Scholar J https://doi.org/10.5281/zenodo.20448700

Review Article

Evolution and Real-world Applications of Generative Artificial Intelligence and Data Mining in a Digital World

Uriah Auzhioluka Nwocha and Musa Ndagie Muhammad

Department of Computer Science,

Federal University of Education, Kontagora, Nigeria.

Received 12 April 2026; Acceptance 20 May 2026; Published 29 May 2026.

Abstract

By learning from patterns in vast datasets using models like GPT (for text), DALL·E (for images), and Codex (for code), generative artificial intelligence (Gen-AI) has become a veritable tool in data mining. This paper provides a detailed analysis of the evolution and creative impact of data mining in Gen-AI. It traces the origin of AI to the 1956 Dartmouth Conference, highlighting the subsequent advancements in technologies such as machine learning and data mining that have progressively sped-up the digital growth of AI. The paper also highlights some real-world applications of data mining in Gen-AI such as healthcare, finance, manufacturing, social media, retail and cybersecurity. It reveals how data mining powers Gen-AI and how Gen-AI enhances data mining. Owing to the fact that Gen-AI relies on data mining to learn patterns and generate new content, the paper raises several concerns on the relationship such as bias and discrimination, data privacy and security and intellectual property and copyright. The paper also illustrates a design for Gen-AI and data mining use case tailored for field of computer science.

Key words: Machine Learning, Generative-Artificial Intelligence, Data and Data Mining.

Introduction

Generative AI (GenAI) and data mining are two powerful technologies that complement each other in transformative ways [1-2]. Firstly, it is essential to understand the concepts of AI and ML. John McCarthy, who is considered the father of AI, describes it as the science and engineering involved in creating intelligent machines, particularly intelligent computer programs [3-4. In general, AI involves transforming a computer into a robot or using programming codes to make computers think and act with human-like intelligence.

Correspondence addressed to Uriah Auzhioluka Nwocha, email:nwochaau@yahoo.com

Russell and Norvig in [5] averred that AI involves creating computer software and systems that can replicate human behavior and problem-solving abilities by studying human thinking patterns, learning methods, and cognitive capabilities. In essence, Artificial Intelligence is enabling computers to develop both intellectual and emotional capabilities [6-7].

While AI and ML are often used as synonyms, they actually represent different concepts. A common belief is that the basis of machine learning is the idea that machines have the ability to learn. Historically, machine learning and Artificial Intelligence have been in existence for over fifty years, as indicated by their fundamental principles [8]. In 1959, IBM scientist Arthur Samuel was the first to utilize Machine Learning by creating a solution for a checkers game and publishing it. He described how, for the first time, a computer was able to play checkers against humans and emerge victorious. Over time, programmers have developed increasingly advanced systems that allow machines to perform tasks typically done by humans. Another well-known example is the ancient board game "Go," which has existed for over 2500 years. It is more complex and strategic than chess, leading to the belief that no computer could surpass a human in the game of "Go" Four years ago, the long-held belief was shattered when a computer program called AlphaGo defeated an 18-time world champion in a decisive 4 to 1 victory [9-11]. Machine learning has sped up the transformation of banking operations. Machine learning in this area uses past data and behaviors to forecast trends and assist in making decisions [12].

In art and design, AI-generated paintings, fashion prototypes and architectural renderings exist; in medical Imaging, creating synthetic scans to train diagnostic models without compromising patient privacy is achieved; Gen-AI is used in gaming and animation for Generating textures, characters, and environments. Advanced models like StyleGAN and CycleGAN have pushed the boundaries of realism and control in image generation. Transformers such as GPT and BERT are used for powering natural language generation and understanding. Behind the latest breakthroughs in high-fidelity image generation are the diffusion models. Generative AI doesn't just perform the task of analysis, it also creates, making it a leap forward from traditional AI.

Generative artificial intelligence (generative AI) has sparked significant concerns and challenges across various domains. These apprehensions have prompted protests, legal actions, and calls for the pause of AI experiments, leading multiple governments to take regulatory actions. Secretary-General António Guterres highlighted the dual potential of generative AI in a July 2023 United Nations Security Council briefing, acknowledg ing its enormous capacity for both positive and negative impacts on a global scale. He emphasized the potential for AI to contribute trillions to the global economy by 2030 but warned of catastrophic consequences if misused. One major concern revolves around job losses [13]. The rise of generative AI, particularly in image generation, has led to significant unemployment in certain sectors. For instance, in China, 70% of jobs for video game illustrators were reportedly lost due to image generation AI. The 2023 Hol lywood labor disputes also saw generative AI contributing to concerns, with industry f igures expressing fears that artificial intelligence poses an existential threat to creative professions, impacting jobs in areas such as voice acting and video game illustration [14-15].

Clarification of conceptual terms

Machine Learning

Machine learning is a subfield of artificial intelligence that enables computers to learn from data and improve their performance over time without being explicitly programmed. Machine learning (ML) involves designing algorithms that can identify patterns, make predictions, or take actions based on data. According to [16], instead of following fixed instructions, ML systems adapt and evolve as they process more information. It is widely used in applications such as spam detection, recommendation systems, image recognition, and autonomous vehicles. The three main types of machine learning are supervised learning - models learn from labeled data; unsupervised learning -models find hidden patterns in unlabeled data; and reinforcement learning - agents learn by interacting with an environment and receiving feedback.

Artificial Intelligence

Artificial Intelligence refers to the development of computer programs and systems that emulate human behavior and decision-making by analyzing cognitive processes, learning strategies, and thought patterns. Analytics Insight in [17] opined that, fundamentally, AI aims to equip machines with both rational and affective capabilities, allowing them to perform tasks that require intellectual and emotional intelligence.

Generative Atificial Intelligence (Gen-AI)

Generative AI refers to models that can create new content—text, images, audio, code, and more—based on learned patterns. These models include Generative Adversarial Networks (GANs), Transformers (e.g., GPT, BERT) and Diffusion Models. GANs have become a cornerstone of modern AI, especially in image synthesis and the creation of deepfakes. Introduced by Ian Goodfellow in 2014, GANs consist of two neural networks – Generator and Discriminator. Generator creates fake data (e.g., images) while discriminator evaluates whether the data is real or fake. They're trained in a zero-sum game where the generator tries to fool the discriminator, and the discriminator tries to catch the generator [18]. Over time, the generator becomes incredibly good at producing realistic outputs. GANs are widely used to generate high-resolution, photorealistic images from noise or structured input.

Data Mining

Data mining is the process of discovering patterns, correlations, and anomalies in large datasets using techniques like, Classification & Clustering, Association Rule Learning, Anomaly Detection and Sequential Pattern Mining. Fayyad in [19] emphasised that the data is normally accessed from one or more databases. The technique is known as knowledge discovery in databases (KDD). It's the backbone of predictive analytics, helping businesses and researchers make data-driven decisions.

Intersection of Gen-AI and Data Mining

Gen-AI and data mining are increasingly intertwined in training data, synthetic data generation and pattern discovery. In training data, data mining helps curate and preprocess massive datasets used to train generative models. During synthetic data generation, Gen-AI creates synthetic datasets to augment real data, especially in privacy-sensitive fields like healthcare. In pattern discovery, generative models can simulate scenarios or generate hypotheses that data mining then tests and validates.

Historical Evolution of Data Mining in AI

Data mining has been a cornerstone of artificial intelligence since the mid-20th century. Its evolution can be traced through several key phases, according to [20-22]:

- *Early AI Foundations (1950s–1980s):* The Dartmouth Conference in 1956 marked the birth of AI, but data mining remained rudimentary, limited to rule-based systems and symbolic reasoning. Early expert systems used structured databases but lacked the ability to extract patterns autonomously.
- *Rise of Machine Learning (1990s–2000s):* Algorithms like decision trees, clustering, and neural networks enabled machines to learn from data. At this point, data mining became central to training models for classification, prediction, and pattern recognition.
- *Big Data Era (2010s):* The explosion of digital data led to scalable data mining techniques using distributed computing (e.g., Hadoop, Spark). Deep learning emerged, allowing for more nuanced feature extraction from unstructured data like images and text.
- *Generative AI Emergence (2020s–Present):* Gen-AI models like GPT, DALL·E, and Stable Diffusion rely heavily on large-scale data mining to learn linguistic, visual, and multimodal patterns. Text and data mining (TDM) became a legal and ethical focal point due to copyright concerns over training data.

Creative Impact of Data Mining in Gen-AI

Data mining has profoundly influenced the creative capabilities of Gen-AI across multiple domains:

a. *Automated Content Generation:* Gen-AI models mine vast corpora to generate human-like text, music, and art. This has democratized creativity, enabling non-experts to produce high-quality outputs.
b. *Personalized Creativity:* Data mining enables Gen-AI to tailor outputs to individual preferences, enhancing user engagement in applications like chatbots, design tools, and storytelling platforms.
c. *Cross-Modal Innovation:* Mining multimodal datasets allows Gen-AI to generate content that blends text, image, and audio—e.g., generating a video from a script or an illustration from a poem.
d. *Creative Industries Transformation:* Advertising, fashion, gaming, and film industries use Gen-AI for ideation, prototyping, and even final production. This has shifted creative workflows from manual to hybrid human-AI collaboration.

e. *Challenges and Ethical Considerations:* In Copyright and Data Ownership**,** the legality of mining copyrighted content for training Gen-AI remains contested, especially in jurisdictions with strict database rights. While considering Bias and Representation**,** data mining can perpetuate societal biases if training data is unbalanced or discriminatory. On transparency and explainability considerations**, t**he complexity of mined data and model architectures makes it difficult to trace how Gen-AI arrives at specific outputs [23-24].

Real-World Applications of Generative AI

Generative AI is no longer just a lab curiosity. Its applications are ubiquitous. In healthcare, Gen-AI models generate molecular structures with desired properties in drug discovery. It enhances resolution and fills in missing data for diagnostics in medical imaging. In creative industries, tools like DALL·E and RunwayML help artists generate visuals from prompts. In music composition, AI can compose melodies, harmonies, and even full tracks. Education and research apply Gen-AI in Tutoring systems. Personalized learning experiences using AI-generated explanations. In research assistance, it is used in summarizing papers, generating hypotheses and automating literature reviews [16,25].

In business and productivity**,** blog posts, marketing copy, and reports generated at scale are used for content creation. Whereas, customer service utilizes chatbots and virtual assistants powered by LLMs. In Engineering and Software Development, tools like GitHub Copilot assist developers by generating code snippets. To ease simulation and modelling, AI generates synthetic data for testing and training. In Gaming and Virtual Worlds, levels, characters, and narratives are created dynamically aiding procedural content generation [23]. In Non-Player Characters (NPC) dialogue - conversations or interactions between non-player characters and players; they are characters controlled by the game or virtual world, rather than by a human player ensuring more natural and varied interactions in games. NPC dialogue can include storytelling, quests and missions, information and guidance, trading and commerce and role-playing. In Media and Journalism, Gen-AI is used to facilitate sports scores, financial summaries and weather updates, using automated reporting. In deepfake detection, ironically, generative AI also helps detect manipulated content [5].

Real-World Applications of Generative AI in Computer Science

Real world application of Generative AI in computer Science is in software development, Datat Science and machine learning, Cybersecurity, Human-Computer Interaction, Computer Vision, Cloud computing and DevOps [7,26,27].

a. **Software Development:** In code generation, tools like GitHub Copilot and Amazon Code-Whisperer assist developers by generating code snippets, documentation, and even entire functions based on natural language prompts. For bug detection and resolution, AI models can identify vulnerabilities and suggest fixes by learning from vast code repositories.

b. **Data Science & Machine Learning:** In synthetic data generation, GANs and VAEs produce realistic datasets for training models when real data is scarce or sensitive. In feature engineering, generative models help automate the creation of meaningful features for predictive modelling.

c. **Cybersecurity:** In threat simulation, AI generates attack scenarios to test system defences. During anomaly detection, generative models learn normal behaviour patterns and flag deviations as potential threats.

d. **Human-Computer Interaction:** In Natural language interfaces, Large Language Models (LLMs) - advanced AI systems designed to understand and generate human-like language using deep learning techniques, enable conversational agents that understand and respond to human queries with context-aware precision. Used in personalized user experiences, AI adapts interfaces and content based on user behaviour and preferences.

e. **Computer Vision:** During image enhancement, generative models improve resolution, fill missing data, and simulate realistic environments. In object recognition and segmentation, AI generates labelled data to train vision models more efficiently.

f. **Cloud Computing & DevOps:** In Infrastructure as code, AI can generate deployment scripts and automate cloud configurations. Monitoring and alerting, generative models predict system failures and recommend preventive actions.

A model for Gen-AI and data mining use case tailored for field of computer science

A powerful use case combining Generative AI and data mining in computer science is the creation of an intelligent academic research assistant that automates literature review, trend analysis, and hypothesis generation. This system leverages Gen-AI to synthesize knowledge and data mining to extract patterns from large datasets, streamlining research workflows for students and scientists.

Use Case: Intelligent Research Assistant for Computer Science

The specific objective is to assist researchers in identifying emerging trends, generating summaries, and proposing novel research directions by mining academic databases and generating insights using Gen-AI.

System Design Overview

a. Sources of Data

- *Academic databases*: IEEE Xplore, ACM Digital Library, arXiv
- *Citation networks*: Google Scholar, Semantic Scholar
- *Code repositories:* GitHub, GitLab

b. Data Mining Layer

- *Text mining:* Extract keywords, topics, and sentiment from abstracts and papers
- *Clustering:* Group papers by themes (e.g., deep learning, cybersecurity)
- *Trend analysis:* Identify rising topics using time-series analysis
- *Association rule mining:* Discover co-occurring concepts (e.g., blockchain + privacy)

c. Generative AI Layer

- *Summarization:* Generate concise summaries of papers or clusters
- *Question generation:* Create research questions based on gaps in literature
- *Hypothesis suggestion:* Propose testable ideas based on mined patterns
- *Code synthesis:* Generate prototype code snippets for proposed methods

d. User Interface

- *Search dashboard*: Query by topic, author, or keyword
- *Visualization tools:* Graphs of topic evolution, citation networks
- *Interactive assistant:* Chat-based interface for Q&A and suggestions

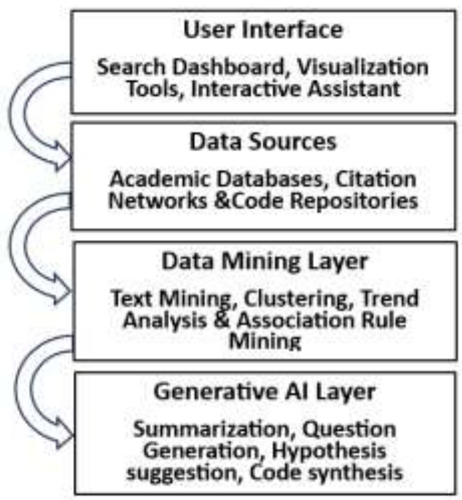

Figure 1. Model for Gen-AI and Data mining use case for computer science

In figure 1, the model is designed to assist researchers in identifying emerging trends, generating summaries, and proposing novel research directions by mining academic databases and generating insights using Gen-AI.

Technologies Involved

Each of the layers in the design model has tools/technologies used for optimal performance.

Table 1. Technologies Involved

Table 1 shows the technologies involved with respect to data mining, Gen-AI, Backend and Frontend layers.

Conclusion

In conclusion, Generative AI (Gen-AI) and data mining have become transformative technologies driving innovation, productivity, and intelligent decision-making across diverse sectors including healthcare, education, business, media, and computer science. Their ability to analyze vast datasets and generate personalized, multimodal content has enhanced creativity, automated complex processes, and improved research and learning outcomes. However, concerns relating to ethics, transparency, bias, privacy, and copyright remain significant challenges that require urgent attention. Therefore, governments should establish adaptive regulatory policies and support public-sector innovation, while industry stakeholders invest in secure and scalable infrastructure as well as collaborative innovation platforms. Academic institutions should integrate Gen-AI and data mining into interdisciplinary curricula, and all sectors should promote inclusive, multilingual, and culturally adaptive AI systems to ensure equitable and responsible digital transformation worldwide.

References

1. Sharma, R., & Patel, K. (2024). The convergence of big data and generative AI: A paradigm shift in data analytics. *International Journal of Future Management Research*, 6(1), 42–55. https://www.ijfmr.com/papers/2024/6/32142.pdf
2. O'Connor, B. (2024). Copyright, text & data mining and the innovation dimension of generative AI. *Journal of Intellectual Property Law & Practice*, 19(7), 557–566. https://academic.oup.com/jiplp/article/19/7/557/7624901
3. Kumar, V., & Chandrakala, D. R. (2016). *Artificial intelligence*. Nalanda Publishing House.
4. Taher-Uz-Zaman, M., & Rahman, M. M. (2024). Generative models for synthetic data: Transforming data mining in the age of AI. *arXiv preprint arXiv:2508.19570*. https://arxiv.org/pdf/2508.19570
5. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
6. Tretter, M. (2024). Equipping AI decision-support systems with emotional capabilities? Ethical perspectives. *Frontiers in Artificial Intelligence*, 7, Article 1398395. https://doi.org/10.3389/frai.2024.1398395

7. Alloghani, M., Thron, C., & Subair, S. (2023). Cognitive computing, emotional intelligence, and artificial intelligence in healthcare. In *Artificial Intelligence for Data Science in Theory and Practice* (pp. 109–118). Springer. https://link.springer.com/chapter/10.1007/978-3-030-92245-0_5

8. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (3rd ed.). O'Reilly Media.

9. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359. https://doi.org/10.1038/nature24270

10. Deloitte. (2022). *AI and machine learning in banking: Transforming operations and customer experience*. Retrieved from https://www2.deloitte.com

11. DeepMind. (2023). *AlphaGo: The story so far*. A current overview of AlphaGo's development and its influence on AI research. https://deepmind.google/alphago

12. Open MedScience (2025, January 28). *How AI is improving medevac response times and Patient care*. https://openmedscience.com/how-ai-is-improving-medevac-response-times-and-patient-care/

13. United Nations (2023, July 18). *Secretary-General's remarks to the Security Council on Artificial Intelligence*.https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence

14. Karpathy, A.; Abbeel, P.; Brockman, G.; Chen, P.; Cheung, V., Duan, Y.; Goodfellow, I.; Kingma, D.; Ho, J.; Rein, H.; Tim, S.; John, S.; Ilya, S. & Wojciech, Z. (2016). Generative Models: A Branch of Unsupervised Learning Techniques in Machine Learning. Retrieved from https://openai.com/research/generative-models.

15. Metz C. (2023). Open AI plans to up the ante in tech's AI race. The New York Times, New York

16. GeeksforGeeks. (2025, July 23). *Generative AI applications*. https://www.geeksforgeeks.org/artificial-intelligence/generative-ai-applications/

17. Analytics Insight (2025, August 19). *AI and emotional intelligence: Can machines truly Understand us?* https://www.analyticsinsight.net/artificial-intelligence/ai-and-emotional-intelligence-can-machines-truly-understand-us

18. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial nets*. In *Advances in neural information processing systems* (Vol. 27). https://papers.nips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

19. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From data mining to knowledge*

discovery in databases. AI Magazine, 17(3), 37–54. https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230

20. Chooramun, R. (2025). Evolution of data mining: From statistical foundations to big data and deep learning. *International Journal of Computer Applications, 187*(4), 12–20. https://www.ijcaonline.org/archives/volume187/number4/evolution-of-data-mining-from-statistical-foundations-to-big-data-and-deep-learning/
21. Trigka, M., & Dritsas, E. (2025). The evolution of generative AI: Trends and applications. *IEEE Access, 13*, 98504–98529.https://doi.org/10.1109/ACCESS.2025.3574660
22. Li, D., Huang, Y., Li, M., Zhou, T., Zhang, X., & Liu, H. (2025). Generative models for synthetic data: Transforming data mining in the GenAI era. *arXiv preprint arXiv:2508.19570*. https://arxiv.org/abs/2508.19570
23. Oloruntoba, O. (2024). Generative AI for creative data management: Optimizing database systems in the creative industry. *Iconic Research and Engineering Journals, 7*(7), 588–597. https://www.irejournals.com/paper-details/1707435
24. He, R., Cao, J., & Tan, T. (2025). Generative artificial intelligence: A historical perspective. *National Science Review, 12*(5), nwaf050. https://doi.org/10.1093/nsr/nwaf050
25. Salloum, S. A., Al-Marzouqi, A., Gaber, T., Masa'deh, R., & Shaalan, K. (2025). Foundations and frontiers: The evolution and impact of generative AI technologies. In *Generative AI in Creative Industries* (pp. 3–12). Springer. https://link.springer.com/chapter/10.1007/978-3-031-89175-5_1
26. Tyagi, K. (2024). Copyright, text & data mining and the innovation dimension of generative AI. *Journal of Intellectual Property Law & Practice*, 19(7), 557–570. https://doi.org/10.1093/jiplp/jpae028
27. Wu, Z. (2025). Data mining in AI: Evolution, applications, and future directions. *ACE Proceedings*.https://www.ewadirect.com/proceedings/ace/article/view/16882/pdf

Publisher's note Scholar J remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.