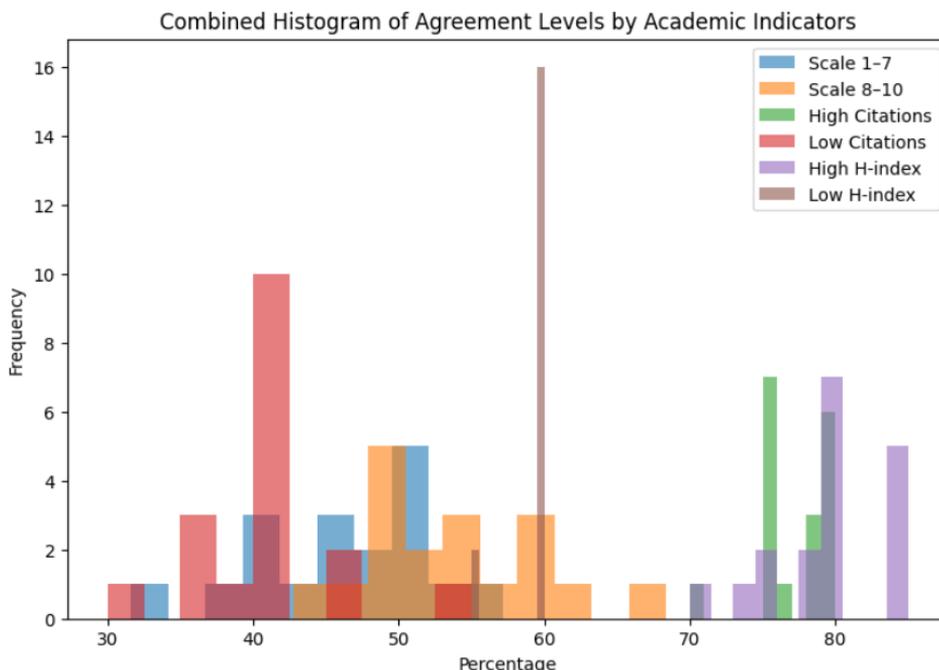

Fig. 1. Effect of Title, Certificate, No. of Citations, and H-Index on Percentage of Admission

Fig. 1 shows that the multiple histograms reveal a substantial variation in agreement levels based on the experience and scientific influence of the participants. Those with high citations and a high H-index tend to show strong and consistent support for the research findings, while those with less influence exhibit moderate or cautious agreement. These results support the robustness and construct validity of the research instrument, as well as the reliability of the conclusions drawn.

Table 2. Effect of Academic Title, Certificate, Number of Citations, and H-Index on Percentage of Admission

N.	Indicators	Scale	Size	Title				Citation#		H-index$		Acceptance
		01.07.25	08.10.25	Prof.	Lecturer	PhD	Master	High	Low	High	Low	
1	Digital Presence	49.50%	50.50%	75%	50%	80%	60%	70%	40%	75%	55%	+
2	Data transparency	40.00%	60.00%	65%	45%	70%	50%	75%	55%	80%	60%	+
3	Global Cooperation	45.00%	55.00%	70%	40%	75%	55%	80%	45%	85%	60%	+
4	Technology Adoption	41.23%	58.77%	68%	45%	76%	54%	78%	38%	80%	60%	+
5	Digital Impact	52.04%	47.96%	65%	40%	72%	50%	76%	45%	74%	55%	-
6	Performance Analysis	37.50%	62.50%	70%	50%	78%	50%	75%	40%	80%	60%	+
7	Open Access	50.50%	49.50%	65%	30%	70%	40%	78%	40%	75%	60%	-
8	Social Innovation	44.50%	55.50%	60%	30%	70%	45%	75%	30%	78%	60%	+
9	Use of digital media	44.00%	56.00%	60%	35%	70%	45%	75%	40%	80%	60%	+
10	Ethical behavior	31.62%	68.38%	70%	30%	75%	50%	80%	40%	85%	60%	+
11	Inclusive Practices	49.62%	50.38%	68%	35%	75%	50%	78%	40%	70%	60%	+
12	Research Impact	53.54%	46.46%	65%	30%	70%	40%	75%	35%	78%	60%	-
13	Data Management	41.00%	59.00%	70%	45%	75%	50%	80%	40%	80%	60%	+
14	Interdisciplinary research	57.21%	42.79%	70%	30%	75%	50%	80%	40%	85%	60%	-
15	Data Privacy	48.84%	51.16%	65%	40%	70%	50%	75%	35%	80%	60%	+
16	Research Tools	52.00%	48.00%	70%	30%	75%	50%	80%	40%	85%	60%	-
17	Research Funding	51.26%	48.74%	68%	40%	72%	45%	75%	35%	80%	60%	-
18	Research Ethics	45.14%	54.86%	70%	30%	75%	50%	80%	40%	85%	60%	+

Columns of Table 2:

Title for Professor (Prof.) and Lecturer indicate the percentage of scientific titles that support the ratio 8-10

Title for PhD and Master Indicate the percentage of academic degrees that gave an agreement ratio of 8-10

\# - indicates the percentage of individuals who agreed by 8-10, depending on the number of citations indicator

$ - refers to individuals with a high H-Index who agreed to 8-10

Digital presence: Digital presence reflects the ability of researchers to appear

on digital platforms. (49.5%) of the study, the sample chose a score of 1-7, which indicates poor presence, while those who gave a score of 8-10 were (50.5%), which shows the possibility of using the platforms well. The results of demographic data have been added to this indicator. The scientific title shows that the respondents with the rank of professor and assistant professor contributed (75%) to this indicator. This percentage shows a better digital presence for those with experience and knowledge compared to academic staff who voted on this indicator (50%). The results of the academic certificate reflected that those who obtained a PhD have a digital presence by (80%) compared to those who obtained a master's degree, which reached (60%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (70%) show a greater digital presence compared to those with a low citation count (40%). Finally, the H-Index criterion shows that those with a high H-Index (75%) are people with a higher digital presence than those with a low H-Index (55%).

Data abundance and transparency: The concept of data abundance and transparency reflects the willingness of researchers to share their data and research openly. (40%) of the study sample chose a score of 1-7, an assessment that indicates poor abundance and transparency in data, while those who gave a score of 8-10 were (60%), an assessment that shows the great importance of transparency in enhancing the quality of scientific research, and the results of the demographic data for this indicator indicate that the scientific title has the degree of professor and assistant professor Positive impact (65%) contributed to this indicator, which indicates that those with industry experience are more inclined to support transparency compared to academic staff (45%) When it comes to academic degrees, those with a PhD have a greater interest (70%) in abundance and transparency compared to master's degree holders, who support it (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflected a high interest in transparency, compared to those with a low citation count (55%). Finally, the H-Index criterion shows high H-Index holders (80%), suggesting that these researchers have a greater incentive to support abundance and transparency in data than those with a low H-Index (60%).

Global Collaborations and Research Networks: Global collaborations and research networks represent the extent to which researchers can collaborate across borders to improve results. 45% of respondents rated poorly (1-7), while 55% felt that collaboration was very important (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (70%) to this index, indicating that those with experience in the field are more inclined to international cooperation compared to Lecturers (40%). In terms of academic degrees, PhDs have a greater interest in international cooperation (75%), compared to master's degree holders (55%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of collaboration in improving the quality of research, compared to those with a low citation count (45%). Finally, the H-Index criterion shows that high H-Index holders (85%), suggesting that these researchers have more support for promoting global collaboration than those with a low H-Index (60%).

Technology and innovation adoption: Technology and innovation are key to improving the quality of research. We found that (41.23%) of respondents rated this

aspect twice (1-7), while (58.77%) saw it as critical (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (68%) to this index, indicating that those with experience in the field are more inclined to rely on technology compared to Lecturers (45%). In terms of academic degrees, Ph.D. holders have a greater interest in technology (76%), compared to master's degree holders (54%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (78%) reflected their greater interest in using technology, compared to those with a low citation count (38%). Finally, the H-Index standard shows that high H-Index holders (80%), suggesting that these researchers have more support for technology and innovation adoption than those with a low H-Index (60%).

Digital impact and response: Digital influence refers to the ability of researchers to influence their scientific communities. (52.04%) of the study sample chose a score from 1 to 7, indicating a weakness in the digital effect, while (47.96%) gave higher scores (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (65%) to this index, indicating that those with industry expertise are more aware of the importance of a digital impact than Lecturers (40%). In terms of academic degree, PhDs have a greater interest in digital impact (72%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (76%), reflecting the importance of digital influence in improving the quality of research, compared to those with a low citation count (45%). Finally, the H-Index standard shows that high H-Index holders (74%), suggesting that these researchers have more support for enhancing digital impact than those with a low H-Index (55%).

Performance measurement and data analysis: Performance measurement and data analysis are the cornerstones of improving the quality of research. (37.50%) of the study sample showed weakness in measuring performance (1-7), while (62.50%) considered it an essential element (8-10). The scientific title shows that respondents with the rank of professor and assistant professor contributed (70%) to this indicator, which indicates that those with experience in the field are more likely to appreciate performance measurement compared to lecturers (50%). In terms of academic degree, PhDs have a greater interest in measuring performance (78%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of measuring performance in improving the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients (80%), suggesting that these researchers have more support for enhanced performance measurement and data analysis than those with a low H-Index (60%).

Open Access and Community Engagement: Open access reflects the desire of researchers to make their research findings available to the public. (50.50%) showed a poor rating (1-7), while (49.50%) gave higher scores (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (65%) to this indicator, indicating that those with industry expertise are more likely to support open access than lecturers (30%). In terms of academic degree, PhDs have a greater interest in open access (70%), compared to master's degree holders (40%). On the other hand, the number of citations criterion indicated that researchers with a high

citation count (78%), reflecting the importance of open access in improving the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients were (75%), suggesting that these researchers have more support for promoting open access and community engagement than those with a low H-Index (60%).

Social innovation and social impact: Social innovation refers to the ability of researchers to make a positive impact in society. (44.50%) showed weakness in this aspect (1-7), while (55.50%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed 60% to this indicator, indicating that those with industry expertise are more likely to support social innovation than lecturers (30%). In terms of academic degree, PhDs have a greater interest in social innovation (70%), compared to master's degree holders (45%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of social innovation in improving the quality of research, compared to those with a low citation count (30%). Finally, the H-Index standard shows that high H-Index recipients (78%), suggesting that these researchers have more support for promoting social innovation and social impact than those with a low H-Index (60%).

Ability to use digital media and digital culture: The ability to use digital media is an essential skill in modern research. (44%) showed weakness in this aspect (1-7), while (56%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed 60% to this index, indicating that those with industry expertise are more likely to support the use of digital media than lecturer (35%). In terms of academic degree, PhDs have a greater interest in using digital media (70%), compared to master's degree holders (45%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of digital culture in improving the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index holders (80%), suggesting that these researchers have more support for promoting digital media use and digital culture than those with a low H-Index (60%).

Ethical behavior and research integrity: Ethical behavior is essential to research success. (31.62%) showed weakness in ethics (1-7), while (68.38%) gave a higher rating (8-10). The scientific title shows that respondents with the degree of professor and assistant professor contributed (70%) to this index, which indicates that those with experience in the field are more aware of the importance of ethics than lecturers (30%). In terms of academic degrees, PhDs have a greater interest in ethics (75%) compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of ethics in enhancing the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients (85%), suggesting that these researchers have greater support for promoting ethical behavior and research integrity than those with a low H-Index (60%).

Inclusive and diverse research practices: Holistic research practices refer to the ability of researchers to integrate multiple perspectives into their research. (49.62%) showed weakness in this aspect (1-7), while (50.38%) gave higher grades

(8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (68%) to this indicator, indicating that those with industry expertise are more likely to support inclusiveness than lecturers (35%). In terms of academic degrees, Ph.D. holders have a greater interest in inclusive practices (75%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (78%), reflecting the importance of diversity in enhancing the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients (70%), suggesting that these researchers have more support for promoting inclusive and diverse research practices than those with a low H-Index (60%).

Research Impact Assessment: Research Impact Assessment allows researchers to measure the results of their research and its impact on society. (53.54%) showed weakness in this aspect (1-7), while (46.46%) gave higher grades (8-10). The scientific title shows that respondents with the rank of professor and assistant professor contributed (65%) to supporting impact evaluation, indicating that those with experience in the field are more aware of the importance of impact evaluation compared to lecturer (30%). In terms of academic degrees, Ph.D. holders have a greater interest in impact evaluation (70%), compared to master's degree holders (40%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of evaluating the impact in improving the quality of research, compared to those with a low citation count (35%). Finally, the H-Index criterion shows that high H-Index recipients (78%), suggesting that these researchers have more support to enhance research impact assessment compared to those with a low H-Index (60%).

Data management and reiteration: Data management is an essential part of scientific research. (41%) showed weakness in this aspect (1-7), while (59%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (70%) to supporting data management, indicating that those with experience in the field are more aware of the importance of this aspect than lecturers (45%). In terms of academic degree, PhDs have a greater interest in data management (75%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of data management in enhancing the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients (80%), suggesting that these researchers have greater support for enhanced data management and redundancy than those with a low H-Index (60%).

Interdisciplinary research: Interdisciplinary research encourages the integration of knowledge in different fields. (57.21%) showed strength in this aspect (1-7), while (42.79%) gave higher grades (8-10). The scientific title shows that respondents with the rank of professor and assistant professor contributed (70%) to supporting interdisciplinary research, indicating that those with expertise in the field are more aware of the importance of this type of research than lecturer (30%). In terms of academic degree, Ph.D. holders have a greater interest in interdisciplinary research (75%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of interdisciplinary research in enhancing the quality

of research, compared to those with a low citation count (40%). Finally, the H-Index criterion shows that those with a high H-Index (85%), suggesting that these researchers have more support for promoting interdisciplinary research compared to those with a low H-Index (60%).

Data privacy and security: Data privacy is vital to protecting information. (48%) showed weakness in this aspect (1-7), while (52%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (65%) to supporting privacy, indicating that those with industry expertise are more aware of the importance of privacy than lecturers (40%). In terms of academic degree, PhDs have a greater interest in data privacy (70%), compared to master's degrees holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of privacy in enhancing the quality of research, compared to those with a low citation count (35%). Finally, the H-Index standard shows that high H-Index holders (80%), suggesting that these researchers have greater support for enhancing data privacy and security than those with a low H-Index (60%).

Research Tools: The quality of research results reflects the accuracy and reliability of research. (52%) showed poor quality results (1-7), while (48%) gave higher scores (8-10). The scientific title shows that the respondents with the degree of professor and assistant professor contributed (70%) to supporting the quality of results, which indicates that those with experience in the field are more aware of the importance of good results than lecturers (30%). In terms of academic degree, Ph.D. holders have a greater interest in the quality of results (75%), compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of quality, results in enhancing research, compared to those with a low citation count (40%). Finally, the H-Index criterion shows that high H-Index scorers (85%), suggesting that these researchers have more support for enhancing the quality of search results than those with a low H-Index (60%).

Research Funding and Resource Allocation: Research funding and resource allocation reflect researchers' ability to obtain the necessary financial support for their research. (51.26%) showed weakness in funding (1-7), while (48.74%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (68%) to funding support, indicating that those with industry expertise are more aware of the importance of funding than lecturers (40%). In terms of academic degree, Ph.D. holders have a greater interest in funding (72%) compared with master's degrees holders (45%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (75%), reflecting the importance of funding in enhancing the quality of research, compared to those with a low citation count (35%). Finally, the H-Index criterion shows that those with a high score were (80%), suggesting that these researchers have more support to enhance research funding and resource allocation than those with a low H-Index (60%).

Research Ethics and Responsible Behavior: Ethical behavior represents researchers' commitment to the fundamental principles of research. (45.14%) showed weakness in ethics (1-7), while (54.86%) gave higher grades (8-10). The scientific title shows that Professor and Assistant Professor respondents contributed (70%) to supporting ethics, indicating that those with expertise in the field are more

aware of the importance of ethics than lecturers (30%). In terms of academic degree, PhDs have a greater interest in ethics (75%) compared to master's degree holders (50%). On the other hand, the number of citations criterion indicated that researchers with a high citation count (80%), reflecting the importance of ethics in enhancing the quality of research, compared to those with a low citation count (40%). Finally, the H-Index standard shows that high H-Index recipients (85%), suggesting that these researchers have greater support for promoting ethics and responsible behavior in research than those with a low H-Index (60%).

4. CONCLUSION

This study is a pioneering step towards enhancing the quality of scientific research in the digital age by providing an integrated set of indicators that reflect the requirements of modern research. The study started with the analysis of a sample of 200 lecturers at the University of Mosul, where accurate statistical methods were used to evaluate the 18 indicators affecting the quality of research collected based on the literature in this field.

The results showed that 12 indicators affect the quality of scientific research in the digital age, including digital presence, transparency in data, global cooperation, technological adoption, performance analysis, social innovation, use of digital media, ethical behavior, inclusive practices, data management, data privacy, and research ethics. The data showed that these indicators are positively correlated with the quality of research and contribute to enhancing the competitiveness of scientific research. In contrast, some indicators did not show Others, such as digital impact, open access, research impact, interdisciplinary research, research tools, and research funding. Resilience in the face of rapid technological changes did not reflect a significant impact in improving the quality of research, suggesting the need to reassess or adjust it to suit the contemporary academic context.

Acceptable indicators contribute to guiding educational institutions toward adopting effective policies that enhance the quality of scientific research. The study also highlights the importance of enhancing the digital presence and opening channels of international cooperation to raise the level of scientific research and achieve more impactful results.

In conclusion, this research provides a comprehensive framework that contributes to improving the quality of academic research at the University of Mosul and other academic institutions. These institutions must adopt the recommended indicators to enhance research practices and create a learning and research environment that is compatible with the rapid developments in the digital age.

ACKNOWLEDGEMENT

REFERENCES

1. Aga, F.J., *Determinants of academic research quality: The case at three Ethiopian public universities.* Social Sciences & Humanities Open, 2023. **7**(1): p. 100409.
2. Lee, D.J., et al., *Toward the Conceptualization of Data Quality Assurance in Research Data Repositories.* Proceedings of the Association for Information Science and Technology, 2023. **60**(1): p. 1019-1021.
3. Yang, X., X. Lin, and X. Ma, *Does interdisciplinarity improve research performance? An empirical study of faculty members at Chinese universities.* Scientometrics, 2024. **129**(12): p. 7645-7666.
4. Meyer, E.T. and R. Schroeder, *Knowledge machines: Digital transformations of the sciences and humanities*. 2023: Mit Press.
5. Margherita, A., G. Elia, and C.J.S. Petti, *What Is Quality in Research? Building a Framework of Design, Process and Impact Attributes and Evaluation Perspectives.* 2022. **14**(5): p. 3034.
6. Lokesh, C., et al., *Digital Transformation and Hybrid Model in Engineering Education.* 2023. **36**(Special Issue 2).
7. Robson, S.G., et al., *Promoting open science: a holistic approach to changing behaviour.* 2021. **7**(1): p. 30137.
8. Pontika, N., et al., *Indicators of research quality, quantity, openness, and responsibility in institutional review, promotion, and tenure policies across seven countries.* Quantitative Science Studies, 2022. **3**(4): p. 888-911.
9. Herwix, A., et al., *Ethics in information systems and design science research: five perspectives.* 2022. **50**(1): p. 589-616.
10. Aithal, P.S. and S. Aithal, *Key Performance Indicators (KPI) for Researchers at Different Levels & Strategies to Achieve it.* International Journal of Management, Technology, and Social Sciences, 2023. **8**(3): p. 294-325.
11. Andreassen, H.K., *Research assessments*, in *Gender Diversity, Equity, and Inclusion in Academia*. 2023, Routledge. p. 193-203.
12. Steccolini, I., *Methodological Paper What counts as "good" qualitative accounting research? Researchers' perspectives on assessing and proving research quality.* Accounting, Auditing & Accountability Journal, 2022. **36**(3): p. 1032-1057.
13. Raju, J. and A. Mfengu, *Challenges with measures used for assessing research impact in higher education institutions.* South African Journal of Science, 2024. **120**(7/8): p. 1-6.
14. Kaplan, D.J., *Editorial Commentary: Medical Research Study Quality Could Improve if Authors Consider General and Condition-Specific Study Quality Scoring Systems When Designing Their Investigations.* Arthroscopy, 2024. **40**(3): p. 970-971.
15. Wang, X. and X. Feng, *Research on the relationships between discourse leading indicators and citations: perspectives from altmetrics indicators of international multidisciplinary academic journals.* Library Hi Tech, 2024. **42**(4): p. 1165-1190.
16. Yarborough, M.J.B.O.S., *Moving towards less biased research.* 2021. **5**(1).
17. Zhu, L., et al., *Citation of updated and co-published Cochrane Methodology Reviews.* Systematic Reviews, 2023. **12**(1): p. 120.
18. Prasad, M., *Introduction to the GRADE tool for rating certainty in evidence and recommendations.* Clinical Epidemiology and Global Health, 2024. **25**: p. 101484.
19. Verma, M.K., M. Tripathi, and B. Basumatary, *Does Altmetric Attention Score Correlate with Citations of Articles Published in High CiteScore Journals.* DESIDOC Journal of Library & Information Technology, 2023. **43**(06): p. 432-440.
20. Ukidve, S., et al. *Evaluating Research Impact: A Comprehensive Overview of Metrics and Online Databases*. Springer.

21. Andrea, B.-A., et al. *Visibility of Scientific Production and Digital Identity of Researchers through Digital Technologies*. Education Sciences, 2022. **12**, DOI: 10.3390/educsci12120926.

22. Wagner, E., *How to Create a Digital Presence: A Review of Digital Leadership in Higher Education by Josie Ahlquist.* eLearn, 2023. **5**(2023).

23. Cinelli, M., et al., *Promoting engagement with quality communication in social media.* PLoS One, 2022. **17**(10): p. e0275534.

24. Conrad, T.O.F., et al., *Making Mathematical Research Data FAIR: Pathways to Improved Data Sharing.* Scientific Data, 2024. **11**(1): p. 676.

25. Colavizza, G., et al., *An analysis of the effects of sharing research data, code, and preprints on citations.* PLoS One, 2024. **19**(10): p. e0311493.

26. Hermann, S. and J. Fehr, *Documenting research in simulation science to enhance understanding for reusability.* R Soc Open Sci, 2024. **11**(10): p. 240776.

27. Vanz, S.A.d.S. and D. Docampo, *The Influence of International Scientific Collaboration with English-speaking Countries on the Research Performance of Brazilian Academic Institutions.* Journal of Scientometric Research, 2023. **11**(3): p. 358-370.

28. Pepe, C.G.E., M.V.A. Fonseca, and C.F. Silva Marques, *International collaboration towards innovation management: a network perspective and the Global Innovation Index.* Journal of Innovation and Entrepreneurship, 2024. **13**(1): p. 32.

29. Neu, C. *Exploring Opportunities for the Enhancement of International Academic Collaboration*. in *Chances and Challenges of Digital Management*. 2023. Cham: Springer Nature Switzerland.

30. Joshi, P. and N.P. Nyaupane, *Adoption and Integration of Modern Teaching Technologies in Higher Education: A Quantitative Study of Faculty Practices.* NPRC Journal of Multidisciplinary Research, 2024. **1**(6): p. 142-167.

31. Adianto and R.V. Vani, *The Adoption Model of Technological Innovations That Have an Impact on the Welfare of the Community in the Village of Koto Masjid, Riau Province.* International Journal of Social Science and Business, 2022. **6**(3): p. 446-451.

32. Yayha, M.A., et al., *Innovation and Technology Management in Modern Enterprises.* Journal of Ecohumanism, 2024. **3**(5): p. 895-908.

33. Adriaanse, L., C.J.P. Niemand, and C. Rensleigh, *Altmetric and bibliometric indicators for academic social networks: Academia.edu & ResearchGate.* South African Journal of Information Management, 2024. **26**(1): p. 1869.

34. Dutt, E. and J.R. Hurst, *Comparing traditional bibliometrics and Altmetric assessments of research impact in respiratory disease.* ERJ Open Res, 2023. **9**(6).

35. Hermanu, A.I., et al., *Higher education research performance parameters classified by systems theory: antecedents for the development of assessment models.* 2023.

36. Krüger, A.K. and S.J.V.S. Petersohn, *From Research Evaluation to Research Analytics. The digitization of academic performance measurement.* 2022. **9**(1): p. 11-46.

37. Ullah Khan, M.S., *Performance Measurement in the Digital Era: Review Literature at the Intersection of Management Accounting, Big Data Analytics and Emerging Technologies' Impact on Performance Evaluation.* International Journal of Research and Scientific Innovation, 2024. **XI**(VIII): p. 94-101.

38. Thompson, D.L., *Data Analysis*, in *Conducting Undergraduate Research in Education*. 2022, Routledge. p. 94-108.

39. Susi, T., et al., *Centrality of researchers in reforming research assessment.* 2022. **162**: p. 218-223.

40. Heredia, A. and E. Viggiani, *Research and Collaborative Working and Sharing Online*, in *Encyclopedia of Libraries, Librarianship, and Information Science*, D. Baker and L. Ellis, Editors. 2025, Academic Press: Oxford. p. 155-167.

41. Cundeva-Blajer, M., *Research infrastructure open access approach – Implementation challenges and solutions for accredited calibration laboratories.* Measurement: Sensors, 2024: p. 101613.

42. de Koning, J. and M. van der Bijl-Brouwer, *Value Dimensions in Creative Collaborations for Social Innovation.* She Ji: The Journal of Design, Economics, and Innovation, 2024. **10**(3): p. 286-307.

43. Sadic, S., E. Demir, and J. Crispim, *Towards a connected world: Collaborative networks as a tool to accomplish the SDGs.* Journal of Cleaner Production, 2024. **462**: p. 142726.

44. Chen, H., *Research Progress of Digital Social Innovation.* International Journal of Social Sciences and Public Administration, 2024. **3**(3): p. 359-364.

45. Bhattacharya, M. *Enhancing the Quality of Research using Smart eResearch Tools and Technologies*. in *Proceedings of the 13th International Conference on Education Technology and Computers*. 2021. Association for Computing Machinery.

46. Jungerth, M. and L. Andersson, *Developing Skills Through Digital Project-Based Learning: Dual case study of Linköping University and WSP Sweden AB*. 2022.

47. Oey-Gardiner, M., F. Rahardi, and C.K. Can, *Ethics During Research*, in *Ethics in Social Science Research in Indonesia*. 2023, Springer. p. 59-92.

48. Vie, K.J.J.A.i.R., *Can research integrity prevail in the market? Lessons from commissioned research organizations.* 2022. **29**(7): p. 415-441.

49. Zhaksylyk, A., et al., *Research Integrity: Where We Are and Where We Are Heading.* J Korean Med Sci, 2023. **38**(47): p. e405.

50. Smyth, B.M., M.A. Martin, and M. Downing, *The Routledge Handbook of Human Research Ethics and Integrity in Australia*. 2024, Taylor & Francis.

51. Hernández-Lloret, C.M., et al., *Diversity and Research in Spanish Universities. Weaknesses, Strengths, And Good Practices for Inclusive Research.* Journal of Education Culture and Society, 2024. **15**(2): p. 723-745.

52. Moreira-Bouchard, J.D., et al., *Creating diversity, equity, inclusion, and accessibility statements for your CV: a resource guide to effectiveness and comprehensiveness.* American Journal of Physiology-Heart and Circulatory Physiology, 2024. **327**(6): p. H1376-H1383.

53. Sivertsen, G. and I. Meijer, *Evaluating and Improving the Societal Impact of Research*, in *Challenges in Research Policy*, G. Sivertsen and L. Langfeldt, Editors. 2025, Springer Nature Switzerland: Cham. p. 21-28.

54. Wood, T. and A. Wilner, *Research impact assessment: Developing and applying a viable model for the social sciences.* Research Evaluation, 2024: p. rvae022.

55. Abudu, R. and K. Oliver, *Identifying Research Impacts Within a Global Cancer Survivorship Grants Portfolio.* JCO Global Oncology, 2024. **10**(Supplement_1): p. 97-97.

56. Cunha-Oliveira, T., J.P.A. Ioannidis, and P.J. Oliveira, *Best practices for data management and sharing in experimental biomedical research.* Physiol Rev, 2024. **104**(3): p. 1387-1408.

57. Fouad, K., et al., *A practical guide to data management and sharing for biomedical laboratory researchers.* Exp Neurol, 2024. **378**: p. 114815.

58. Thelwall, M., et al., *Does the perceived quality of interdisciplinary research vary between fields?* 2023.

59. Leahey, E. and S.N.J.R.P. Barringer, *Universities' commitment to interdisciplinary research: To what end?* 2020. **49**(2): p. 103910.

60. Victor, B.G., et al., *Recommendations for Social Work Researchers and Journal Editors on the Use of Generative AI and Large Language Models.* 2023. **14**(3): p. 000-000.

61. Jansen, J. and D. Visser, *On Becoming a Scholar: What Every New Academic Needs to Know*. 2022, African Minds.

62. Mutlu, M.A., E.E. Ulku, and K. Yildiz, *A web scraping app for smart literature search of the keywords.* PeerJ Comput Sci, 2024. **10**: p. e2384.

63. Bendeck, A., D. Bromley, and V. Setlur, *SlopeSeeker: A Search Tool for Exploring a Dataset of Quantifiable Trends*, in *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 2024, Association for Computing Machinery: Greenville, SC, USA. p. 817-836.

64. Redkina, N.S., *Impact of Open Science Infrastructure on the Development of the World Information Resources Market.* Scientific and Technical Information Processing, 2024. **51**(2): p. 161-172.

65. Garg, S., et al., *An Efficient Summarisation and Search Tool for Research Articles.* Procedia Computer Science, 2024. **235**: p. 2215-2226.

66. Chen, B., et al., *Do Financial Investment, Disciplinary Differences, and Level of Development Impact on the Efficiency of Resource Allocation in Higher Education: Evidence from China.* Sustainability, 2023. **15**(9): p. 7418.

67. Zhang, Q., et al., *Research on the impact of digital finance on the high-quality development of education.* Finance Research Letters, 2024. **70**: p. 106188.

68. Ou, G., et al., *Effects of research funding on the academic impact and societal visibility of scientific research.* Journal of Informetrics, 2024. **18**(4): p. 101592.

69. Alizadeh, P. and M. Gholipour Domyeh, *How to optimize government R&D budget allocation? A review of recent studies and research gaps.* Journal of Science and Technology Policy Management, 2024. **ahead-of-print**(ahead-of-print).

70. Wazim, R.S., *Ethical integrity in doctoral research: A focus on classroom remediation.* World Journal of Advanced Research and Reviews, 2024. **21**(2): p. 1190-1194.

71. Weber-Mandrin, M., M. Daphinoff, and I.C. Weber, *Legal and Operational Aspects of Compliance with Scientific Integrity.* Chimia (Aarau), 2024. **78**(9): p. 589-593.

72. Butt, F.A., et al., *The quality and reporting of recommendation documents to enhance the integrity of clinical trials: A systematic review and critical appraisal.* Semergen, 2024. **51**(2): p. 102333.

73. Smith, S., et al., *Development of person‐centred quality indicators for aged care assessment services in Australia: A mixed methods study.* Health Expectations, 2024. **27**(1): p. e13958.

74. Bellotti, P., et al., *Recent advances in the chemistry and applications of N-heterocyclic carbenes.* Nature Reviews Chemistry, 2021. **5**(10): p. 711-725.

75. Sarstedt, M., et al., *Progress in partial least squares structural equation modeling use in marketing research in the last decade.* Psychology & Marketing, 2022. **39**(5): p. 1035-1064.

76. Rosário, A. and R. Raimundo, *Consumer marketing strategy and E-commerce in the last decade: a literature review.* Journal of theoretical and applied electronic commerce research, 2021. **16**(7): p. 3003-3024.

77. Chang, Y., et al., *A survey on evaluation of large language models.* ACM Transactions on Intelligent Systems and Technology, 2024. **15**(3): p. 1-45.

78. Camilleri, M.A., *Evaluating service quality and performance of higher education institutions: a systematic review and a post-COVID-19 outlook.* International Journal of Quality and Service Sciences, 2021. **13**(2): p. 268-281.

79. Tang, K.-Y., C.-Y. Chang, and G.-J. Hwang, *Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998–2019).* Interactive Learning Environments, 2023. **31**(4): p. 2134-2152.

80. Wu, H., et al., *Mapping knowledge structure and themes trends of osteoporosis in rheumatoid arthritis: a bibliometric analysis.* Frontiers in Medicine, 2021. **8**: p. 787228.

81. Rejeb, A., et al., *The Internet of Things (IoT) in healthcare: Taking stock and moving forward.* Internet of Things, 2023. **22**: p. 100721.

82. Geng, C., Y. Zhang, and L. Yang. *Construction of Scientific Research Management Platform in Higher Vocational Colleges Based on Big Data Technology*. in *Proceedings of the 2024 9th International Conference on Cyber Security and Information Engineering*. 2024.

83. Kang, Y. and R. Liu, *Does the merger of universities promote their scientific research performance? Evidence from China.* Research Policy, 2021. **50**(1): p. 104098.

84. Li, Z. and Z. Qi, *[Retracted] Identification of Scientific Research Evaluation Indicators of College Teachers Based on Wireless Communication Network.* Mobile Information Systems, 2022. **2022**(1): p. 5408382.

85. Yiqing, J., *OVERVIEW OF KEY INDICATORS OF THE DEVELOPMENT OF THE ACADEMIC INDUSTRY IN CHINA.* Государственное и муниципальное управление. Ученые записки, 2022(2): p. 142-151.

86. Wang, Y. and J. Huang, *Research on the present situation and improvement countermeasures of the performance evaluation of scientific research of music teachers in colleges and universities.* Journal of Roi Kaensarn Academi, 2023. **8**(7): p. 672-680.

87. Ban, O.I., et al., *Evaluation of the Quality of Higher Education Services by Revised IPA in the Perspective of Digitization.* Sustainability, 2024. **16**(7): p. 3017.

A Semi-Automated Technique for Cadastral Boundary

Binyam Zeray Abrham, Tebarek Lika, Asnake Mekuriaw

Addis Ababa University, Addis Ababa, Ethiopia

Abstract. The 2030 Agenda for Sustainable Development recognizes that land is essential to achieving the Sustainable Development Goals globally. Yet, a large portion of land rights worldwide are still unregistered in government-sanctioned systems. To mitigate this problem, the Fit-for-Purpose (FFP) approach to land administration was introduced, aiming to simplify cadastral mapping and lower the cost and time required by traditional surveying techniques. This study investigates the application of UAV imagery coupled with deep learning models, specifically BDCN and HED, for automated cadastral boundary extraction. Results demonstrate that both models can accurately identify boundaries, with BDCN achieving 0.68 precision, 0.80 recall, 0.67 F1 score, and 0.85 IoU, while HED scored 0.66 precision, 0.68 recall, 0.73 F1 score, and 0.88 IoU. This research shows the potential of deep learning and UAV-based workflows to support efficient and affordable land registration, especially in rural settings, in alignment with FFP principles.

Keywords: Land administration; Cadastral mapping; UAV imagery; Deep learning; BDCN; HED; Fit-for-Purpose

1. Introduction

Land serves as a crucial driver for sustainable development and is integrated into the core objectives of the UN 2030 Agenda (United Nations, 2015). However, globally, only about 30% of land rights are formally recorded in government systems, leaving most landholders, especially in the Global South, vulnerable to tenure insecurity, land disputes, and exclusion from formal economic systems (Bennett & Alemie, 2016).

In many developing countries, inefficient and costly land registration processes further exacerbate these challenges. Traditional cadastral mapping relies heavily on extensive field surveys and manual processing, which are not only time-consuming but also require considerable technical resources and skilled personnel (Zevenbergen, 2004; Enemark & McLaren, 2017). These constraints are particularly acute in rural areas where formal land information systems are weak or nonexistent, and spatial data coverage is often limited (Metaferia et al., 2023b).

To address these issues, the Fit-for-Purpose (FFP) approach to land administration was proposed. This framework emphasizes affordability, flexibility, and scalability over technical rigidity, promoting the use of general boundaries, participatory methods, and readily available geospatial technologies (Enemark & Mclaren, 2017) By leveraging high-resolution imagery from UAVs and adopting modern digital tools, FFP enables countries to rapidly and cost-effectively build inclusive land administration systems(Crommelinck et al., 2017).

Remote sensing and photogrammetry play a crucial role in this transition, Providing alternatives to labor-intensive fieldwork. UAVs, in particular, provide timely, high-resolution imagery suitable for small-scale cadastral mapping. UAVs are especially effective in rural and peri-urban contexts, where traditional surveying is

often impractical or prohibitively expensive (Bennett et al., 2021; Tuladhar Arbind Man, 2004)

However, extracting meaningful cadastral information from UAV imagery remains a technical challenge. Manual digitization is still commonly used and introduces operator biases and inefficiencies. Recent advancements in deep learning and computer vision offer promising pathways toward automation. Convolutional Neural Networks (CNNs), known for their ability to learn complex spatial patterns, are now being used to perform semantic segmentation and edge detection for geospatial applications, including cadastral boundary extraction (Carrio et al., 2017).

This study builds on these developments by evaluating the performance of two deep learning models—Bilateral Divisive Contrast Network (BDCN) and Holistically-Nested Edge Detection (HED)in extracting cadastral boundaries from UAV orthophotos. Conducted in central Ethiopia, this article aims to assess the feasibility, accuracy, and efficiency of semi-automated boundary extraction for rural land administration, aligning with the core principles of Fit-for-Purpose land governance.

2. Literature Review

Automated feature extraction has been a growing area of interest in the domain of land administration, particularly with the increased availability of high-resolution imagery from Unmanned Aerial Vehicles (UAVs). Researchers have explored various approaches to automate tasks such as road network extraction, building roof delineation, river boundary mapping, and, more recently, cadastral boundary identification. Among these applications, cadastral mapping has gained prominence due to the urgent need to register vast areas of unrecorded land efficiently and accurately (Crommelinck et al., 2016).

Several studies have assessed the effectiveness of UAV imagery in extracting cadastral boundaries. For example, (Fetai et al., 2022) conducted a comparative evaluation of different automated techniques, including edge detection, object-based image analysis (OBIA), and machine learning on UAV-derived imagery. Their results indicated that traditional edge detection methods, while useful, lacked the precision and contextual understanding necessary for reliable cadastral mapping.

More advanced studies, such as (Xia, 2019)Applied Fully Convolutional Networks (FCNs) to UAV datasets and demonstrated a significant improvement over the classical method. The FCNs showed enhanced performance in learning complex spatial features and segmenting boundary lines from dense image data. Similarly, (Crommelinck et al., 2019). performed a comparative analysis of traditional computer vision, machine learning, and deep learning models, finding that CNN-based approaches yielded better accuracy and computational efficiency for feature extraction.

A foundational technique in computer vision, Canny Edge Detection has historically been applied to land mapping but shows limitations when used with satellite or UAV imagery. For instance, (Wassie, 2016) reported only 19.47% overall accuracy in applying Canny edge detection to satellite data, highlighting its inadequacy for precise cadastral tasks.

In response, the field has shifted towards more robust and adaptive models. (Yan et al., 2022), for instance, proposed a single-task semantic segmentation model to delineate farmland boundaries in satellite images, achieving modest Intersection over Union (IoU) scores around 11.31% and 11.72%, underscoring the need for improvement in segmentation accuracy. Building on this, more recent work has adopted Link Net and D-LinkNet architecture models designed for fast, accurate image segmentation. In one such study, Yusen Xie (2016) achieved impressive scores of 0.7398 (IoU) and 0.8505 (F1), validating the use of deep CNNs for this task.

Among the latest models tailored for edge detection are BDCN (Bilateral Divisive Contrast Network) and HED (Holistically Nested Edge Detection). BDCN is known for its multi-scale feature learning and fine-grained edge prediction, while HED leverages hierarchical supervision to learn edge representations at different layers of a CNN. These models have shown strong performance in benchmark computer vision datasets and are now being extended to geospatial applications (Xie & Tu, 2015).

This study adds to this body of knowledge by implementing BDCN and HED for cadastral boundary extraction from UAV orthophotos in a rural Ethiopian context. By comparing these two models, we aim to contribute evidence on their suitability and limitations for semi-automated cadastral mapping tasks.

3. Data and Methodology

3.1. Study Area

The study was conducted in Sodo (Fig. 1), located in the Gurage Zone of Central Ethiopia. The research focused on the area between Firsha kebele and Bue Town, geographically situated at approximately 8°26'5.6"N latitude and 38°36'43.56"E longitude. The total area covered is 487.81 hectares, comprising 59 kebeles, of which 54 are rural, and 5 are urban.
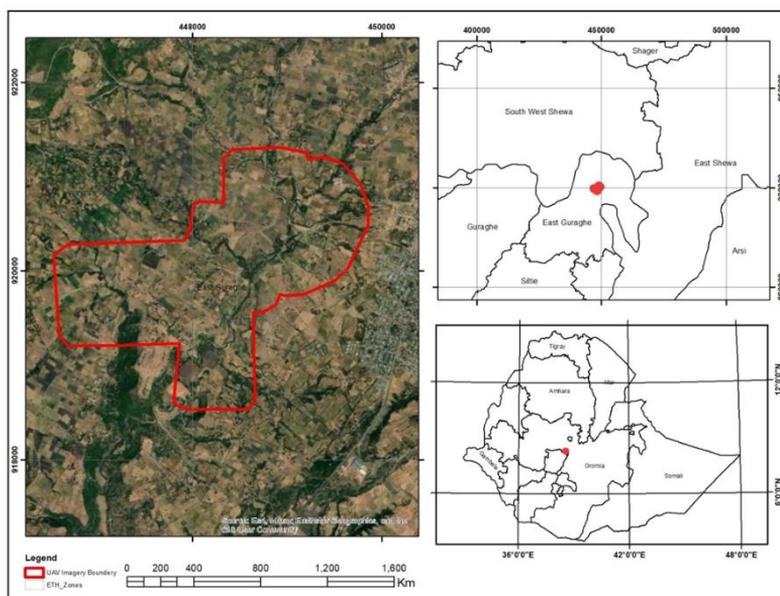

Fig. 1 Location map of the study area

According (Tadesse et al., 2020) The region's economy is dominantly agrarian, with about 85% of households practicing mixed crop-livestock farming. Although agriculture is largely rain-fed, approximately 25% of households situated near the Meki River practice small-scale irrigation for vegetable production to supply local markets (Berhanu, 2021).

53.1% of households own between 0.5 and 2 hectares, and 25% have less than 0.5 hectares. This means land owns b. Around 7.4% of households are landless. The principal crops include maize, wheat, teff, and enset (false banana), with most of the produce consumed domestically and some traded locally. The area was chosen for this study due to its access to high-quality UAV imagery and its small, well-defined

parcel sizes, ideal conditions for testing automated boundary extraction.

3.2. Data Description

A high-resolution UAV orthophoto was used for cadastral boundary extraction. The imagery, captured in January 2023 by the Ethiopian Construction Design and Supervision Works Corporation (ECDSWC), has a spatial resolution of 0.20 m × 0.20 m and three spectral bands (RGB). The images were orthorectified using PIX4D software and georeferenced with 20 Ground Control Points (GCPs) for geometric accuracy.

The timing of the data capture post-harvest ensured clear visibility of parcel boundaries. A vector shapefile representing the cadastral boundaries was created by manually digitizing these boundaries from the UAV orthophoto. This shapefile served as the ground truth reference for evaluating the automated extraction models.

3.3. Proposed Workflow

The methodology involved several systematic steps:

1. Preprocessing and Annotation

The UAV imagery underwent orthorectification and color correction to standardize inputs. A team of trained experts manually annotated visible boundaries to create a labeled dataset, which was used to train and validate the deep learning models.

2. Model Selection and Training

Two deep learning models were selected: BDCN and HED, both built on VGG11 backbone architectures with batch normalization. The selection was based on their proven performance in edge detection tasks and their compatibility with remote sensing imagery.

Each model was trained using the annotated UAV dataset. Key training parameters included:
- Learning Rates: BDCN (2.29e-03), HED (3.02e-04)
- Epochs: Both models trained up to 20 epochs; HED showed early convergence at epoch 14.
- Batch Size: 8
- Data Augmentation: Applied during the testing phase to increase generalizability.

3.4. Evaluation Metrics

Model performance was evaluated using standard computer vision metrics:
- Precision (Equation 1): Measures the proportion of true boundary pixels among all predicted boundary pixels.
- Recall (Equation 2): Measures the proportion of correctly identified boundary pixels among all ground truth boundary pixels.
- F1 Score (Equation 3): Harmonic mean of precision and recall.
- Intersection over Union (IoU): Quantifies the overlap between predicted and actual boundary areas.

$$Recall\ Equation\ (1)\quad Recall = \frac{TP}{(TP+FN)}$$

$$\text{Precision Equation (2) } Precision \ = \ \frac{TP}{(TP+FN)}$$

$$\textit{F1 score Equation (3) } Precision = F1\ score1\ score \ = \ 2 \ * \ \frac{recall * precision}{recall+ precision}$$

Where TP = True Positives, FP = False Positives, and FN = False Negatives.

4. Results

4.1. Comparison of Ground Truth Data with UAV Reading

To proceed to the next step, the first task is to validate the UAV data using ground reference data. This comparison helps to identify the similarities or differences in terms of implementable and scalability and meet the needs of the targeted user groups (Chipofya et al., 2020). The International Organization for Standardization (ISO 19157) guideline states that horizontal accuracy is the measure of how well the position of a dataset aligns with a horizontal datum (2023). To evaluate horizontal accuracy, it is recommended to use root mean square error (RMSE) statistics. In this study for comparison, 20 GCP were taken from the ground using Lica differential GPS and then compared to the corresponding point in the UAV orthophoto.

Table 1. RMSE Statistics

Point_Id	FIXED GCP READING		UAV READING		RMSE	
	Easting	Northing	Easting2	Northing3	Residual_E	Residual_N
GCP 5	448004.263	920191.94	448004.2691	920191.9221	-0.0061	0.0159
GCP 8	449621.373	920418.26	449621.3735	920418.2719	-0.0005	-0.0079
GCP2	446155.047	917348.76	446155.047	917348.761	0	0
GCP 4	447015.36	919691.7	447015.36	919691.7	0	0
GCP 26	446728.353	917576.56	446728.3569	917576.5688	-0.0039	-0.0078
GCP 3	446769.02	918036.72	446769.0205	918036.7361	-0.0005	-0.0151
GCP 27	446923.688	918342.51	446923.6852	918342.5098	0.0028	-0.0008
GCP 29	446850.961	919434.92	446850.9605	919434.921	0.0005	0.001
GCP 1	446775.463	916889.17	446775.4613	916889.1872	0.0017	-0.0202
GCP 13	447524.052	917103.11	447524.142	917103.102	-0.09	0.007
GCP 11	447167.207	916916.28	447167.2117	916916.2749	-0.0047	0.0091
GCP 12	448040.443	917972.65	448040.4373	917972.6521	0.0057	-0.0071
GCP 9	448831.394	919913.22	448831.4061	919913.218	-0.0121	0.002
GCP 23	448401.132	919318.44	448401.1317	919318.4437	0.0003	-0.0057
GCP 32	448185.259	919342.39	448185.2606	919342.3804	-0.0016	0.0076
GCP 31	447794.005	919085.59	447794.0168	919085.6064	-0.0118	-0.0154
GCP 30	447392.607	919372.76	447392.6187	919372.8344	-0.0117	-0.0784
GCP 24	449007.352	921104.59	449007.3466	921104.598	0.0054	-0.005
GCP 10	447785.644	918247.78	447785.5947	918247.8179	0.0493	-0.0389
GCP 28	447090.886	917949.31	447090.8876	917949.3125	-0.0016	-0.0015

RMSE	0.02356	0.02146
Total RMSE		0.031874975

The total residual in X is 0.02 in Y 0.021, and the total RMSE error is 0.03 (Table 1). Based on Geospatial Positioning Accuracy Standards, the result is an RMSE value that is less than about 1/3 to 1/2 of the orthophoto resolution, which is considered good or acceptable accuracy. In this case, the RMSE is only about 1/6 of the 20 cm

resolution, which indicates the geolocation accuracy of the orthophoto is quite high. Additionally, the individual RMSE values for the Residual E and Residual N columns are also very low, all being less than 0.01. This suggests the geometric accuracy of the orthophoto is well within the expected tolerance for a 20 cm resolution product. The low RMSE values demonstrate the UAV-derived data are closely aligned with the fixed GCP reference points.

4.2.   Model Performance Overview

The performance of the two models—**BDCN** and **HED**—was evaluated using the UAV dataset and the metrics described previously.

Table 2. Key training parameters and accuracy results

Model	Backbone	Learning Rate	Epochs Trained	Batch Size	Data Augmentation	Accuracy (%)
BDCN	VGG11 (Batch Normalized)	2.29×10^{-3}	20	8	Testing phase	8.11
HED	VGG11 (Batch Normalized)	3.02×10^{-4}	14 (stopped early)	8	Testing phase	8.28

The BDCN model used a higher learning rate and completed 20 epochs, whereas the HED model converged earlier at 14 epochs with a lower learning rate, suggesting a more stable training process.

4.3.   Accuracy Assessment

In this study, the BDCN and HED models were evaluated to determine their accuracy in detecting land boundaries from UAV imagery. The assessment (Fig.2) utilized standard performance metrics, including precision, recall, F1-score, average precision, and Intersection over Union (IoU).
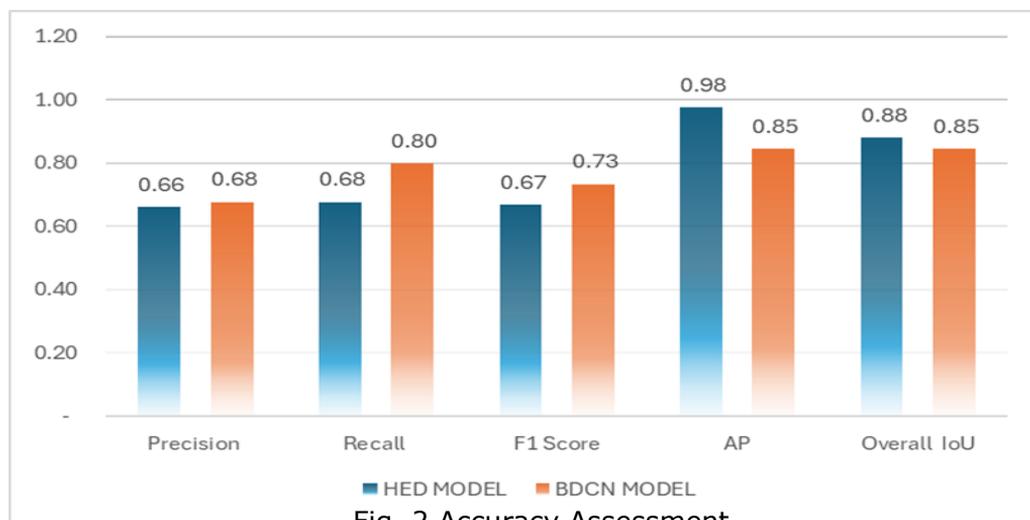

Fig. 2 Accuracy Assessment

When comparing the two architectures, the BDCN model achieved a precision of 0.68 and a high recall of 0.80. This suggests that while the model is highly effective at identifying most actual boundaries within the data (80%), it also carries a slightly higher risk of identifying non-boundary areas. Its F-score of 0.67 reflects this balance. The HED model, by contrast, showed a precision of 0.66 and a recall of 0.68. Although it identifies fewer boundaries overall compared to BDCN, its strength lies in spatial accuracy. This is evidenced by an IoU score of 0.88, which surpasses the 0.85 achieved by BDCN. The higher IoU indicates that the edges predicted by HED align more precisely with the ground truth. Consequently, HED proves to be a reliable choice for rural cadastral mapping where the precision of the primary boundary is more critical than the density of detected lines.

Overall, both models demonstrate that deep learning combined with UAV data can significantly automate land administration. By reducing the time and cost associated with manual surveying, these tools directly support the Fit-For-Purpose (FFP) land administration approach, making land services more accessible and affordable.

4.4. Qualitative Analysis

To further assess the model's effectiveness, it is recommended to evaluate its performance on a separate test set or real-world data to ensure accurate predictions in practical scenarios. The qualitative results align with the quantitative metrics. As illustrated in Figure 4, the BDCN model (yellow) captures significantly more detail and complex structures. In contrast, the HED model shown in Figure 3 (red) produces sparser outputs but remains highly accurate along the main boundary lines.
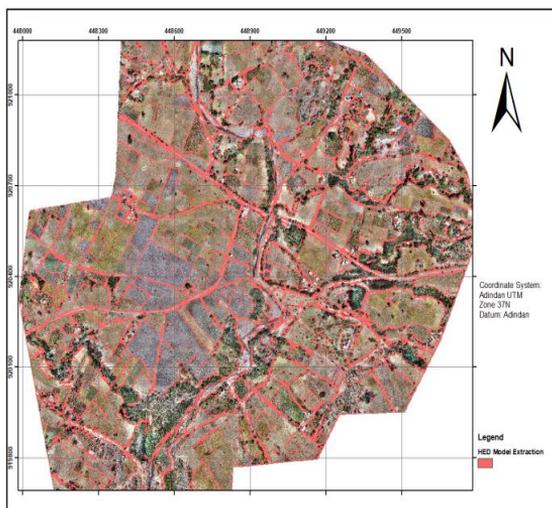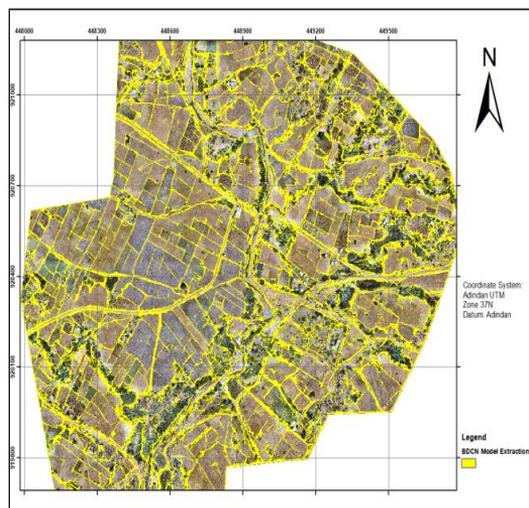

Fig. 3. HED model extraction

Fig. 4 BDCN model extraction

4.5. Strengths and Limitations

Strengths:
- Both models use end-to-end deep learning, removing the need for manual feature engineering.
- High accuracy in extracting visible cadastral boundaries enables semi-automated workflows.

- Potential to reduce time and costs associated with traditional cadastral surveying methods.

Limitations:
- Model performance is sensitive to UAV image quality, including resolution and lighting.
- Generalization to different geographic areas or datasets may require retraining or transfer learning.
- Deep learning models demand significant computational resources (e.g., GPUs) for training and inference.

4.6. Implications for Fit-for-Purpose Land Administration

The successful application of BDCN and HED in this study supports the **Fit-for-Purpose (FFP)** approach by providing a scalable, cost-effective means of cadastral mapping in rural contexts. The ability to semi-automate boundary extraction aligns with FFP's goals of affordability and adaptability, offering governments and land agencies practical tools for accelerating land registration and tenure security (Enemark & Mclaren, 2017; Metaferia et al., 2023).

5. Conclusion

This study demonstrated the promising potential of deep learning models, specifically the Bilateral Divisive Contrast Network (BDCN) and Holistically-Nested Edge Detection (HED), for automated cadastral boundary extraction from high-resolution UAV imagery. Both models effectively extracted boundaries in vector polygon format, providing outputs that can be integrated into rural cadastral mapping systems with minimal post-processing and field validation.

The BDCN model showed higher precision and recall, effectively capturing detailed edges, particularly in complex boundary areas. Meanwhile, the HED model delivered slightly better overall alignment with ground truth boundaries, as reflected in higher IoU and F1 scores. This complementary performance suggests that either model could be adapted for semi-automated cadastral workflows depending on the specific mapping context and priorities.

The approach outlined here aligns closely with the Fit-for-Purpose (FFP) land administration paradigm by offering a method that is cost-effective, scalable, and adaptable to local conditions. By integrating UAV imagery with advanced deep learning techniques, land administration agencies can overcome many traditional challenges, such as high costs and slow turnaround times associated with manual cadastral surveying (Bennett et al., 2021; Enemark & Mclaren, 2017).

Future work should explore improvements through transfer learning, hyperparameter optimization, and expansion to larger and more diverse geographic areas. Additionally, integrating other data sources, such as LiDAR or satellite imagery, could further enhance boundary extraction accuracy and robustness. Overall, this study contributes to the growing field of geospatial AI applications and demonstrates a clear pathway for modernizing cadastral mapping to support land tenure security and sustainable development goals globally.

REFERENCE

1. Bennett, R. M., & Alemie, B. K. (2016). Fit-for-purpose land administration: Lessons from urban and rural ethiopia. Survey Review, 48(346), 11–20. https://doi.org/10.1080/00396265.2015.1097584

2. Bennett, R. M., Koeva, M., & Asiama, K. (2021). Review of remote sensing for land administration: Origins, debates, and selected cases. Remote Sensing, 13(21). https://doi.org/10.3390/rs13214198

3. Berhanu, H. (2021). Landscape action plan for sodo and sodo guragie woredas, snnp regional state. www.bagerconsult.com,

4. Carrio, A., Sampedro, C., Rodriguez-Ramos, A., & Campoy, P. (2017). A review of deep learning methods and applications for unmanned aerial vehicles. In Journal of Sensors (Vol. 2017). Hindawi Limited. https://doi.org/10.1155/2017/3296874

5. Crommelinck, S., Bennett, R., Gerke, M., Nex, F., Yang, M. Y., & Vosselman, G. (2016). Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping. In Remote Sensing (Vol. 8, Issue 8). MDPI AG. https://doi.org/10.3390/rs8080689

6. Crommelinck, S., Bennett, R., Gerke, M., Yang, M. Y., & Vosselman, G. (2017). Contour detection for UAV-based cadastral mapping. Remote Sensing, 9(2). https://doi.org/10.3390/rs9020171

7. Crommelinck, S., Koeva, M., Yang, M. Y., & Vosselman, G. (2019). Application of deep learning for delineation of visible cadastral boundaries from remote sensing imagery. Remote Sensing, 11(21). https://doi.org/10.3390/rs11212505

8. Enemark, S. ;, & Mclaren, R. (2017). Aalborg Universitet Fit-For-Purpose Land Administration Developing Country Specific Strategies for Implementation.

9. Fetai, B., Grigillo, D., & Lisec, A. (2022). Revising Cadastral Data on Land Boundaries Using Deep Learning in Image-Based Mapping. ISPRS International Journal of Geo-Information, 11(5). https://doi.org/10.3390/ijgi11050298

10. Metaferia, M. T., Bennett, R. M., Alemie, B. K., & Koeva, M. (2023a). Fit-for-Purpose Land Administration and the Framework for Effective Land Administration: Synthesis of Contemporary Experiences. In Land (Vol. 12, Issue 1). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/land12010058

11. Metaferia, M. T., Bennett, R. M., Alemie, B. K., & Koeva, M. (2023b). Furthering Automatic Feature Extraction for Fit-for-Purpose Cadastral Updating: Cases from Peri-Urban Addis Ababa, Ethiopia. Remote Sensing, 15(17). https://doi.org/10.3390/rs15174155

12. Tadesse, H., Melese, M., & Tamene, N. (2020). Determinants of Rural Women Participation in Agriculture Activities in Gurage Zone, Ethiopia. In International Journal of Research in Agricultural Sciences (Vol. 7, Issue 2).

13. Tuladhar Arbind Man. (2004). Parcel-based Geo-Information System: Concepts and Guidelines Arbind Man Tuladhar.

14. United Nations. (2015). 70/1. Transforming our world: the 2030 Agenda for Sustainable Development Preamble.

15. Wassie, Y. A. (2016). Toward automatic cadastral boundary mapping from satellite imagery.

16. Xia, X. (2019). Extracting cadastral boundaries from uav images using fully convolutional networks.

17. Xie, S., & Tu, Z. (2015). Holistically-Nested Edge Detection. http://arxiv.org/abs/1504.06375

18. Yan, Z., Ma, L., He, W., Zhou, L., Lu, H., Liu, G., & Huang, G. (2022). Comparing Object-Based and Pixel-Based Methods for Local Climate Zones Mapping with Multi-Source Data. Remote Sensing, 14(15). https://doi.org/10.3390/rs14153744

Explainable AI Approaches for Detecting and Mitigating Phishing Attacks: A Review

Nipuna Sankalpa Thalpage [1,2][0009-0001-3374-1927], Eranga Jayarathne [3][0009-0004-8915-4467]

[1]Cardiff Metropolitan University, Wales, United Kingdom;
[2]Institute of Cited Scientists, Agia Napa, Cyprus
[3]University of Peradeniya, Sri Lanka

Abstract. Phishing remains one of the most pervasive and sophisticated cybersecurity threats, increasingly leveraging social engineering, AI-driven content generation, and multi-vector delivery methods. While machine learning (ML) and deep learning (DL) models have significantly advanced phishing detection capabilities, their "black-box" nature often limits transparency, trust, and practical adoption in real-world security environments. Explainable Artificial Intelligence (XAI) offers a solution by providing interpretable insights into model decisions, enabling analysts and stakeholders to understand, validate, and act upon automated classifications. This semi-systematic review examines contemporary XAI techniques applied to phishing detection, focusing on studies published between 2017 and 2025. Searches conducted across Scopus, IEEE Xplore, and Google Scholar yielded peer-reviewed literature integrating explainability into ML/DL-based phishing detection. The selected studies were synthesized to identify the types of models used, the XAI methods employed, and their contributions to interpretability, operational value, and human–AI collaboration.

Findings show that feature attribution methods such as SHAP, LIME, and Integrated Gradients are the most widely adopted, offering both global and local explanations for text-based and URL-based phishing detection. Attention mechanisms and visualization techniques further enhance transparency in deep learning models, while interpretable models—such as decision trees and logistic regression, remain valuable for contexts requiring high clarity. However, gaps persist in real-world validation, dataset diversity, standard metrics for evaluating explanations, and deployment feasibility.

Overall, XAI strengthens phishing mitigation by improving user trust, supporting analyst decision-making, and enabling more accountable AI-driven security systems. The review highlights the need for scalable, human-centred, and adversarially robust XAI approaches to support the next generation of phishing detection frameworks.

Keywords: Phishing, Explainable AI, Cybersecurity.

1. Introduction

Phishing has become a significant global cybersecurity threat, affecting individuals, businesses, and governments across all sectors. As a form of social engineering, phishing attacks deceive users into disclosing sensitive information or performing harmful actions by impersonating trustworthy entities. With the rapid expansion of digital services, cloud platforms, mobile communication, and remote work environments, phishing incidents have increased in both frequency and

sophistication. Attackers now employ targeted spear-phishing, business email compromise (BEC), and multi-vector delivery methods that exploit human behavior rather than system vulnerabilities, making phishing one of the most challenging cybersecurity issues worldwide.

Despite continuous advancements in cybersecurity technologies, traditional phishing detection methods face several limitations. Signature-based filters struggle to identify newly crafted phishing emails, as attackers frequently modify URLs, payloads, and content to evade detection. Rule-based systems depend on predefined patterns and are often ineffective against zero-day attacks or dynamically generated phishing pages. Moreover, conventional spam filters can produce high false-positive and false-negative rates, reducing reliability and causing user fatigue. These limitations create a critical gap in existing defense mechanisms, allowing sophisticated phishing campaigns to bypass conventional security tools and compromise users [1].

In response to these challenges, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as promising approaches for enhancing phishing detection. ML-based systems can automatically learn discriminative features from large datasets, enabling more accurate identification of malicious emails, URLs, and websites. AI-driven models such as deep learning, natural language processing (NLP), and anomaly detection techniques provide adaptive and scalable solutions capable of identifying previously unseen threats [2]. These intelligent systems can analyze behavioural, structural, and linguistic patterns, offering a more robust and proactive defense against evolving phishing tactics. As a result, AI/ML-based phishing detection has become a key focus in contemporary cybersecurity research.

1.1 Need for Explainability

The increasing reliance on machine learning for phishing detection has introduced new challenges related to the interpretability of these systems. Many state-of-the-art ML and deep learning models operate as "black boxes," providing high accuracy but offering little insight into how decisions are made. This opacity becomes problematic in cybersecurity applications, where understanding the rationale behind a model's classification is crucial. Without clear explanations, security analysts may struggle to validate alerts, investigate incidents, or identify model weaknesses, ultimately limiting the practical adoption and trustworthiness of such systems [3].

Transparency, trust, and accountability are therefore essential when deploying automated threat-detection tools. In high-risk environments, such as financial institutions, critical infrastructure, and governmental systems, stakeholders require assurance that AI-driven decisions are consistent, unbiased, and aligned with organizational policies. Explainable systems enable users to understand why an email or URL is classified as phishing, improving confidence in automated decisions and supporting compliance with regulatory standards. Additionally, explainability reduces the likelihood of misclassifications, facilitates human oversight, and enhances the collaborative interaction between analysts and intelligent detection systems.

The importance of explainable artificial intelligence extends beyond cybersecurity into broader digital transformation initiatives. Recent review studies highlight that integrating machine learning with explainable AI significantly enhances organizational trust, accountability, and decision-making across data-driven systems. These findings emphasize that explainability is a key enabler for the responsible adoption of AI technologies, reinforcing the relevance of XAI in high-risk domains such as phishing detection, where transparency and human oversight are critical [4].

Explainable Artificial Intelligence (XAI) plays a vital role in addressing these concerns within cybersecurity. The need for explainable models is further reinforced

by calls for a rigorous scientific foundation for interpretability, particularly in high-stakes decision-making domains such as cybersecurity [5].

XAI techniques aim to make ML model behavior more interpretable by identifying key features, patterns, or reasoning processes that influence predictions. In the context of phishing detection, XAI can highlight suspicious linguistic cues, structural anomalies in URLs, or abnormal sender behaviors that contribute to an alert. These insights help analysts refine detection strategies, understand emerging attack vectors, and improve model robustness against adversarial manipulation. Consequently, XAI has become a critical component of modern cybersecurity frameworks, enabling more secure, transparent, and reliable AI-driven defense mechanisms.

1.2 Purpose of the Review

The purpose of this review is to examine the role of Explainable Artificial Intelligence (XAI) in enhancing phishing detection systems. As phishing attacks continue to evolve in complexity, many organizations have turned to machine learning and deep learning models to strengthen their defensive capabilities. However, the opaque nature of these black-box models raises significant concerns regarding trust, interpretability, and operational reliability. By evaluating the current landscape of XAI techniques and their applications in phishing detection, this review aims to identify how explainability can address these limitations and support more informed, transparent decision-making in cybersecurity environments.

This review contributes to the body of knowledge in several ways. First, it synthesizes existing research on AI/ML-based phishing detection and highlights the specific challenges associated with model interpretability. Second, it provides a structured analysis of leading XAI methods, such as feature attribution, model-agnostic explanations, and visualization frameworks—and examines their effectiveness in cybersecurity contexts. Third, the review identifies gaps in current literature, including issues related to usability, scalability, adversarial robustness, and real-world deployment. Through this comprehensive evaluation, the review offers insights that can guide future research, support the development of more transparent detection systems, and promote the responsible integration of AI within cybersecurity practices.

In contrast to existing reviews that primarily focus on detection accuracy or algorithmic performance, this study enriches the subject area by foregrounding explainability as a central analytical lens. Specifically, it advances current literature by comparatively synthesising XAI techniques in phishing detection, examining their contribution to transparency, human–AI collaboration, and real-world deployment feasibility. By explicitly linking explainability methods to operational decision-making and trust, this review extends beyond technical surveys and provides interdisciplinary value relevant to cybersecurity, human-centered AI, and digital governance.

1.3 Research Questions

This review is guided by the following research questions, which aim to explore the application, effectiveness, and limitations of Explainable Artificial Intelligence (XAI) within phishing detection systems:

RQ1: What XAI methods have been applied in phishing detection?
This question seeks to identify and categorize the range of XAI techniques used in existing studies, including model-specific and model-agnostic approaches. It explores how various explainability methods, such as LIME, SHAP, attention mechanisms, and rule-based explanations, have been integrated into phishing detection models.

RQ2: How do XAI techniques enhance transparency and decision-making in phishing mitigation?

This question examines the role of explainability in improving analyst understanding, trust, and operational effectiveness. It investigates how XAI contributes to clearer interpretations of model outputs, better incident response, and more informed cybersecurity decision-making.

RQ3: What limitations and research gaps exist in current literature?

This question identifies shortcomings in present research, such as challenges in evaluating explanation quality, scalability concerns, adversarial vulnerabilities, dataset biases, and limited real-world deployment. It highlights areas where further investigation is needed to advance the development of robust, transparent AI-based phishing detection systems.

2. Methodology

This study employs a semi-systematic literature review to examine how Explainable Artificial Intelligence (XAI) techniques are applied in phishing detection and how interpretability contributes to cybersecurity decision-making. This review method was selected because it allows a structured yet flexible examination of emerging interdisciplinary research, particularly where technological, behavioural, and security aspects intersect.

A targeted search was conducted across three major academic databases: Scopus, IEEE Xplore, and Google Scholar. The search strategy used combinations of key terms such as: *"Explainable AI," "XAI," "phishing detection," "email phishing," "URL phishing," "interpretable machine learning,"* and *"explainability in cybersecurity."*

To ensure relevance and quality, the following inclusion criteria were applied:
• Peer-reviewed journal or conference publications
• Published between 2017 and 2025
• Focused on machine learning–based phishing detection with an XAI component
• Written in English

Studies were excluded if they: (1) Did not include any explainability method; (2) Focused solely on phishing awareness or user training, were non-peer-reviewed preprints, abstracts, or posters; (3) Used traditional rule-based detection without ML/XAI.

The screening procedure consisted of an initial title and abstract review to identify relevant papers, followed by a detailed full-text evaluation. To reduce subjectivity and strengthen reliability, all authors participated in the screening and data extraction process. This multidisciplinary collaboration ensured balanced interpretation by combining expertise from cybersecurity, AI/ML, and user-centered perspectives.

The final set of selected papers was examined using thematic analysis. Each study was reviewed to identify:
• The type of phishing detection model used (ML/DL),
• The XAI technique applied (e.g., SHAP, LIME, attention mechanisms),
• The nature of explanations provided (local/global),
• Key findings, strengths, and limitations related to interpretability in phishing detection.

This review process aligns with established practices for semi-systematic and mapping reviews, offering both methodological transparency and flexibility for exploring rapidly developing research areas [6].

Compared to purely narrative reviews, the semi-systematic approach adopted in this study provides improved methodological rigor by combining structured

database searches, explicit inclusion and exclusion criteria, and thematic synthesis. This methodology enables reproducibility while retaining flexibility to analyse emerging interdisciplinary research that spans explainable AI, cybersecurity, and human-centered design. By involving multiple authors in screening and synthesis, the approach also reduces selection bias and strengthens the reliability of the findings.

3. Background and Theoretical Foundation

3.1 Overview of Phishing Attacks

Phishing is a form of cyberattack in which adversaries deceive users into revealing confidential information or performing malicious actions by impersonating legitimate entities. Traditionally delivered through email, phishing has expanded into multiple channels, including fraudulent websites, SMS-based attacks (smishing), and voice-based scams (vishing). Spear-phishing—highly targeted attacks tailored to specific individuals or organizations, has become increasingly prevalent due to the availability of personal data on social platforms and public sources [7]. Large-scale empirical studies have shown that phishing campaigns continuously evolve in linguistic structure, delivery mechanisms, and visual deception strategies, complicating static detection approaches [8].

Modern phishing trends reflect a shift toward more sophisticated and evasive techniques. Attackers now use AI-generated content, dynamic phishing websites, URL obfuscation, and advanced social engineering strategies to bypass traditional security filters. Additionally, multi-stage and multi-vector phishing campaigns combine email, malicious links, and spoofed authentication pages to increase success rates. The rise of business email compromise (BEC), credential-harvesting kits, and adversarial tactics has made phishing a persistent and evolving cybersecurity threat [3].

3.2 ML Models in Phishing Detection

Machine learning (ML) has become a central mechanism for improving phishing detection by analyzing complex patterns in emails, URLs, and website structures. Feature-based models such as decision trees, random forests, support vector machines, and logistic regression rely on hand-crafted features, for example, URL length, domain age, sender attributes, or HTML structure, to differentiate phishing attempts from legitimate communication. These models often provide fast and explainable decisions but may struggle to adapt to rapidly evolving attack vectors [1]. Traditional supervised machine learning models using lexical and host-based features remain effective baselines for phishing detection and are frequently compared against deep learning approaches [9].

NLP-based approaches utilize natural language processing techniques to analyze the textual content of emails or messages. Models can detect linguistic anomalies, sentiment patterns, and semantic relationships indicative of phishing. Techniques such as TF-IDF, word embeddings, and transformer architectures have improved the ability to capture subtle cues in phishing messages [10].

Deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), LSTMs, and transformer-based models, have further advanced detection capabilities by automatically learning complex, high-dimensional representations. These models excel in classifying phishing websites, analyzing visual layouts, and processing unstructured textual data. However, their high accuracy often comes at the cost of interpretability, making them challenging to trust and validate in real-world settings [3].

Convolutional neural network architectures have been successfully applied to phishing website detection by learning discriminative visual and structural patterns directly from web content [11].

3.3 Explainable AI in Cybersecurity

Explainable Artificial Intelligence (XAI) aims to address the opacity of complex ML models by providing insights into how and why decisions are made. In cybersecurity, XAI is essential for supporting human analysts, ensuring transparency, and enabling reliable deployment of ML-driven detection systems.

One core distinction in XAI is between global and local explanations. Global explanations describe the overall behavior of a model, how features generally influence predictions, while local explanations focus on individual decisions, clarifying why a specific email or URL was flagged as phishing. Both perspectives are valuable for different security tasks, including model auditing and incident investigation [2].

Another key distinction lies between post-hoc and intrinsic interpretability. Post-hoc methods, such as LIME, SHAP, counterfactual explanations, and saliency maps, generate explanations after model training without altering the underlying architecture. In contrast, intrinsic interpretability involves building models, such as decision trees, rule-based models, or attention-enabled architectures, with explanations inherently embedded. While intrinsic models tend to be more transparent, post-hoc techniques allow high-performing deep learning models to be used without sacrificing interpretability [12, p. 2019].

XAI plays a critical role in increasing trust among users and security analysts. By revealing the reasoning behind detection outcomes, XAI improves confidence in automated decisions, supports regulatory compliance, and facilitates more effective threat investigation. It also enhances resilience against adversarial manipulation by allowing analysts to identify model weaknesses and potential exploitation paths. As phishing threats continue to grow, the integration of XAI into detection frameworks becomes essential for ensuring reliable, transparent, and actionable cybersecurity defenses [3].

4. XAI Techniques in Phishing Detection — Review Findings

The findings presented in Sections 4 and 5 represent the results of this semi-systematic review. Rather than reporting experimental outcomes, these sections synthesise evidence from existing studies to identify dominant XAI approaches, their interpretability characteristics, and their implications for phishing mitigation.

4.1 Feature Attribution Methods

Feature attribution methods were the most common XAI techniques found in phishing detection literature. These methods help identify which features—such as URL structure, token positions, domain characteristics, or suspicious keywords—contributed most to the model's prediction.

SHAP (Shapley Additive Explanations):

SHAP is widely used due to its strong theoretical foundation and ability to provide consistent global and local explanations. Studies showed that SHAP effectively identified high-impact phishing indicators such as URL length, abnormal characters, and deceptive lexical cues [13].

LIME (Local Interpretable Model-Agnostic Explanations):

LIME was applied to explain individual phishing samples, highlighting influential tokens or words in email bodies and URL segments. Although sometimes less stable than SHAP, it remained popular for its simplicity and intuitive visual outputs [14].

Permutation Feature Importance:

Permutation-based importance methods were used mainly in traditional ML models, such as Random Forests, to evaluate how shuffling a feature affects prediction performance. Features like "presence of IP in URL" and "domain age" commonly ranked high in phishing detection tasks.

Integrated Gradients:

Deep learning–based phishing detectors (e.g., LSTMs and transformers) applied integrated gradients to reveal token-level contributions across email content or URL structures. This method provided fine-grained explanations for complex neural models [15].

4.2 Interpretable Models

Some studies adopted models that are transparent by design, eliminating the need for post-hoc explainability tools.

Decision Trees:
Decision trees remained useful due to their rule-based, human-readable nature. They generated explicit decision paths such as "URL contains multiple subdomains → classify as phishing."

Rule-Based Models:
Rule-based systems offered deterministic explanations using if–then constructs, supporting environments where compliance and interpretability are essential.

Logistic Regression:

Logistic regression allowed direct interpretation of feature coefficients, enabling analysts to understand which features increased or decreased the likelihood of phishing.

While these models achieved strong interpretability, their performance often lagged behind deep learning models on sophisticated phishing attacks.

4.3 Visualization and Model Transparency Tools

Visualization-based explanation methods, including relevance heatmaps and saliency mapping, have been widely adopted to interpret deep learning decisions by highlighting influential input regions [16].

Heatmaps:
Heatmaps were used to emphasize influential regions in emails or webpage screenshots, highlighting suspicious forms, links, or layout structures. Fig. 1 illustrates a heatmap used against a phishing email.

Fig.1. Illustrative Heatmap Highlighting Suspicious Regions in a Phishing Email

Visualization-based XAI techniques provided clear insight into how DL models processed and classified phishing content.

Attention Mechanisms:

Transformer-based phishing detectors visualized attention weights to show which tokens or URL parts influenced classification outcomes most. Attention-based XAI is particularly effective for text-rich phishing emails [17].

Saliency Maps:

Saliency maps were employed in visual phishing detection to highlight critical UI elements, logos, login fields, and fake certificates that the model found suspicious.

These tools improved clarity for cybersecurity analysts, supporting human–AI collaboration in investigation workflows.

4.4. Hybrid and Emerging XAI Methods

Emerging studies explored hybrid approaches that combine symbolic reasoning, multiple explainers, or alternative forms of interpretability.

Neural-Symbolic XAI:
These approaches integrated rule-based logic with neural network architectures, providing a balance between accuracy and interpretability.

Ensemble Explainability:
Some works combined SHAP (global) with LIME (local) or added attention maps to produce multi-layered explanations, offering richer interpretive insights.

Counterfactual Explanations:
Counterfactual methods showed how small changes, such as modifying a suspicious keyword or altering a URL parameter, could shift a classification from phishing to legitimate. This method helped reveal model sensitivity and decision boundaries.

Although promising, hybrid methods require higher computational resources and remain less common in phishing-specific literature.

4.5. Comparative Synthesis

Table 1 presents a comparative synthesis of dominant XAI techniques applied in phishing detection, highlighting trade-offs between interpretability, performance, and real-world deployment feasibility.

Table 1. Comparative synthesis of XAI techniques applied in phishing detection

XAI Technique	Model Type	Explanation Level	Strengths	Limitations	Deployment Suitability
SHAP	ML / DL	Global & Local	Stable, theoretically grounded	Computational overhead	Medium
LIME	Model-agnostic	Local	Intuitive, flexible	Explanation instability	Medium
Attention Mechanisms	DL	Local	Context-aware explanations	Not causally grounded	High
Decision Trees	ML	Global	Fully interpretable	Lower detection accuracy	High
Counterfactuals	ML / DL	Local	Actionable explanations	Hard to generate	Low–Medium

A cross-study comparison reveals several important insights:

Best Performing XAI Methods

- **SHAP** provided the most reliable and widely applicable explanations across both ML and DL models.
- **Attention mechanisms** excelled for deep learning models processing text-based phishing content.

Most Interpretable Approaches

- **Decision trees** and **rule-based models** were easiest for analysts to interpret.
- **LIME** remained valuable for quick, instance-level explanations.
  Key Trade-offs
- **Interpretable models offer transparency** but may sacrifice detection accuracy.
- **Deep models achieve higher accuracy** but rely heavily on post-hoc XAI.
- **Hybrid methods provide richer explanations** but at higher computational cost.

Overall, no single XAI technique is universally optimal. The literature suggests that combining feature attribution, visual interpretability, and interpretable model

design can provide the most balanced approach for real-world cybersecurity applications.

5. Impact of XAI in Phishing Mitigation – Review Findings

Explainable Artificial Intelligence (XAI) plays a critical role in enhancing the security, usability, and operational effectiveness of phishing detection systems. As phishing attacks continue to evolve in complexity, XAI enables both technical and non-technical stakeholders to understand the reasoning behind automated decisions, supporting better trust, coordination, and response mechanisms.

Further, Explainable AI methods can be broadly categorized into intrinsic and post-hoc approaches, offering different trade-offs between transparency, fidelity, and usability [18]. This chapter summarizes the impact of XAI on transparency, human–AI collaboration, and real-world deployment considerations.

5.1 Improving Transparency and Trust

One of the most significant contributions of XAI in phishing mitigation is its ability to improve transparency in automated decision-making. Explanations help analysts understand *why* a particular email or URL was classified as phishing, making model behaviour more predictable and interpretable. Human-centered explainability research indicates that explanations aligned with human reasoning significantly enhance trust, confidence, and decision quality in AI-assisted systems [19].

Influence on Analyst Decision-Making

XAI-generated explanations, such as SHAP plots, attention maps, or feature contribution scores, allow cybersecurity analysts to verify whether the model is focusing on meaningful indicators, such as deceptive keywords, suspicious sender domains, or structural anomalies in URLs. This validation helps analysts make confident decisions, reduce false alarms, and identify potential blind spots in the model's reasoning. Studies indicate that transparent ML models improve analysts' ability to detect misclassifications and refine incident response workflows [20].

User Trust Improvements

Trust is essential when deploying AI systems in high-risk cybersecurity environments. When end-users and analysts can clearly see why a system flagged content as malicious, they are more likely to accept and rely on model outputs. Evidence shows that interpretability increases perceived reliability and promotes responsible adoption of AI in security operations [21].

5.2 Human–AI Collaboration

XAI strengthens the interaction between human analysts and automated phishing detection systems, allowing both to work more effectively together.

Helping Cybersecurity Teams

By providing clear, interpretable insights into phishing indicators, XAI reduces cognitive load for analysts and accelerates triage processes. For example, attention-based explanations can highlight critical tokens or patterns that require immediate review, helping security teams prioritize alerts. This collaborative advantage enhances

overall detection performance and reduces manual analysis time.

Supporting End-User Training

XAI explanations can also be used in phishing awareness and training programs. Highlighting which elements of an email prompted its classification,such as unusual URL obfuscation, manipulated brand names, or mismatched sender information, helps educate users on real-world phishing characteristics. Prior work suggests that visual, example-based explanations improve user comprehension and retention [14]**.**

XAI therefore contributes not only to technical defense but also to behavioural and educational aspects of phishing mitigation.

5.3 Practical Deployment Considerations

While XAI provides substantial benefits, deploying explainable phishing detection systems in real-world environments introduces several operational challenges.
Real-Time Processing Constraints:
a) Generating explanations—especially for deep neural networks—can be computationally demanding. Methods like SHAP or Integrated Gradients may introduce latency that is incompatible with real-time email filtering or URL scanning. Lightweight explanation methods or approximation-based techniques are often necessary to maintain system responsiveness [15].
b) Explainability vs. Model Performance

5.4 A recurring challenge highlighted in the reviewed studies is the trade-off between interpretability and accuracy.

- Interpretable models (e.g., decision trees, logistic regression) are easier to understand but may fail to capture the complexity of sophisticated phishing attacks.
- High-performing deep learning models achieve superior detection rates but require post-hoc XAI to remain transparent.

Balancing these factors is crucial for building reliable, deployable phishing detection systems that meet both security and operational requirements.

6. Discussion

6.1 Discussion in Relation to RQ1: XAI Methods Applied in Phishing Detection

In relation to RQ1, the reviewed literature demonstrates that feature attribution–based XAI methods are the most widely adopted approaches in phishing detection systems. Techniques such as SHAP, LIME, Integrated Gradients, and attention mechanisms dominate existing research due to their flexibility and compatibility with both traditional machine learning and deep learning models. Among these, SHAP emerged as the most consistently applied method, offering both global and local explanations grounded in solid theoretical principles. Its ability to quantify feature contributions has been particularly effective in identifying phishing indicators such as URL obfuscation patterns, suspicious lexical tokens, and anomalous sender attributes.

Attention mechanisms were predominantly used in deep learning architectures, especially transformer-based models, to provide token-level explanations for text-based phishing detection. These mechanisms enable visualization of which parts of an email or URL the model prioritizes, thereby enhancing interpretability without significantly compromising detection performance. In contrast, interpretable models such as decision trees and logistic regression offer intrinsic transparency through rule-based reasoning and coefficient analysis but generally underperform when faced with sophisticated, evolving phishing attacks.

Overall, the findings indicate a clear research preference for post-hoc, model-agnostic XAI techniques that preserve high detection accuracy while providing meaningful explanations. However, the dominance of post-hoc methods also reflects the ongoing challenge of achieving both interpretability and performance within a single unified model architecture.

6.2 Discussion in Relation to RQ2: Enhancing Transparency and Decision-Making in Phishing Mitigation

Addressing RQ2, the review confirms that XAI techniques play a critical role in enhancing transparency, trust, and decision-making in phishing mitigation. Explainability mechanisms—such as feature importance plots, attention visualizations, saliency maps, and heatmaps—enable cybersecurity analysts to understand why a particular email, URL, or webpage is classified as phishing. This transparency allows analysts to validate model outputs, identify false positives, and assess whether the model's reasoning aligns with known phishing characteristics.

The reviewed studies highlight that XAI significantly improves human–AI collaboration by reducing cognitive load and supporting faster incident triage. Rather than treating AI systems as opaque decision-makers, analysts can interact with explanations to refine detection strategies and prioritize alerts more effectively. This is particularly important in operational security environments where analysts must justify actions, comply with regulatory requirements, and respond to threats in real time.

Beyond analyst support, XAI also contributes to user trust and awareness. Several studies suggest that explanation-driven insights can be integrated into phishing awareness and training programs, helping end-users recognize deceptive patterns in real-world attacks. By illustrating how specific elements—such as manipulated brand names, mismatched URLs, or urgent language—trigger detection, XAI enhances both technical defenses and human resilience against phishing.

Collectively, these findings demonstrate that explainability is not merely an auxiliary feature but a foundational requirement for responsible and effective AI-driven phishing mitigation.

6.3 Discussion in Relation to RQ3: Limitations and Research Gaps

In response to RQ3, the review identifies several persistent limitations and research gaps that constrain the practical impact of XAI-based phishing detection systems. The absence of standardized evaluation methodologies for explainability remains a critical challenge, particularly in high-risk domains such as cybersecurity [22] .A major concern is the overreliance on public benchmark datasets, such as PhishTank and static email corpora, which often fail to capture the diversity, context, and rapid evolution of real-world phishing attacks. This raises questions about the generalizability and robustness of both detection models and their explanations.

Another critical limitation is the lack of human-centered evaluation. Most studies assess XAI effectiveness using technical metrics or qualitative examples, with minimal empirical evaluation of how explanations are perceived and used by analysts

or end-users. Without systematic user studies, it remains unclear whether explanations genuinely improve understanding, trust, or decision quality in operational environments.

The absence of standardized metrics for explanation quality further complicates comparative evaluation across studies. Explanation clarity, usefulness, stability, and trustworthiness are often measured inconsistently or not at all, limiting the ability to draw strong conclusions about the relative effectiveness of different XAI approaches.

From a deployment perspective, computational overhead and latency remain significant challenges. Techniques such as SHAP and Integrated Gradients, while informative, may introduce delays incompatible with real-time email filtering or large-scale enterprise systems. Moreover, few studies address the adversarial robustness of XAI, despite the risk that attackers could manipulate inputs to exploit or mislead explanation mechanisms.

These limitations highlight the need for more holistic, deployment-aware, and user-focused research to advance explainable phishing detection beyond experimental settings.

6.4 Implications for Future Explainable Phishing Detection Research

The findings of this review suggest several important implications for future research. First, there is a growing need to develop intrinsically interpretable models that integrate transparency directly into the detection process while maintaining competitive accuracy. Second, human-centered evaluation frameworks should be prioritized to assess how explanations support real-world decision-making. Third, research should focus on lightweight and scalable XAI techniques suitable for real-time deployment. Finally, greater attention must be given to adversarial threats against explainable systems, ensuring that explanations enhance security rather than introduce new vulnerabilities.

7. Conclusion

This review demonstrates that Explainable Artificial Intelligence has substantially enriched phishing detection research by transforming opaque classification systems into transparent, accountable, and analyst-supportive tools. Evidence from the reviewed studies shows that XAI improves interpretability, supports informed decision-making, and strengthens trust in AI-driven cybersecurity systems. These findings directly address the central research objective by confirming that explainability is not a peripheral enhancement, but a foundational requirement for effective and responsible phishing mitigation.

The review examined the role of Explainable Artificial Intelligence (XAI) in phishing detection, highlighting how explainability enhances transparency, trust, and operational effectiveness in cybersecurity environments. Across the reviewed studies, it is evident that XAI contributes significantly to the interpretability of machine learning and deep learning models, offering insights into model behaviour that are essential for both analysts and end-users. Feature attribution methods, interpretable models, visualization tools, and emerging hybrid techniques each play a distinct role in supporting clearer, more accountable decision-making.

The findings demonstrate that while deep learning approaches continue to deliver strong predictive performance, their complexity requires robust explainability mechanisms to ensure they can be used responsibly and effectively in security operations. At the same time, interpretable models offer high levels of transparency, though often at the expense of detection accuracy. Hybrid frameworks provide a

promising balance, but require further refinement and computational optimisation before they can be widely adopted.

The review also revealed persistent challenges, including limited real-world evaluation, a lack of standardized metrics for explanation quality, reliance on static datasets, and constraints in deploying XAI-enabled models in real-time systems. Addressing these gaps will be crucial for advancing practical, trustworthy phishing detection solutions. Future research should focus on developing lightweight, scalable XAI methods; conducting human-centered usability studies; improving dataset diversity; and exploring adversarial robustness within explainable frameworks.

Overall, XAI represents a critical component of modern phishing mitigation strategies. As phishing attacks continue to evolve, integrating explainability into detection systems will not only strengthen cybersecurity defenses but also support more informed, collaborative, and trustworthy decision-making across both technical and non-technical users.

REFERENCES

1. N. Alsuqayh, A. Mirza and A. Alhogail, "Exploring feature engineering and explainable AI for phishing website detection: a systematic literature review," International Journal of Electrical and Computer Engineering (IJECE), pp. DOI: 10.11591/ijece.v15i6.pp5863-5878, 2025.
2. P.R. Chandre, P. Bhujbal, A. Jadhav, B. D. Shendkar, A. Wangikar and R. Sachdeo, "A comprehensive review of interpretable machine learning techniques for phishing attack detection," IAES International Journal of Artificial Intelligence, pp. DOI: 10.11591/ijai.v14.i4.pp3022-3032, 2025.
3. M. Tawfik, A. A. Abu-Ein, A. Abdelhaliem and S. FathiIslam, "Explainable few-shot learning with modern BERT for detecting emerging phishing attacks using XF PhishBERT: Explainable few-shot learning with modern BERT...M. Tawfik et al.," Scientific Reports, pp. DOI: 10.1038/s41598-025-27500-0, 2025.
4. Thalpage, N. The Integration of Machine Learning and Explainable AI in Business Digitization: Unleashing the Power of Data - A Review. Journal of Digital Science, 6(1), 2023. https://doi.org/10.33847/2686-8296.6.1_2 .
5. F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Computer Science, Philosophy, 2017. https://api.semanticscholar.org/CorpusID:11319376.
6. N. S. Thalpage and T. A. D. Nisansala, "Exploring the Opportunities of Applying Digital Twins for Intrusion Detection in Industrial Control Systems of Production and Manufacturing – A Systematic Review," Data Protection in a Post-Pandemic Society, pp. https://doi.org/10.1007/978-3-031-34006-2_4, 2023.
7. M. Mehdi, Y. Farzaneh, S. Farzaneh, S. Elham, S. Elham and H. Gharaee, "An Adaptive Machine Learning Based Approach for Phishing Detection Using Hybrid Features," in Conference: 2019 5th International Conference on Web Research (ICWR), 2019. DOI: 10.1109/ICWR.2019.8765265
8. S. Baki and R. Verma, "Sixteen Years of Phishing User Studies: What Have We Learned?," IEEE Transactions on Dependable and Secure Computing, 2022, DOI: 10.1109/TDSC.2022.3151103
9. O. K. Sahingoz, E. Buber, O. Demir and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, 2019, https://doi.org/10.1016/j.eswa.2018.09.029.
10. R. Basnet, S. Mukkamala and A. H. Sung, "Detection of Phishing Attacks: A Machine Learning Approach," Studies in Fuzziness and Soft Computing, pp. DOI: 10.1007/978-3-540-77465-5_19, 2008.
11. M. Adebowale and K. Lwin, "Deep Learning with Convolutional Neural Network and Long Short-Term Memory for Phishing Detection," in Deep Learning with Convolutional Neural Network and Long Short-Term Memory for Phishing Detection, 2019, DOI: 10.1109/SKIMA47702.2019.898242.
12. U. Bhatt, A. Xiang and P. Eckersley, "Explainable machine learning in deployment," in Computer Science Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, https://doi.org/10.1145/3351095.3375624 .

13. S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in NIPS, 2017, DOI: 10.48550/arXiv.1705.07874.

14. M. T. Ribeiro, S. Singh and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in the 22nd ACM SIGKDD International Conference , 2016, https://doi.org/10.1145/2939672.2939778.

15. M. Sundararajan, A. TalyAnkur and T. Yan, "Axiomatic Attribution for Deep Networks," p. DOI: 10.48550/arXiv.1703.01365, 2017.

16. W. Samek, T. Wiegand, Klaus-Robert and M.-R. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," 2017, DOI: 10.48550/arXiv.1708.0829.

17. Vaswani, N. Shazeer, N. Parmar and I. Polosukhin, "Attention Is All You Need," p. DOI: 10.48550/arXiv.1706.03762, 2017.

18. G. Riccardo, A. Monreale, F. Turini and F. Giannotti, "A Survey of Methods for Explaining Black Box Models," ACM Computing Surveys , p. DOI: 10.1145/3236009, 2018.

19. T. Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," p. DOI: 10.1016/j.artint.2018.07.007, 2017.

20. F. Doshi-Velez and B. Kim, "A Roadmap for a Rigorous Science of Interpretability," p. DOI: 10.48550/arXiv.1702.08608, 2017.

21. F. Charmet, H. Chandra, Tanuwidjaja, S. Ayoubi and Z. Zhang, "Explainable artificial intelligence for cybersecurity: a literature survey," annals of telecommunications - annales des télécommunications, pp. DOI: 10.1007/s12243-022-00926-7, 2022.

Application of Virtual Reality in Occupational Health and Safety in Enterprises

Adéla Nováková[0009-0005-5623-0360], Lilia Dvořáková[0000-0001-6389-381X]

University of West Bohemia, Pilsen, Czech Republic

Abstract. The aim of this article is to evaluate the perceived benefits of virtual reality–based training in occupational health and safety (OHS) and first aid in a real workplace setting. The study addresses the need for more effective safety training approaches that support employees' preparedness for crisis and emergency situations, with an emphasis on practical skills development. The research employs an applied case study across seven organizations from different industries. The VR training was delivered as part of mandatory OHS programs and complemented by immersive simulations of emergency scenarios. Data were collected via a post-training questionnaire administered to 298 participants over a six-month period (May–October 2025). The analysis focuses on participants' subjective evaluation of the training, particularly perceived practical preparedness and self-confidence in providing first aid. The findings indicate high participant acceptance and consistently positive evaluations. Perceived realism and a strong sense of presence emerged as key benefits, fostering engagement and strengthening participants' perceived readiness to respond to real-world crisis situations.

Keywords: occupational health and safety, virtual reality, safety training, first aid, immersive learning, corporate education, employee preparedness.

1. Introduction

Occupational health and safety represent a fundamental obligation of every organization regardless of industry, as employers are required by applicable legislation to ensure safe working conditions and to regularly train employees in occupational safety and first aid. Legal regulations impose an obligation on companies to identify risks, prevent occupational accidents, and systematically prepare employees to respond to crisis and emergency situations (European Agency for Safety and Health at Work, 2023).

Despite these legal requirements, occupational accidents and crisis events remain a significant global issue. Millions of workplace accidents occur each year, with a substantial proportion linked to human error, insufficient training, or inadequate preparedness for non-standard situations (International Labour Organization, 2022).

Traditional forms of occupational safety training, such as classroom lectures, printed materials, or instructional videos, are commonly used to meet legislative requirements. However, these methods often fail to realistically simulate the stress, complexity, and time pressure associated with real crisis situations. As a result, employees may formally fulfill training requirements without acquiring sufficient practical preparedness or confidence to respond effectively in real-world emergencies.

Recent advances in digital technologies have opened new opportunities in corporate education, with virtual reality being identified in the literature as a promising tool for safety training. Virtual reality enables employees to undergo realistic simulations of hazardous situations in a safe and controlled environment and

allows repeated practice of critical procedures without risking participants' health or disrupting operational processes (Akdere et al., 2021).

Although existing research suggests a positive impact of virtual reality on participant engagement, motivation, and learning outcomes, empirical evidence from real corporate practice remains limited, particularly in the area of occupational safety and first aid training (Poyade et al., 2020). At the same time, the relationship between the level of immersion in virtual environments and performance indicators, such as reaction speed or error rates, has not yet been sufficiently explored.

The aim of this study is to empirically assess the perceived benefits of occupational health and safety and first aid training delivered through the VR Rescue system in a real corporate environment. Specifically, the study examines whether participants evaluate their practical preparedness for emergency situations and their self-confidence in providing first aid above the neutral value of the measurement scale. It further explores experiential characteristics of virtual reality—particularly perceived realism and sense of presence, which are considered key mechanisms for shaping user responses and learning in immersive environments (Slater, 2009). The contribution of this work lies in its applied perspective on using virtual reality as a complement to legally mandated occupational safety training and in providing empirical evidence from practice to support implementation decisions and future research.

2. Literature Review

Organizations today employ a wide range of training methods within occupational health and safety (OHS) programs, including face-to-face lectures, instructor-led training, e-learning modules, on-the-job instruction, and blended learning formats. These approaches are well suited to meeting legislative requirements and ensuring standardized knowledge transfer. However, for topics that require crisis response and decision-making under time pressure, such methods may encounter limitations, including limited opportunities for realistic practice, a lower level of experiential engagement, and difficulties in objectively measuring practical performance. For these reasons, the use of immersive methods—particularly virtual reality—has been increasingly discussed in the academic literature, as VR enables the simulation of hazardous situations in a controlled and safe environment and allows for repeated practice of critical procedures without real-world risk (Akdere et al., 2021; Kaplan et al., 2021; Babalola, 2023; Scorgie et al., 2024).

Virtual reality provides interactive, scenario-based training that can closely replicate complex work environments and emergency situations. Compared to traditional methods, VR supports experiential learning, standardization of training conditions, and immediate feedback. In the context of corporate training, VR has been associated with higher levels of participant engagement and greater training effectiveness, particularly for procedural tasks and situations requiring decision-making under pressure (Akdere et al., 2021). Empirical studies focusing on safety training indicate that participants in VR-based training may demonstrate higher motivation and engagement compared to less immersive formats, such as video-based instruction or slide presentations (Joshi et al., 2021; Haj-Bolouri & Katende, 2024). Review studies further suggest that immersive technologies achieve comparable or even higher effectiveness than traditional training methods, while offering added value in terms of safety, scalability, and training flexibility (Stefan et al., 2023; Kaplan et al., 2021).

Beyond participant engagement, it is essential to assess whether virtual reality leads to improvements in areas that are practically relevant for OHS, particularly self-confidence, perceived preparedness, and observable performance. Existing research suggests that VR-based training can enhance psychological readiness and self-

confidence in safety-related scenarios (Poyade et al., 2020; Carvalho et al., 2025). OHS training does not primarily aim to transmit theoretical knowledge alone, but rather to develop procedural skills, such as rapid risk identification, selection of appropriate actions, and accurate execution of procedures under stressful conditions. Immersive environments may support the development of these skills by exposing participants to realistic stimuli and time pressure that more closely resemble real emergency situations (Kaplan et al., 2021; Scorgie et al., 2023). At the same time, the literature highlights that the magnitude of these effects may vary depending on task complexity, scenario design quality, and participant characteristics (Akdere et al., 2021; Stefan et al., 2023).

Immersion represents one of the core characteristics of virtual reality and is often considered a key mechanism through which VR influences learning and performance. Higher levels of immersion may enhance attention, emotional engagement, and perceived realism, thereby facilitating better transfer of learned procedures into practice (Kaplan et al., 2021; Babalola, 2023). However, direct empirical evidence linking immersion to objective performance metrics—such as reaction time, procedural accuracy, or error rates—remains limited in real corporate environments (Kaplan et al., 2021; Poyade et al., 2020; Scorgie et al., 2023). An additional limitation is that many existing studies are conducted in laboratory settings or with educational samples, rather than within mandatory training delivered under routine operational conditions in organizations.

This article addresses these gaps through a case study of the VR Rescue project, which supports first aid and fire protection training as part of occupational health and safety programs. The system enables scenario-based simulations of emergency situations and simultaneously generates objective performance metrics during training. For the purposes of this study, however, only subjective data were analyzed. These data were collected immediately after training through a standardized online questionnaire; participants accessed the questionnaire via a QR code provided on-site and completed it directly after the VR training session. Objective performance metrics are presented solely as contextual information and as an illustration of the system's evaluative potential for internal organizational use and future research. By combining insights from the literature with applied evidence from practice, this study contributes to the ongoing discussion on the role of immersive methods in legislatively mandated OHS training and their potential to enhance perceived training quality and effectiveness.

3. Data and Methodology

The study is based on an applied multiple-case study of occupational health and safety (OHS) and first aid training delivered using the VR Rescue system. Data were collected as part of mandatory OHS training conducted in seven organizations operating across different sectors, including manufacturing, logistics, services, and public administration.

The VR Rescue system was integrated into existing, legally required training programs, which also included traditional internal training formats such as instructor-led classroom training, presentations, printed materials, and instructional videos. The research design therefore reflects real corporate training conditions, in which virtual reality complements rather than replaces traditional training methods.

The research design combines the collection of subjective questionnaire-based data with the potential collection of objective performance metrics generated by the VR Rescue system. However, in this publication, the analysis primarily focuses on subjective data obtained through the questionnaire survey.

Data collection took place between May and October 2025. Each participating organization implemented VR-based training as part of its regular cycle of mandatory

OHS training. Observations were conducted at the level of individual participants. For each participant, data were recorded during a single VR training session, while questionnaire data were collected immediately after the completion of the training in order to capture participants' immediate perceptions and experiences. Objective performance metrics were recorded continuously throughout the duration of the VR simulation.

The dataset includes both subjective and objective variables. Subjective variables were obtained through a questionnaire administered after the completion of the training and focused on participants' perceptions of the training. These variables included:

- perceived preparedness to respond to emergency and crisis situations,
- level of self-confidence in applying first aid procedures,
- perceived usefulness of the training,
- perceived level of immersion and realism experienced during the VR simulation.

Responses were measured using Likert-type scales. Within the VR Rescue system, objective performance metrics are automatically generated during simulations (e.g., quality of cardiopulmonary resuscitation, reaction times, error rates, or success in fire protection scenarios). These metrics are available within the system as supplementary data for internal training evaluation. However, in this study, these objective indicators were not included in the analysis and were not subject to empirical testing, as the primary aim of the study was to evaluate participants' subjectively perceived training benefits and self-confidence immediately after completing the VR training.

Descriptive statistical methods (mean, median, and standard deviation) were used to summarize questionnaire-based variables. Performance metrics generated by the VR Rescue system (e.g., reaction times, error rates, task completion time) are presented in this study only as contextual information and were not included in the statistical analysis.

Given the applied nature of the study and the heterogeneity of the participating organizations, the analysis primarily focuses on identifying patterns and tendencies rather than on strict causal inference. Based on the literature review and the study objectives, the following hypotheses were formulated:

H1: Participants evaluate training delivered through the VR Rescue system as beneficial for their practical preparedness for emergency situations (i.e., perceived benefit is statistically significantly above the neutral scale value).

H2: Participants evaluate training delivered through the VR Rescue system as beneficial for strengthening their self-confidence in providing first aid (i.e., above the neutral scale value).

Data analysis was primarily based on descriptive methods and one-sample inferential tests. First, basic descriptive statistics (mean, median, standard deviation) were calculated for key variables capturing perceived training benefits and participants' self-confidence.

To test hypotheses H1 and H2, one-sample statistical tests were applied to assess whether the mean rating of a given item or scale differed significantly from the neutral value of the Likert scale (e.g., the midpoint). Depending on the distributional characteristics of the data, either a one-sample t-test or its non-parametric alternative (the Wilcoxon signed-rank test) was used. Due to the exploratory and application-oriented nature of the study, no complex regression models or causal comparisons with other training methods were estimated.

The questionnaire consisted of closed-ended questions focusing on perceived training benefits and participants' self-confidence in relation to practical management of emergency situations and first aid provision. The key items and scales corresponded to the study hypotheses:

• perceived training benefit for practical preparedness (H1),
• perceived training benefit for strengthening self-confidence in providing first aid (H2).

Responses were recorded using a Likert scale, with the midpoint representing a neutral position. The full wording of the questionnaire is provided in the appendix of the article. The questionnaire was administered immediately after the completion of the training to minimize bias caused by time delay and to capture participants' immediate impressions and self-assessments. Data collection took place directly on the premises of the participating organizations. A total of 298 participants completed the VR simulation using the VR Rescue system. The response rate was 100%, and all 298 completed questionnaires were included in the analysis.

For the purpose of testing hypotheses H1 and H2, only complete responses for the relevant questionnaire items and scales were included in the analysis. The study involved adult participants attending mandatory OHS training. Participation in the research component of the study was voluntary, and all responses were anonymized. No personally identifiable data were collected, and results were analyzed exclusively at an aggregated level.

4. Results

This section presents the results of the empirical analysis focused on the perceived benefits of training delivered through the VR Rescue system. The analysis concentrates on participants' subjective evaluations of perceived practical preparedness, self-confidence in providing first aid, and experiential characteristics of virtual training, particularly the level of immersion and perceived realism of the simulations.

The results are presented using descriptive statistics followed by hypothesis testing for H1 and H2. Special attention is given to the role of immersion and realism, which emerged as key mechanisms influencing the overall perceived benefits of the training. Table 1 summarizes the basic descriptive statistics of key variables capturing perceived training benefits, self-confidence in providing first aid, and experiential aspects of virtual reality training. Responses were recorded using a five-point Likert scale (1–5), where the value of 3 represents a neutral position.

The results indicate that participants evaluated the VR Rescue training predominantly very positively. The mean values of all observed variables are substantially above the neutral midpoint of the scale, suggesting an overall favorable reception of this type of training.

The highest mean values were observed for variables related to perceived realism of the simulation and the sense of presence in the simulated situation. These findings suggest that participants did not perceive the VR training merely as a technological demonstration, but rather as a realistic simulation of emergency situations that closely resembles real-life interventions.

Table 1. Descriptive statistics of VR Rescue training evaluation (N = 298)

Variable	Mean	Standard Deviation	Median
Perceived benefit for practical preparedness	4.32	0.61	4
Strengthening of self-confidence in providing first aid	4.21	0.66	4
Perceived realism of the scenario	4.56	0.52	5
Feeling of "being directly at the scene of the incident"	4.59	0.49	5
Overall evaluation of the training	4.47	0.54	4

Note: Responses were recorded using a five-point Likert scale, where the value of 3 represents a neutral position.

Participants rated the feeling of "being directly at the scene of the incident" (M = 4.59) and the perceived realism of the scenario (M = 4.56) particularly highly, indicating a strong immersive character of the virtual reality training. These experiential characteristics can be interpreted as one of the key mechanisms supporting the perceived benefits of the training and its practical relevance in the context of occupational health and safety.

Hypotheses H1 and H2 were tested using one-sample statistical tests to assess whether the mean ratings of individual variables differ significantly from the neutral scale value (value = 3).

The results of hypothesis testing confirm H1, indicating that participants evaluated VR Rescue training as statistically significantly beneficial for their practical preparedness for emergency situations. The mean value of this variable substantially exceeds the neutral midpoint of the scale, and the difference is statistically significant at the $p < 0.001$ level.

Similarly, hypothesis H2 was confirmed. Participants perceived VR training as a tool that enhances their self-confidence in providing first aid, particularly when dealing with crisis situations and making decisions under pressure.

Table 2. Results of one-sample tests for hypotheses H1 and H2

Hypothesis	Variable	Mean	Test Value	p-value
H1	Perceived benefit for practical preparedness	4.32	3	< 0.001
H2	Strengthening of self-confidence in providing first aid	4.21	3	< 0.001

Note: The test value corresponds to the neutral midpoint of the Likert scale.

The findings suggest that the key benefit of VR Rescue training lies not only in information transfer, but primarily in the strong experiential component associated with a sense of realism and presence in the simulated environment. These characteristics of virtual reality likely contribute to higher participant engagement and to strengthening their subjective preparedness and self-confidence.

The immersive nature of the training appears particularly important for topics where rapid response, decision-making under time pressure, and stress management are critical. The results therefore support the assumption that virtual reality can serve as an effective complement to traditional forms of occupational health and safety training, especially in areas focused on emergency and crisis situations.

5. Conclusion

The aim of this article was to assess the perceived benefits of occupational health and safety (OHS) and first aid training delivered through the VR Rescue system. The study focused on participants' subjective evaluations, particularly their perceived practical preparedness and self-confidence after completing VR-based training. The objective was to determine whether the use of virtual reality can contribute to more effective OHS training, especially in the context of emergency situations.

An applied case study design was employed, involving seven organizations operating across different sectors, including manufacturing, logistics, and public

administration. Data were collected through a questionnaire survey completed by 298 participants who attended VR training as part of mandatory OHS programs. The questionnaire focused on perceived training benefits and the strengthening of self-confidence in providing first aid. Hypotheses were tested using one-sample statistical tests to examine whether ratings of key variables (perceived benefit and self-confidence) significantly exceeded the neutral scale value.

The results demonstrate that training delivered through the VR Rescue system was evaluated very positively. Participants perceived the use of virtual reality as significantly beneficial for their practical preparedness for emergency situations. The realism of the scenarios was highlighted as a key strength, providing participants with a sense of "being at the scene" and enabling them to respond to situations that would require immediate and correct decision-making in real life. The findings also confirm that VR-based training strengthens participants' self-confidence in providing first aid.

These results are highly relevant for managers and HR professionals, particularly when considering innovations in employee training and development. The use of VR technology in OHS training provides a realistic and safe environment for practicing crisis scenarios, leading to improved employee preparedness and higher training effectiveness. This approach has strong potential to increase participant engagement, optimize training time, and ensure compliance with legislative requirements.

Integrating VR training into standard training processes may also help reduce costs associated with real-life emergency simulations, as scenarios can be repeated without risk to participants and without disrupting organizational operations.

Several limitations of the study should be considered when interpreting the results. The primary limitation is the absence of a control group, which prevents direct comparison of VR training effectiveness with other training methods, such as instructor-led sessions or e-learning modules. Furthermore, data were collected immediately after training, which does not allow assessment of long-term transfer of learned skills into practice. In addition, the study focused exclusively on subjective evaluations and did not assess objective changes in behavior or performance during real emergency situations.

Future research should therefore focus on longitudinal study designs to examine the long-term impact of VR training on participant performance in real-world emergency situations. Research incorporating control groups would allow direct comparison of VR-based training with traditional training methods. Another promising direction is the inclusion of objective performance measurements, such as real-time simulated emergency tasks, to assess actual behavioral change.

Further research could also explore different virtual reality scenarios targeting specific situations (e.g., fires, workplace injuries, traffic accidents) and their effects on particular skills and competencies within the OHS domain.

Acknowledgments

REFERENCES

1. Akdere, M., Acheson, K., & Jiang, Y. (2021). Virtual reality–based training effectiveness in corporate learning. *Human Resource Development Review, 20*(3), 258–281. https://doi.org/10.1177/15344843211007045

2. Babalola, S. O. (2023). Virtual reality safety training: A systematic review. *Cognitive Robotics, 3*, 226–241. https://doi.org/10.1016/j.cogr.2023.10.004

3. Haj-Bolouri, A., Katende, J., & Rossi, M. (2024). Gamified immersive safety training in virtual reality: A mixed methods approach. *Journal of Workplace Learning, 36*, 516–538. https://doi.org/10.1108/JWL-01-2024-0008

4. Joshi, D., Yadav, R., & Sharma, A. (2021). Evaluating virtual reality–based safety training: Motivation and learning outcomes. *Safety Science, 139*, Article 105256. https://doi.org/10.1016/j.ssci.2021.105256

5. Kaplan, A. D., Cruit, J., Endsley, M., Beers, S., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods. *Human Factors, 63*(4), 706–726. https://doi.org/10.1177/0018720820904229

6. Poyade, M., Morris, J., & Taylor, J. (2020). Virtual reality training for occupational safety: Presence, confidence, and performance. *Computers & Education, 148*, Article 103809. https://doi.org/10.1016/j.compedu.2020.103809

7. Scorgie, M., Paes, M., Feng, A., Parisi, L., Yiu, Z., & Lovreglio, M. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis. *Safety Science, 171*, Article 106372. https://doi.org/10.1016/j.ssci.2023.106372

8. Slater, M. (2009). *Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments*. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Measuring the Level of Implementation of Quality 4.0 Dimensions: A Case Study at the College of Administration and Economics – University of Mosul

Zaid Khaleel Ibrahim[0009-0006-0810-595X]
Omar Ali Ismael[0000-0002-8384-6532]
Amal Sarhan Sulaiman[0000-0003-0169-1002]

University of Mosul, Nineveh, Iraq

Abstract. The concepts of quality management have witnessed significant development in recent decades, evolving from traditional quality focused on products and services, through total quality encompassing all organizational processes, to Quality 4.0, which represents the integration of modern quality practices with digital technologies and continuous innovation. This study aims to measure the level of implementation of Quality 4.0 dimensions at the College of Administration and Economics – University of Mosul, and to analyze the gaps between the current reality and the optimal requirements, to provide practical recommendations that enhance institutional performance and student experience. The research adopted a case study approach, and data were collected using a checklist comprising 25 items covering the five dimensions of Quality 4.0: technological dimension, process and procedures dimension, human resources and training dimension, customer satisfaction and experience dimension, and sustainability and innovation dimension. Descriptive and inferential statistical analyses were employed to calculate weighted averages, percentages, and implementation gaps. The results indicated that the overall implementation rate of the five dimensions reached 76%, with a total execution gap of 24%, reflecting a good level of Quality 4.0 application, albeit with variation across dimensions. The human resources and training dimension achieved the highest implementation rate at 84%, whereas the sustainability and innovation dimension recorded the lowest at 68%, highlighting uneven gaps and limited challenges in certain areas. Accordingly, the study concluded that the college partially implements Quality 4.0 dimensions, with a need to enhance digital strategies, develop technological infrastructure, increase investment in sustainable innovation, and expand training and awareness programs for academic and administrative staff to ensure full integration of all dimensions.

Keywords: Quality 4.0, Digital College, Institutional Performance, Technological Dimension, Sustainable Innovation, University Quality Improvement, Student Experience.

1. Introduction

Quality has become one of the fundamental pillars for achieving institutional excellence and ensuring sustainability in higher education organizations, particularly in an environment characterized by rapid change and increasing competition. The concepts of quality management have witnessed significant evolution through multiple stages, beginning with traditional quality, which focused on inspecting the final product, followed by Total Quality Management (TQM), which emphasized process improvement and the involvement of all staff, and culminating in the Quality 4.0 stage, representing a transformative shift in management thinking by integrating quality principles with Industry 4.0 technologies [7].

Quality 4.0 is defined as an integrated framework that combines modern quality management practices with advanced digital technologies such as artificial intelligence, big data, the Internet of Things, cloud computing, and predictive analytics, aiming to support decision-making, enhance institutional performance, and increase value delivered to customers and stakeholders [1]. Thus, quality management is no longer solely dependent on traditional methods or historical indicators; it now relies on real-time data, intelligent analysis, and rapid responsiveness to change.

In the context of higher education, universities face increasing challenges, including improving the quality of education, developing curricula, enhancing the efficiency of academic and administrative services, and enriching the learning experience, in line with digital transformation requirements and stakeholder expectations from students, society, and employers. Recent literature indicates that adopting Quality 4.0 concepts in universities contributes to improved governance, greater transparency, enhanced institutional innovation, and supports academic accreditation, thereby promoting both local and international competitiveness [12].

The dimensions of Quality 4.0 provide an important analytical framework for assessing the readiness of educational institutions for digital transformation. These dimensions include the technological dimension, focusing on digital infrastructure and the use of intelligent systems; the process dimension, concerned with streamlining and automating procedures; the human resources dimension, emphasizing the development of digital capabilities and fostering a quality culture; the customer satisfaction dimension, which measures the satisfaction of students and stakeholders with educational services; and the sustainability and innovation dimension, ensuring continuous improvement and adaptability to future changes [1].

Despite the growing importance of Quality 4.0, its implementation in universities—particularly in developing countries—still faces several challenges, including limited technological readiness, constrained financial resources, shortage of digitally qualified human capital, and organizational resistance to change [19]. Moreover, the absence of clear standardized tools to measure the level of implementation of Quality 4.0 dimensions complicates the identification of gaps between current practices and desired standards.

In this context, there is a pressing need for field studies that measure the level of implementation of Quality 4.0 dimensions in higher education institutions, identify strengths and weaknesses, and provide practical, actionable recommendations. This study aims to contribute to this area by assessing the level of implementation of Quality 4.0 dimensions at the College of Administration and Economics – University of Mosul, as a key institution in preparing administrative and economic personnel, and by analyzing the gaps between current practices and modern digital quality requirements, thereby supporting decision-makers in enhancing institutional performance, improving stakeholder experience, and achieving sustainable academic excellence.

2. Literature Review

1) Concept of Quality 4.0

Quality 4.0 represents an advanced stage in the evolution of quality management thought, emerging as a response to the radical transformations brought about by the Fourth Industrial Revolution in business and service environments, particularly within higher education institutions. Whereas traditional quality focused on inspecting the final product, Total Quality Management expanded to encompass processes and organizational participation. Quality 4.0 has emerged as an integrated framework that combines modern quality principles with smart digital technologies to achieve superior data- and knowledge-driven institutional performance [7].

Quality 4.0 is viewed as a management philosophy that relies heavily on advanced digital technologies, including artificial intelligence, big data, the Internet of Things, cloud computing, predictive analytics, and blockchain, to support planning, control, and continuous quality improvement activities [1]. From this perspective, quality management is no longer conducted through traditional methods relying on periodic reports and lagging indicators; instead, it has become a dynamic process based on real-time data, intelligent analysis, and proactive decision-making [10].

Antony et al. (2021) note that Quality 4.0 does not replace the principles of Total Quality Management; rather, it extends them by enhancing classical quality tools such as process maps, cause-and-effect diagrams, and Six Sigma through digital technologies that increase accuracy, speed, and predictive capabilities. Consequently, Quality 4.0 integrates human and technological elements within a comprehensive framework focused on creating sustainable value for customers and stakeholders.

In the context of higher education, Quality 4.0 is defined as a university's institutional capability to leverage smart digital technologies to enhance educational quality, develop academic and administrative processes, and improve the experience of stakeholders, including students, faculty members, and the community, thereby achieving academic excellence and institutional sustainability [11,15]. This concept emphasizes the transition from traditional to smart digital education, from paper-based to electronic administration, and from lagging evaluation to continuous data-driven assessment[23,20].

Quality 4.0 is also based on several core principles, including data-driven decision-making, systeem integration, organizational agility, continuous innovation, and a focus on customer satisfaction. These principles contribute to faster responsiveness to change, reduced resource waste, enhanced institutional performance, and increased trust in educational outcomes [19].

Recent literature highlights that implementing Quality 4.0 in universities requires several essential prerequisites, including digital infrastructure, intelligent quality management systems, digitally qualified human resources, and an organizational culture supportive of digital transformation and continuous improvement [21,22]. In the absence of these prerequisites, implementing Quality 4.0 may face challenges related to resistance to change, poor coordination among organizational units, and underutilization of available technologies [11].

Accordingly, Quality 4.0 represents a transformative shift in quality management, moving educational institutions from a focus on compliance with standards to smart excellence, from incremental improvement to predictive and proactive enhancement, and from merely satisfying customers to creating a comprehensive and sustainable learning experience. From this standpoint, measuring the level of implementation of Quality 4.0 dimensions in higher education institutions emerges as a crucial step to diagnose the current reality and identify pathways for future development.

2) Importance of Quality 4.0

Quality 4.0 is considered one of the modern management approaches that has gained increasing significance due to rapid digital transformation and rising expectations of customers and stakeholders. It contributes to integrating quality management principles with smart digital technologies, thereby enhancing operational efficiency, supporting data-driven decision-making, and enabling organizations to shift from addressing problems after they occur to predicting and preventing them proactively [1,7].

In universities, the importance of Quality 4.0 is reflected in its role in enhancing the quality of teaching and learning through the use of digital learning systems and educational analytics, streamlining and automating academic and administrative processes, and increasing stakeholder satisfaction, including that of students, faculty members, and the broader community. Moreover, Quality 4.0 supports academic

accreditation requirements, enhances institutional competitiveness, and develops human resource capabilities, thereby achieving academic excellence and institutional sustainability in a continuously changing educational environment [20,12,22].

3) Dimensions of Quality 4.0

1. Technological Dimension

The technological dimension constitutes the foundational pillar of Quality 4.0, as it relies on leveraging advanced digital technologies to support quality activities and enhance institutional performance. This includes the use of artificial intelligence for data analysis and intelligent decision-making, big data to understand performance patterns and predict potential issues, the Internet of Things for real-time process monitoring, and cloud computing to facilitate system integration and easy access to information [9,17].

In universities, this dimension contributes to improving the quality of education and academic services through intelligent learning systems, performance dashboards, and digital quality management systems, supporting continuous monitoring and reducing reliance on traditional manual methods [3]. Moreover, the technological dimension promotes a data-driven decision-making culture and enhances the accuracy and transparency of institutional evaluation [18].

2. Process and Operational Dimension

This dimension focuses on redesigning and streamlining academic and administrative processes in line with digital quality requirements by reducing complexity, eliminating waste, and ensuring smooth procedures. Literature emphasizes that Quality 4.0 relies on digitalizing and continuously analyzing processes to identify weaknesses and gaps between actual and targeted performance [4].

It also involves the use of digital continuous improvement tools such as Six Sigma 4.0 and Lean Digital, which combine classical improvement methodologies with digital analysis techniques, contributing to enhanced operational efficiency and improved quality of educational and administrative outputs in universities [5].

3. Human Resource Dimension

The human resource dimension of Quality 4.0 emphasizes that technology alone cannot achieve quality without qualified personnel capable of effectively utilizing it. Implementing Quality 4.0 requires developing the skills of academic and administrative staff, particularly digital, analytical, and system-handling competencies [13].

This dimension also focuses on fostering a culture of quality, innovation, and continuous improvement, encouraging active participation of employees and students in digital improvement initiatives, which enhances shared responsibility and institutional commitment. Universities that invest in digital training and capacity building are better positioned to successfully implement Quality 4.0 and achieve their strategic objectives [8].

4. Customer/Stakeholder Experience Dimension

This dimension focuses on measuring and improving the satisfaction of stakeholders, including students, faculty members, and the academic community, regarding digital educational and administrative services. Quality 4.0 shifts universities from focusing solely on service delivery to prioritizing the customer experience as a central element in assessing institutional performance [14].

Digital technologies facilitate the collection of real-time stakeholder feedback and its analysis using intelligent tools, enabling universities to respond promptly to expectations and enhance service quality, speed, and accuracy. This dimension also strengthens trust in the educational institution and increases stakeholder satisfaction and loyalty [6].

5. Sustainability and Innovation Dimension

This dimension relates to an educational institution's ability to achieve continuous improvement and innovation in a changing environment, integrating digital, environmental, and economic sustainability principles into quality activities. Quality 4.0 focuses not only on immediate improvement but also on building flexible systems capable of adapting to future challenges [2].

In universities, this dimension contributes to developing academic programs, promoting scientific research, and enhancing efficient use of resources, thereby ensuring sustainable institutional performance and strengthening the university's societal role. It also supports innovation in teaching and assessment methods, creating a smart learning environment capable of responding to future transformations [16].

3. Data and Methodology

1) **Study Problem :**Despite the growing emphasis on quality in Iraqi universities, the transition toward Quality 4.0 remains limited. The main problem lies in the lack of knowledge regarding the readiness of the College of Administration and Economics to implement these modern digital quality dimensions, as well as the existing gaps between the current reality and the optimal practices required for effective application of Quality 4.0.

2) **Study Objectives:** The study aims to:
✓Measure the level of implementation of Quality 4.0 dimensions at the College of Administration and Economics – University of Mosul.
✓Analyze the gaps between the current reality and the optimal requirements for implementing Quality 4.0.
✓Provide practical recommendations to enhance the college's readiness for applying modern digital quality practices.

3) **Significance of the Study** The significance of this study lies in:
✓Highlighting the technological, administrative, and human aspects related to the implementation of Quality 4.0 in Iraqi universities.
✓Providing a standardized framework that can be used to assess the readiness of other colleges for implementing Quality 4.0.
✓Offering strategic recommendations to assist decision-makers at the college in enhancing institutional performance and improving the student experience.

4) **Research Hypotheses:** Based on the study's objectives, which aim to measure the level of implementation of Quality 4.0 dimensions at the College of Administration and Economics – University of Mosul and analyze the gaps between current practices and optimal requirements, the research hypotheses are formulated as follows:
✓Hypothesis 1: The college implements Quality 4.0 dimensions partially, with the presence of execution gaps.
✓Hypothesis 2: The lack of technological and human resources affects the level of implementation of the dimensions.

5) **Scope of the Study**
✓Spatial Scope: The study is limited to the College of Administration and Economics at the University of Mosul, which was selected as the field for applying and measuring the Quality 4.0 dimensions.
✓Temporal Scope: The study period covered six months, with data collection beginning in July 2025 and continuing until December 2025. Accordingly, the study results are confined to this time frame.
✓Human Scope: The study population included several department officials and academic staff at the College of Administration and Economics, selected due to their familiarity with academic and administrative processes and their ability to assess the level of implementation of Quality 4.0 dimensions.

6)   Research Methodology: The researchers adopted the case study approach to examine the current status of Quality 4.0 implementation at the college. Descriptive and inferential statistical analyses were employed in the field study, following data collection from the study samples using specially designed instruments. The data were analyzed to test the research hypotheses and derive practical findings.

7)   Field Data Collection Methods: To conduct the field aspect of the study, a checklist was employed to assess the college's capacity to implement Quality 4.0 dimensions. The checklist comprised 25 items covering all Quality 4.0 dimensions: technological, process, human resources, customer/stakeholder satisfaction, and innovation & sustainability. A five-point Likert scale was used to obtain detailed and accurate results, with scores ranging from 1 to 5, where:

✓1 = Not implemented at all
✓2 = Poorly implemented
✓3 = Partially implemented
✓4 = Well implemented
✓5 = Fully implemented

Additionally, personal interviews were conducted with selected managers according to their specializations, responsibilities, and authorities to ensure the accuracy and comprehensiveness of the data.

8)   Statistical Analysis Tools: After collecting field data from the study site, the researchers employed several statistical tools to analyze the data and interpret the results. These tools include:

Weighted Means: Calculated using the following formula:
✓   Weighted Mean = Sum of Scores ÷ Frequency
Frequencies: Number of responses for each of the weights used in the study.
✓Score (Result) = Frequency × Weight
Percentages: Calculated using the following formula:
✓Percentage (%) = (Weighted Mean ÷ Highest Weight) × 100

4. Results

4.1.1. Results of the Technological Dimension Implementation

The statistical analysis results indicated that the weighted mean for the technological dimension was 3.6, while its implementation rate at the College of Administration and Economics – University of Mosul reached 72%, with a gap of 28%. This percentage reflects a good level of integration of digital technologies in supporting academic and administrative processes, contributing to improved performance quality, faster task completion, and reduced waste.

Table 1 shows that most elements of the technological dimension are notably implemented, particularly regarding the use of electronic systems for process automation and enhancing information management efficiency. The results also indicate that the college is making clear efforts to utilize digital analytics tools to improve institutional performance and monitor quality indicators.

However, the observed gap of 28% highlights certain challenges that may hinder achieving optimal implementation of this dimension, including the need for continuous updating of technological infrastructure, expanding the application of smart systems across all college units, and improving the use of available data for advanced and predictive analysis.

Overall, the results suggest that the College of Administration and Economics has largely succeeded in adopting the technological dimension within the Quality 4.0 framework. Nonetheless, narrowing the gap requires strengthening digital strategies, increasing investments in modern technology, and intensifying training and awareness programs for academic and administrative staff, thereby enhancing digital

maturity and ensuring full implementation of modern quality practices (Table 1).

Table 1. Results of the Technological Dimension Implementation

No.	Technological Dimension Items	Not Implemented	Poorly Implemented	Partially Implemented	Well Implemented	Fully Implemented
1	The College of Administration and Economics relies on modern digital systems (e.g., learning management systems or electronic administrative systems) to support educational and administrative processes.				✓	
2	The college uses digital technologies to collect and analyze data related to academic and administrative performance.				✓	
3	The available technological infrastructure contributes to improving the quality of services provided to students and faculty members.			✓		
4	The college relies on automation and electronic systems to reduce errors and expedite administrative procedures.				✓	
5	Technological systems in the college support data-driven academic and administrative decision-making.			✓		
	Weights	1	2	3	4	5
	Frequencies	0	0	2	3	0
	Score (Result)	0	0	6	12	0
	Weighted Mean	3.6				
	Percentage	%72				
	Implementation Gap	%28				

Note: The table was prepared by the researchers.

4.1.2. Results of the Process and Operational Dimension Implementation

The analysis results indicated that the weighted mean for the process and operational dimension was 3.8, with an implementation rate of 76% and a gap of 24%. These results reflect an advanced level of organization in academic and administrative processes, contributing to the simplification of procedures, reduction of bureaucracy, and improvement of performance efficiency.

Table 2 shows that the college implements most elements of this dimension, notably, particularly regarding the reduction of bureaucracy, enhancing service delivery speed, documenting procedures, periodic review, and performance measurement through clear indicators. The data also indicate that the college relies on digital continuous improvement tools to support process development in line with Quality 4.0 standards.

However, the 24% gap highlights the need for further development in some areas, such as expanding the use of digital improvement tools, enhancing integration between academic and administrative processes, and improving the monitoring of performance indicators to ensure the achievement of desired objectives.

Overall, the results demonstrate that the College of Administration and Economics has achieved a good level of implementation for the process and operational dimension. Nevertheless, narrowing the gap requires continuous process development, the use of digital tools for performance analysis, and fostering a culture of continuous improvement among academic and administrative staff (Table 2).

Table 2. Results of the Process and Operational Dimension Implementation

No.	Process & Operational Dimension Items	Not Implemented	Poorly Implemented	Partially Implemented	Well Implemented	Fully Implemented
1	Academic and administrative processes at the College of Administration and Economics are characterized by clarity, documentation, and ease of implementation.				✓	
2	The college continuously reviews and develops its academic and administrative procedures.			✓		
3	The college uses clear performance indicators to measure the efficiency of academic and administrative processes.				✓	
4	The adopted procedures contribute to reducing bureaucracy and improving service delivery speed.					✓
5	The college uses modern methods to analyze gaps between current performance and the desired level.			✓		
	Weights	1	2	3	4	5
	Frequencies	0	0	2	2	1
	Score (Result)	0	0	6	8	5
	Weighted Mean	3.8				
	Percentage	%76				
	Implementation Gap	%24				

Note: The table was prepared by the researchers.

4.1.3. Results of the Human Resource and Training Dimension Implementation

The study results indicated that the weighted mean for the human resource and training dimension was 4.2, while its implementation rate at the College of Administration and Economics – University of Mosul reached 84%, with a gap of 16%. This percentage reflects an advanced level of skill development and capacity-building practices among academic and administrative staff, enabling effective engagement with Quality 4.0 requirements.

Table 3 shows that most items of this dimension were highly implemented, particularly regarding the provision of specialized training programs to enhance technical skills and the ability to use digital systems, in addition to fostering a culture of quality and continuous improvement among staff. The data also indicate clear efforts to encourage staff participation in developmental initiatives and contribute to improving the quality of academic and administrative processes.

Despite this outstanding performance, the 16% gap highlights some challenges, such as the need to expand advanced training in modern digital technologies, enhance student participation in digital improvement programs, and develop mechanisms to motivate staff for continuous innovation.

Overall, the results confirm that the college has achieved a high level of implementation for the human resource and training dimension within the Quality 4.0 framework. Nevertheless, narrowing the remaining gap requires continued investment in staff development, implementation of specialized training programs, and fostering a culture of participation and innovation to ensure effective integration between human resources and available digital technologies.

Table 3. Results of the Human Resource and Training Dimension Implementation

No.	Human Resource & Training Dimension Items	Not Implemented	Poorly Implemented	Partially Implemented	Well Implemented	Fully Implemented
1	Academic and administrative staff at the College of Administration and Economics possess adequate skills to use modern digital technologies.				✓	
2	The college organizes training courses to develop employees' digital skills and technical capacities.				✓	
3	The college fosters a culture that encourages quality, innovation, and continuous improvement.					✓
4	The college administration encourages faculty and staff to propose digital development ideas.				✓	
5	Students are involved in some improvement and development initiatives related to educational services.				✓	
	Weights	1	2	3	4	5
	Frequencies	0	0	0	4	1
	Score (Result)	0	0	0	16	5
	Weighted Mean	4.2				
	Percentage	%84				
	Implementation Gap	%16				

Note: The table was prepared by the researchers.

3.1.4. Results of the Customer/Stakeholder Experience Dimension Implementation

The statistical analysis results indicated that the weighted mean for the customer/stakeholder experience dimension was 4.0, with an implementation rate of 80% and an implementation gap of 20%. These results reflect a good level of application of this dimension at the College of Administration and Economics – University of Mosul, demonstrating the college's attention to measuring student and faculty satisfaction and its efforts to enhance their educational and administrative experience.

Table 4 shows that most items of this dimension are appropriately implemented, particularly regarding enhancing responsiveness to complaints and suggestions, improving the quality of educational and administrative services, and

utilizing various methods to collect student feedback and proposals. The results also indicate that the use of digital technologies has contributed to improving the teaching and learning experience and increasing the speed and accuracy of service delivery within the college.

Despite this good level of implementation, the 20% gap highlights areas still needing development, such as adopting more systematic and comprehensive mechanisms to measure customer satisfaction and expanding the use of interactive digital systems that focus more holistically on the customer experience.

Overall, the results indicate that the College of Administration and Economics has made significant progress in implementing the customer/stakeholder experience dimension within the Quality 4.0 framework. Nevertheless, narrowing the remaining gap requires enhancing satisfaction measurement systems, developing digital communication channels, and improving the design of educational services to align with customer expectations and increase overall satisfaction.

Table 4. Results of the Customer/Stakeholder Experience Dimension Implementation

No.	Customer/Stakeholder Experience Dimension Items	Not Implemented	Poorly Implemented	Partially Implemented	Well Implemented	Fully Implemented
1	The College of Administration and Economics ensures measuring student satisfaction with the provided educational and administrative services.				✓	
2	The digital technologies used in the college contribute to improving the teaching and learning experience.			✓		
3	The college responds appropriately to complaints and suggestions from students and faculty members.					✓
4	The services provided by the college are characterized by speed, accuracy, and acceptable quality.				✓	
5	The college works on developing its services in line with the expectations of students and stakeholders.				✓	
	Weights	1	2	3	4	5
	Frequencies	0	0	1	3	1
	Score (Result)	0	0	3	12	5
	Weighted Mean	4				
	Percentage	%80				
	Implementation Gap	%20				

Note: The table was prepared by the researchers.

4.1.5. Results of the Sustainability and Innovation Dimension Implementation

The statistical analysis results indicated that the weighted mean for the sustainability and innovation dimension was 3.4, with an implementation rate of 68% and an implementation gap of 32%. These results reflect a moderate level of integration of continuous innovation practices and sustainability strategies in the academic and administrative processes at the College of Administration and Economics – University of Mosul.

Table 5 shows that the college implements some items of this dimension

significantly, particularly those related to encouraging scientific research and adopting sustainable practices in resource utilization. The data also indicate the college's capability to provide a flexible learning environment that fosters continuous improvement and supports innovation.

However, the 32% gap highlights challenges that need to be addressed, such as promoting a culture of sustainable innovation among all stakeholders, expanding the application of digital and environmentally sustainable practices, and improving mechanisms for measuring the impact of innovation on the quality of educational and administrative services.

Overall, the results indicate progress in implementing the sustainability and innovation dimension within the Quality 4.0 framework. Narrowing the remaining gap requires increased investment in digital and sustainable innovation, development of motivating training programs for academic and administrative staff, and systematic strategies to transform innovative ideas into ongoing practical practices that provide added value for the college and its stakeholders.

Table 5. Results of the Sustainability and Innovation Dimension Implementation

No.	Sustainability & Innovation Dimension Items	Not Implemented	Poorly Implemented	Partially Implemented	Well Implemented	Fully Implemented
1	The College of Administration and Economics employs innovative methods in developing its academic programs and administrative services.			✓		
2	The college seeks to achieve sustainability in the use of material and technological resources.			✓		
3	The college administration encourages adopting new ideas that contribute to improving institutional performance quality.				✓	
4	The college can adapt to future technological and organizational changes.			✓		
5	The adopted quality practices contribute to achieving continuous and sustainable performance improvement.				✓	
Weights		1	2	3	4	5
Frequencies		0	0	3	2	0
Score (Result)		0	0	9	8	0
Weighted Mean		3.4				
Percentage		%68				
Implementation Gap		%32				

Note: The table was prepared by the researchers

4.1.6. Level of Implementation of Quality 4.0 Dimensions

After analyzing the results of the summary table of the five dimensions, it is evident that the College of Administration and Economics – University of Mosul demonstrates a clear orientation toward implementing the dimensions of Quality 4.0. The overall implementation rate reached 76%, while the total implementation gap

amounted to 24%. These results indicate a good level of progress, with relative variation in the levels of implementation across the dimensions, reflecting differing priorities in resource allocation among them.

Analysis of Each Dimension:

✓ Technological Dimension: The implementation rate reached 72%, with a gap of 28%. This indicates the college's commitment to using digital technologies to support academic and administrative performance. However, there is a need to further strengthen technological infrastructure, expand the application of smart systems, and improve data utilization for advanced analytical purposes.

✓ Processes and Procedures Dimension: This dimension recorded an implementation rate of 76%, with a gap of 24%. The results reflect efficient application of most academic and administrative procedures, with room for improving integration between processes and digital tools to further enhance institutional performance.

✓ Human Resources and Training Dimension: This dimension achieved the highest level of implementation among the five dimensions, at 84%, with a gap of 16%. This reflects the college's success in developing the skills of academic and administrative staff and fostering a culture of quality and continuous improvement. Nevertheless, there remains a need to expand specialized training programs and enhance the active participation of students and staff in digital improvement initiatives.

✓ Customer Satisfaction and Experience Dimension: The implementation rate reached 80%, with a gap of 20%. This indicates a good level of improvement in the experience of students and faculty members, with a need to further develop satisfaction measurement systems and improve responsiveness to feedback and suggestions to ensure a distinguished educational experience.

✓ Sustainability and Innovation Dimension: This dimension recorded the lowest implementation rate among the five dimensions, at 68%, with a gap of 32%. This highlights the importance of deepening a culture of sustainable innovation, expanding the application of digital and environmental practices, and strengthening scientific research and continuous improvement to achieve institutional sustainability.

▪ Conclusions Related to the Research Hypotheses:

✓ First Hypothesis: The college applies the dimensions of Quality 4.0 partially, with existing implementation gaps.

• Hypothesis confirmed: The results reveal variation among the dimensions, with human resources and training showing the highest implementation rate (84%), while sustainability and innovation recorded the lowest (68%). The presence of gaps ranging from 16% to 32% confirms that implementation is partial rather than complete.

✓ Second Hypothesis: The lack of technological and human resources affects the level of implementation of the dimensions.

• Hypothesis confirmed: The gaps observed in the technological dimension (28%) and the sustainability and innovation dimension (32%) indicate the need to enhance technological infrastructure and increase investment in innovation and sustainable practices, supporting the validity of this hypothesis.

Overall, the findings demonstrate that the college has achieved a good level of Quality 4.0 implementation. However, it requires strategic actions to strengthen

integration among the dimensions and reduce implementation gaps to achieve a more comprehensive and efficient application.

Table 6. Results of the Overall Level of Quality 4.0 Dimensions Implementation

Quality 4.0 Dimensions	Weighted Mean	Percentage	Gap Size
Technological	3.6	%72	%28
Processes and Procedures	3.8	%76	%24
Human Resources and Training	4.2	%84	%16
Customer Satisfaction and Experience	4	%80	%20
Sustainability and Innovation	3.4	%68	%32
Overall Average	3.8	%76	%24

Source: Prepared by the researchers.

5. Conclusion

The results of this study demonstrate that the College of Administration and Economics exhibits a commitment to implementing Quality 4.0, achieving a generally good level of adoption across the five examined dimensions. Despite observable implementation gaps ranging from 16% to 23%, the college has made notable progress in digital transformation and quality enhancement. The human resources and training dimension attained the highest implementation level, reflecting effective capacity building; however, further expansion of advanced training and student engagement in digital quality initiatives is needed. The technological dimension and processes and procedures dimension also showed solid implementation but require enhancements in infrastructure, intelligent system utilization, and digital integration to achieve greater operational efficiency. The customer satisfaction and experience dimension highlights the college's efforts to improve stakeholder experience; nonetheless, more robust feedback mechanisms and responsiveness are essential. The sustainability and innovation dimension recorded the lowest level of implementation, underscoring the need for stronger sustainable innovation practices and mechanisms to evaluate their impact.

In light of these findings, the hypotheses validating partial application of Quality 4.0 were supported, with identified constraints related to technological resources and human capital development. Therefore, this study recommends developing a comprehensive strategic plan for Quality 4.0 to minimize implementation gaps, strengthen integration across all dimensions, and promote sustainable digital innovation, ultimately advancing institutional performance and educational excellence.

REFERENCES

1. J. Antony, M. Sony, O. McDermott, and R. Jayaraman, "An exploration of organizational readiness factors for Quality 4.0: An intercontinental study," International Journal of Production Research, vol. 59, no. 5, pp. 1–18, 2021. https://doi.org/10.1080/00207543.2020.1861358
2. S. Bag, S. Gupta, and S. Kumar, "Industry 4.0 adoption and sustainable manufacturing practices," International Journal of Production Economics, vol. 234, 108018, 2021. https://doi.org/10.1016/j.ijpe.2020.108018
3. M. Bond, O. Zawacki-Richter, and M. Nichols, "Mapping research in educational technology," Computers & Education, vol. 148, 103805, 2020. https://doi.org/10.1016/j.compedu.2020.103805
4. S. V. Buer, J. O. Strand Hagen, and F. T. S. Chan, "The link between Industry 4.0 and lean manufacturing," International Journal of Production Research, vol. 56, no. 8, pp. 2925–2940, 2018. https://doi.org/10.1080/00207543.2017.1397904
5. A. M. Carvalho, P. Sampaio, E. Rebentisch, J. Oehmen, and A. Trigo, "An integrative framework for quality management and Industry 4.0," Total Quality Management & Business

Excellence, vol. 31, no. 13–14, pp. 1567–1588, 2020. https://doi.org/10.1080/14783363.2018.1517854

6. T. H. Elsharnouby and A. A. Mahrous, "Customer participation in digital services," Journal of Service Theory and Practice, vol. 31, no. 2, pp. 187–210, 2021 https://doi.org/10.1108/JSTP-09-2020-0201

7. J. R. Evans and W. M. Lindsay, Managing for quality and performance excellence, 11th ed., Boston, MA: Cengage Learning, 2020. https://www.cengage.com/9780357442036

8. M. Hussain, A. Haleem, and S. Luthra, "Industry 4.0 and human resource management," International Journal of Production Research, vol. 59, no. 16, pp. 4804–4820, 2021. https://doi.org/10.1080/00207543.2020.1843714

9. [9] S. S. Kamble, A. Gunasekaran, and S. A. Gawankar, "Sustainable Industry 4.0 framework," Journal of Cleaner Production, vol. 197, pp. 41–54, 2018. https://doi.org/10.1016/j.jclepro.2018.06.202

10. R. R. Kumar, L. S. Ganesh, and C. Rajendran, "Quality 4.0 – a review and framework for quality management in the digital era," International Journal of Quality & Reliability Management, vol. 39, no. 6, 2022. https://doi.org/10.1108/IJQRM-06-2021-0157

11. H. C. Liu, R. Liu, X. Gu, et al., "From total quality management to Quality 4.0: A systematic literature review and future research agenda," Frontiers of Engineering Management, vol. 10, pp. 191–205, 2023. https://doi.org/10.1007/s42524-023-00264-3

12. H. Liu, Y. Li, and Y. Chen, "Digital Transformation and Quality Management in Higher Education Institutions," Journal of Educational Administration, vol. 59, no. 6, pp. 723–741, 2021. https://doi.org/10.1108/JEA-09-2020-0207

13. [13] C. Marnewick and A. Marnewick, "Digital transformation and skills development," South African Journal of Industrial Engineering, vol. 31, no. 3, pp. 1–14, 2020. https://doi.org/10.7166/31-3-2372

14. R. Pillai and B. Sivathanu, "Adoption of AI-based systems in higher education," Education and Information Technologies, vol. 25, pp. 1–23, 2020. https://doi.org/10.1007/s10639-020-10113-4

15. N. M. Radziwill, "Quality 4.0: Let's get digital – The many ways the fourth industrial revolution is reshaping the way we think about quality," arXiv preprint, 2018 https://arxiv.org/abs/1801.07584

16. M. S. Ramírez-Montoya, "Innovation and sustainability in higher education," Sustainability, vol. 12, no. 18 7202, 2020. https://doi.org/10.3390/su12187202

17. A. Saihi, M. Awad, and M. Ben Daya, "Quality 4.0: Leveraging Industry 4.0 technologies to improve quality management practices – a systematic review," International Journal of Quality & Reliability Management, vol. 40, no. 2, 2023. https://doi.org/10.1108/IJQRM-01-2022-0012

18. S. V. Shet, V. Pereira, and S. Shetty, "Quality management in the era of Industry 4.0," The TQM Journal, vol. 33, no. 6, pp. 1359–1376, 2021. https://doi.org/10.1108/TQM-04-2020-0096

19. M. Sony and S. Naik, "Critical factors for the successful implementation of Industry 4.0: A review and future research direction," Production Planning & Control, vol. 30, no. 3, pp. 229–247, 2019. https://doi.org/10.1080/09537287.2018.1492708

20. M. Sony, J. Antony, and J. A. Douglas, "Essential ingredients for the implementation of Quality 4.0," The TQM Journal, vol. 32, no. 4, pp. 779–793, 2020. https://doi.org/10.1108/TQM-03-2020-0060

21. A. Zonnenshain and R. S. Kenett, "Quality 4.0: A framework for integrating quality management in the digital era," Quality Engineering, vol. 32, no. 1, pp. 1–14, 2020. https://doi.org/10.1080/08982112.2019.1672881

22. A. Zulqarnain, M. Wasif, and S. A. Iqbal, "Developing a Quality 4.0 implementation framework and evaluating the maturity levels of industries in developing countries," Sustainability, vol. 14, no. 18, 2022. https://doi.org/10.3390/su141811456

23. J. K. Alsadi, F. Alkhatib, J. Antony, J. A. Garza-Reyes, G. Tortorella, and E. A. Cudney, "A systematic literature review with bibliometric analysis of Quality 4.0," *TQM Journal*, vol. 37, no. 5, pp. 1446–1470, 2024. https://www.ingentaconnect.com/content/10.1108/TQM-02-2024-0050

Aims and Objectives

Published online by Institute of Cited Scientists, Cyprus, two times a year, Journal of Digital Science (JDS) is an international peer-reviewed journal which aims at the latest ideas, innovations, trends, experiences and concerns in the field of digital science covering all areas of the scholarly literature of the sciences, social sciences and arts & humanities.

The principal ambition of this periodical is the efficacious propagation of original insights and outcomes derived from human cerebral activity and exemplified in scholarly treatises through the utilisation of contemporary information and digital technology.

The main topics currently covered include: Artificial Intelligence in Cybersecurity and Decision-making process, Cryptocurrency as Digital Assets, Digital decoding of language, Digital Technology in Health Care.

Contact information

The picture on JDS cover was generated by Andrei Stepanov.