Artificial Intelligence Review (2026) 59:7
https://doi.org/10.1007/s10462-025-11435-z

A systematic review of generative AI: importance of industry and startup-centered perspectives, agentic AI, ethical considerations & challenges, and future directions

Kinjal Patel[1] · Milind Shah[2] · Karishma M. Qureshi[3] · Mohamed Rafik N. Qureshi[4]

Received: 5 July 2025 / Accepted: 19 October 2025 / Published online: 12 November 2025

Abstract

Generative Artificial Intelligence (GenAI) is rapidly redefining the landscape of work organizations and society at large. GenAI has rapidly evolved from rule-based symbolic systems ofThe 1940 s to advanced deep learning architectures capable of producing human-like content across modalities, such as text, images, audio, and video. This review focuses on current emerging trends, such as large concept models and critical comparisons of tools, including ChatGPT, Gemini, and Claude. This study synthesizes evidence of GenAI's essential role across major industries, revealing transformative applications in the finance, cloud and IT, healthcare, education, and energy sectors. The paper also highlights the unique opportunities GenAI offers for start-ups, enabling agile projects to leverage cutting-edge technology for competitive advantage. However, the deployment of GenAI systems through edge devices also raises critical challenges related to ethics, transparency, bias, accountability, computational issues, and many more. To address these complexities, this paper examines emerging approaches such as AI agents, agentic AI, and multi-agent systems that aim to extend the functionality of GenAI through autonomy, goal-directed behavior, and collaborative intelligence. It discovers novel incorporations with agentic AI architecture, such as BabyAGI, and discusses emerging issues of coordination, hallucination, and security risks. The findings reveal persistent challenges related to scalability, interpretability, and regulatory compliance while identifying future research directions toward developing more sophisticated, ethical, and accessible GenAI systems that will continue to reshape technological landscapes and societal interactions. This systematic review informs researchers, academicians, data scientists, and developers about the latest advancements in GenAI and highlights its applications and role across various industries, as well as supporting practitioners and scholars in staying current with the rapidly evolving landscape of generative technologies.

Keywords Artificial intelligence · Generative artificial intelligence · Large language model · Large concept model · AI agents · Agentic AI · Multi agent systems · Industrial GenAI · GenAI evolution · Ethical frameworks · Compliance

Extended author information available on the last page of the article

Abbreviations

GenAI	Generative Artificial Intelligence
IT	Information Technology
AI	Artificial Intelligence
ChatGPT	Chat Generative Pretrained Transformer
Web3	Web 3.0 (Third generation of the web with decentralization)
DALL-E	Named after Salvador Dalí and Pixar's WALL-E; a generative model for images
CAGR	compound annual growth rate
USD	United States Dollar
GAI	General Artificial Intelligence
GANs	Generative adversarial networks
arXiv	Archive (pronounced "archive"; e-print repository for scientific papers)
BioXiv	Biology Archive (preprint server for biology)
e.g.	Exempli Gratia (for example)
PubMed	Public/Publisher Medline (Database of biomedical literature)
IEEE	Institute of Electrical and Electronics Engineers
VAE	Variational Autoencoders
NLP	Natural Language Processing
3D	3-Dimensional
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
ResNet	Residual Network
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
AlexNet	Neural Network architecture named after Alex Krizhevsky
APIs	Application Programming Interfaces
RLHF	Reinforcement learning from human feedback
OpenAI	Open Artificial Intelligence (Organization)
Fig	Figure
DM	Diffusion Models
TRMs	Transformer-based Models
ICO	Information Commissioner Office
UK	United Kingdom
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
GDPR	General Data Protection Regulation
LCMs	Large Concept Models
LLM	Large Language Models
GPT	Generative pretrained transformer
OASIS	Outage Assessment and Summarization using the Intelligent Systems
BERT	Bidirectional Encoder Representations from Transformers
GENTRL	Generative Tensorial Reinforcement Learning
MedTech	Medical Technology
MedVQA	Medical Visual Question Answering

Llama	Large Language Model Meta AI
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
X-rays	X-radiation
EHR	Electronic Health Record
RCM	Revenue Cycle Management
VQA-RAD	Visual Question Answering – Radiology Dataset
PathVQA	Pathology Visual Question Answering Dataset
MedVQA-2019	2019 version of Medical VQA challenge dataset
ImageCLEF-VQA-Med	ImageCLEF Medical Visual Question Answering Challenge
ViT	Vision Transformer
CLIP	Contrastive Language–Image Pretraining
Llama	Large Language Model Meta AI
COVID-19	Coronavirus Disease 2019
FLAN-T5	Fine-tuned LAnguage Net T5 (Text-to-Text Transfer Transformer)
Med-BERT	Medical Bidirectional Encoder Representations from Transformers
MedFlamingo	Multimodal foundation model for medical question answering
K-12	Kindergarten through 12th Grade (school education levels)
RL	Reinforcement Learning
DGM	Deep Generative Models
GenOps	Generative AI Operations
CPU	Central Processing Unit
GPU	Graphics Processing Unit
DRAM	Dynamic Random-Access Memory
IoT	Internet of Things
SPG	Solar power generation
MLOps	Machine Learning Operations
B	Billion
T	Trillion
IIoT	Industrial Internet of Things
UCAs	Unmanned Combat Aerial Systems
FinTech	Financial Technology
SEO	Search Engine Optimization
FAQs	Frequently Asked Questions
Tech	Technology
OT	Operational Technology
LangChain	Language Model + Chaining Framework (LLM orchestration framework)
C-level	Corporate Executive Level – C-suite
A/B Testing	Alpha/Beta Tesing
UI/UX User	Interface/User Experience
edX	Education Exchange
BERTscore	Bidirectional Auto Regressive Decoder
LaMDA	Language Model for Dialogue Applications

PaLM	Pathways Language Model
Q&A	Question and Answer
BLEU	Bilingual Evaluation Understudy
ROUGE	Recall Oriented Understudy for Gisting Evaluation
BERTScore	Bidirectional Encoder Representations Transformers Score
GB	Gigabyte
RAM	Random Access Memory
W	Watts
DAN	Do Anything Now
IPR	Intellectual Property Rights
XAI	Explainable Artificial Intelligence
MAS	Multiagent systems
BabyAGI	Baby Artificial General Intelligence
FAISS	Facebook AI Similarity Search
SMMEs	Small- and medium-sized enterprises
TPUs	Tensor processing units
RAG	Retrieval-Augmented Generation
FactCC	Fact Checking and Claim Classification
FRANK	Factuality and Reliability Assessment for Natural Language Knowledge

1 Introduction

A widespread perspective of artificial intelligence is that it enables prediction and optimization. However, the rapid expansion of Generative Artificial Intelligence (GenAI) has dramatically transformed the nature of work and organizations, opposing this view. The development of GenAI has, on the other hand, marked a significant leap beyond the capabilities of predictive AI, which has enabled primarily automation and analytical decision support. Early findings indicate that GenAI performance has reached or even exceeded human performance in tasks that involve creative thinking, emotional intelligence, and empathic communication, among other aspects (Krakowski 2025).

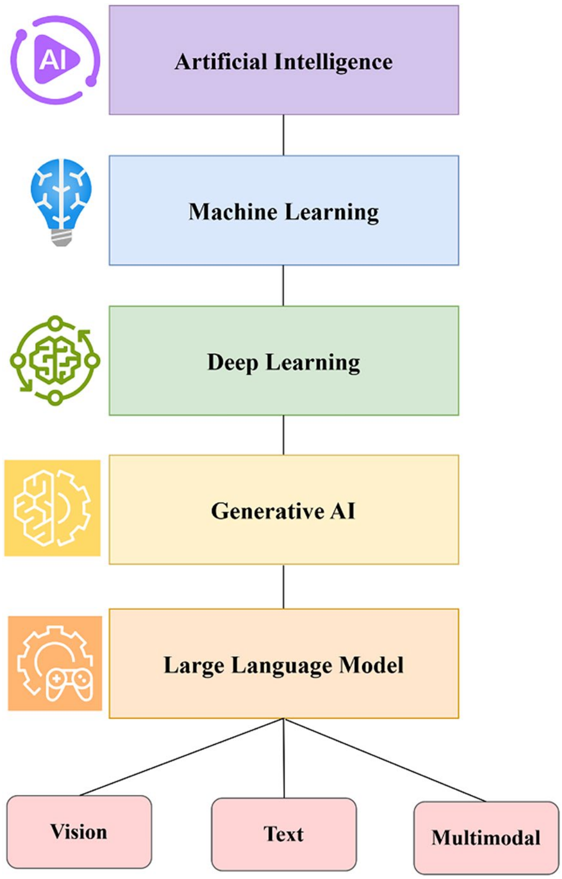
In the age of Web3, GenAI technologies such as the Chat Generative Pretrained Transformer (ChatGPT) have the potential to become productivity aids. These technologies can solve challenges related to digital assets and content production, as well as fill key gaps in the progression of Web3. It is believed that GenAI technologies will accelerate the arrival of the Web3 era by providing Web3 producers and contributors with productivity tools that are more dependable and easier. As a result of the emergence of GenAI tools such as ChatGPT, the industry's focus on the adaptability and creativity of these technologies has significantly increased. ChatGPT, which is based on deep learning models, can produce content in a broad variety of situations and satisfy a wide range of demands. With its assistance, the effectiveness and quality of content generation and distribution may be significantly enhanced (Lv 2023).

GenAI is a term that describes machinery that can generate new ideas and produce outputs that are comparable to human cognition. The modern era has begun with the development of this technology, which provides extraordinary learning capabilities and generates outcomes that are one of a kind. GenAI refers to artificial intelligence algorithms that can generate new content on demand. This includes graphics, text, conversation, and music, among other things, as shown in Fig. 1. GenAI uses deep learning models to filter quantities of unstructured information in search of patterns. It does this by simulating the way people think in regard to making decisions and finding solutions to challenges to provide innovative outcomes. As a result of recent developments, such as Gemini, ChatGPT, and DALL-E, GenAI has become an extremely popular area. The ability of GenAI to generate content, in contrast to the data interpretation-only methodologies utilized by traditional machine learning algorithms, makes it an essential motivator in several different industries (Mohawesh et al. 2025).

With a Compound Annual Growth Rate (CAGR) of 35.6% from 2023 to 2028, the global market for GenAI has the potential to reach USD 51.8 billion by 2028, as stated in research that was published not too long ago. The statistical data illustrates the considerable potential of GenAI as a game changer for businesses that want to plan routes, which will lead to exponential development and increased productivity (Birlasoft 2023).

InThe 1960 s, chatbots were the first to have general artificial intelligence (GAI). Nevertheless, it was not until 2014 that generative adversarial networks (GANs), algorithms for machine learning, were introduced. This was the first time that GenAI was able to produce

Fig. 1 AI with its advanced fields, including GenAI and LLM for image, text, and multimodal [Author's work]



realistic images, videos, and audio for users. The rise in popularity of GenAI may be linked, in large part, to recent technological advancements such as Transformers and the language models that they allowed. Researchers can train increasingly significant models with the use of transformers, which are machine learning tools. This is accomplished without the need to classify all of the data beforehand. Therefore, new models may be trained on many pages of text, which results in more complete responses (Sai et al. 2025).

GenAI models are sophisticated types of artificial intelligence that use advanced architectural components to produce new content on the basis of a large amount of data. Each of these models uses deep neural networks and layers aimed at identifying challenging patterns and making predictions. GANs compare a generator against a discriminator so that authenticity alone distinguishes new data it makes from actual data, leading to progressively more realistic outputs with practice. Transformer models have helped to take NLP a step further by facilitating fine-grained contextual comprehension and processing of long-range textures within textual data. Similar advances in other data modalities have been achieved on the basis primarily of breakthroughs in generative modeling. In image generation, diffusion models have proven to be the current state-of-the-art, allowing the generation of high-quality, realistic visual content and opening up more possibilities with respect to the creative abilities of generative AI systems (Bandi et al. 2023; Aggarwal et al. 2021; Aldausari et al. 2023).

Training for these models makes use of different learning techniques that match the data and applications needed. With supervised learning, data are provided in pairs with input and output labels, but unsupervised learning does not require labels by using models to identify reliable patterns in data by compressing and reconstructing them with autoencoders. GANs are driven by two components (called optimizers) that compete with each other so that each of the generators and the discriminator simultaneously enhances their performance. Fine-tuning trains pretrained models using new data to align them to the newly required needs of a task. This is how generative AI models can create content that matches the trends and patterns found in the data used during training, resulting in coherence and relevance across different purposes (Bandi et al. 2023). Figure 2 shows a visual representation of the advanced capabilities of GenAI, including self-learning algorithms, self-correction mechanisms, and automated decision-making for complex processes against traditional automation.

1.1 Research motivation

As various GenAI services, such as text, image, and video, have recently been launched, GenAI services are no longer limited to searching and providing information but are expanding into learning, problem-solving, and life support services. Hence, there is a need to expand the debate on the acceptance of GenAI technology and perform more focused research on the motivations, attitudes, and intentions of users utilizing GenAI as a service. The use of GenAI at the workplace has also transformed the way professionals spend their time at the workplace, as they can now create quality work in a short period. It can be a performance review draft, an idea generation session, or a marketing email composition session; the outcomes are more efficient and, in most cases, higher quality when humans work with GenAI. However, studies find that there is a hidden trade-off: Although GenAI cooperation improves prompt task execution, it may damage workers' intrinsic inspiration

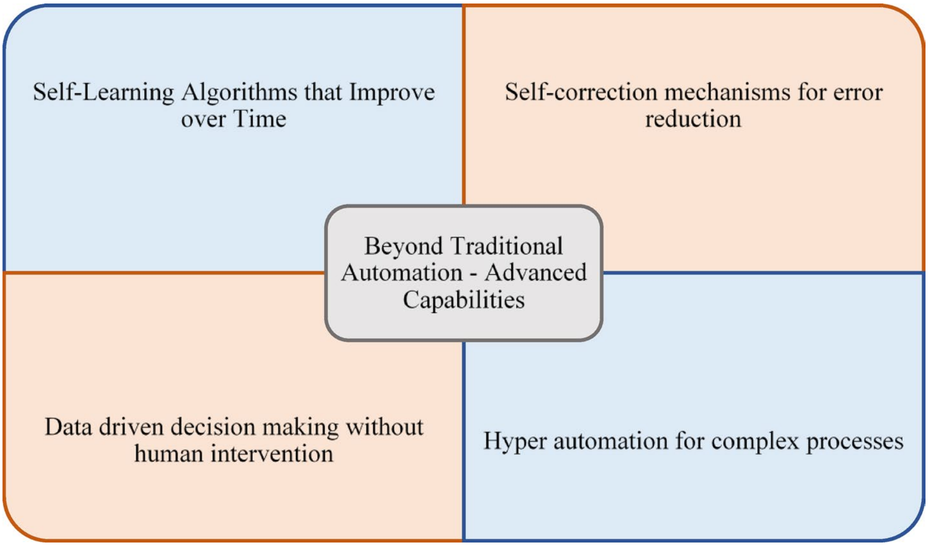


Fig. 2 GenAI advanced capabilities over traditional automation [Author's work]

and augment tediousness when they switch to activities where they lack this technological support.

This transition has fueled an urgent need to understand not only the usability of GenAI technologies but also the underlying motivations, anticipations, and psychological experiences of those who use them. These outcomes indicate that there are severe limitations to our current knowledge about user perception, trust, and GenAI tool integration into the workflow. To align the existing knowledge and clarify the research gaps, a systematic review is needed to guide the generation and implementation of more human-friendly GenAI mechanisms. Because these trends and trade-offs in combination highlight the necessity of a systematic overview of the current literature, a review is needed that not only provides a mapping of the functional advantages of GenAI but also critically reflects its expanded psychological, social, and organizational implications.

In the future, it is estimated that human–GenAI cooperation can significantly advance productivity and performance, yet organizations should be cautious of its psychological implications. Through careful consideration and the development of workflows that embrace GenAI, companies can reap the rewards of this technology without affecting the motivation and interest of workers. The future of work is not simply what AI is capable of after all but what people and AI can do together.

1.2 Intended audience

The intended audience for a systematic review of GenAI would be mainly researchers and academicians, as well as industry professionals and practitioners who are concerned with the applications and implications of technology. It will assist AI developers, data scientists, and machine learning engineers in applying generative models productively to a wide variety of fields. The review will also be useful to interdisciplinary scholars working in various spheres, including healthcare, digital media, NLP, and human–computer interaction, because it will offer a unified perspective on current developments, issues, and research opportunities. By revealing current shortcomings and future potential, this survey hopes to instruct the next generation of researchers to close existing gaps in knowledge and generate breakthroughs in GenAI. Those that work in areas where GenAI is being used, such as education, healthcare, finance, and creative industries, and those studying the wider implications of GenAI in society.

1.3 Inclusion and exclusion criteria

To conduct these systematic reviews, several research articles were analyzed and explored by considering certain standards and regulations, as discussed in Table 1.

1.4 Search strategy

The databases listed in Table 2 were used to find related journal papers and conference papers published by top peer-reviewed sources.

1.5 Article filtering

The selection process was the standard systematic review process. First, the records found in the chosen databases (see Table 2) were imported into a reference management tool to eliminate duplicates. The remaining articles passed through two stages of the screening process:

1. Title and Abstract Screening: Initially, articles were removed on the basis of title and abstract to exclude inappropriate articles, such as those not related to generative models or related to non-AI methods.
2. Full-Text Review: In the case of research that reached the second stage of selection, full texts were analyzed to determine their eligibility according to predetermined inclusion and exclusion criteria.

The inclusion criterion was that studies needed the following:

- Attention should be given to GenAI technologies, applications, and impacts.
- Offer original research with technical information regarding model architecture, training, or evaluation.
- Published in peer-reviewed journals or good-quality conferences.
- Written in English.

Exclusion criteria:

- Review articles, commentaries, or opinion papers that do not offer original contributions to a generative model.
- Research that lacks methodological details.
- Articles that have nothing to do with generative tasks or that concentrate only on dis-

Table 1 Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
This review includes peer-reviewed studies that propose, evaluate, or apply GenAI models such as Generative Adversarial Networks (GANs), Diffusion Models, or Transformer-based architectures for generating data across various modalities (e.g., images, text, audio, or multimodal content). Only studies that present original research with a clear focus on generative modeling techniques and their applications are considered.	Non-Scopus-indexed and Web of Science-indexed journals were eliminated from the review.
Peer-reviewed journal articles and conference papers	Non-peer-reviewed sources (e.g., blogs, opinion pieces, white papers without peer review)
Publications from 2018 to 2025	Publications before 2018 unless seminal foundational works
Articles focused on Generative Artificial Intelligence (GenAI), including LLMs, GANs, and related models	Articles not focused on GenAI or its applications
Studies addressing: evolution, technical background, applications across industry sectors (finance, healthcare, education, energy)	Studies addressing unrelated AI fields or general AI without GenAI focus
Research discussing implementation, deployment, cybersecurity risks, ethical considerations, and governance of GenAI	Studies lacking relevance to implementation, risks, or deployment aspects
Literature reviewing agentic AI, multi-agent systems, AI agents, and large concept models	Articles unrelated to agentic AI or multi-agent systems
Publications covering challenges, future directions, and innovations in GenAI	Articles without substantial empirical or theoretical contributions to GenAI
English language publications	Non-English language publications without available translations
Studies available in full text	Abstracts, conference posters, or incomplete reports
Research pertinent to startups, market impact, and innovation management of GenAI	Papers focused on unrelated business or management topics without AI focus

criminative models.

Figure 3 shows the PRISMA diagram with a graphic representation of the research article selection process for the systematic literature review of generative artificial intelligence (GenAI). Initially, data were obtained through a few databases and registries. After the elimination of duplicates and ineligible records through the use of automated tools or other tools,

Table 2 Search strategy including academic databases and relevant search terms

Database/Publisher	Search Date	Search Terms	Fields Searched	Limits/Filters
Scopus	2025-05-15	("Generative AI" OR "GenAI" OR "Large Language Models" OR "LLM" OR "AI agents" OR "Agentic AI") AND ("industry" OR "finance" OR "healthcare" OR "education" OR "energy" OR "cloud and IT support")	Title, Abstract, Keywords	English; 2018–2025; Peer-reviewed articles
PubMed/MEDLINE	2025-05-18	("Generative AI" OR "AI agents" OR "LLM") AND ("healthcare" OR "cybersecurity" OR "risk factors" OR "ethical implications")	Title, Abstract	English; 2018–2025; Peer-reviewed articles
IEEE Xplore	2025-05-22	("Generative AI" OR "Deep learning" OR "AI convergence" OR "Multi-agent systems") AND ("implementation" OR "deployment" OR "accuracy" OR "computational resources")	All fields	English; 2018–2025; Peer-reviewed articles
Elsevier ScienceDirect	2025-05-24	("Generative AI" OR "GenAI") AND ("industry" OR "implementation" OR "ethics" OR "AI agents")	Title, Abstract, Keywords	English; 2018–2025; Peer-reviewed articles
SpringerLink	2025-05-26	("Generative AI" OR "Large Language Models" OR "Cybersecurity") AND ("deployment" OR "risk factors" OR "ethical implications")	Title, Abstract	English; 2018–2025; Peer-reviewed articles
Wiley Online Library	2025-05-27	("Generative AI" OR "Agentic AI") AND ("computational resources" OR "accuracy" OR "multiagent systems")	Title, Abstract	English; 2018–2025; Peer-reviewed articles
Taylor & Francis Online	2025-06-29	("Generative AI" OR "GenAI") AND ("finance" OR "education" OR "healthcare")	Title, Abstract	English; 2018–2025; Peer-reviewed articles
MDPI Journals	2025-06-28	("Generative AI" OR "Ethical implications" OR "Challenges")	Title, Abstract	English; 2018–2025; Peer-reviewed articles
ACM Digital Library	2025-06-29	("Generative AI" OR "AI agents") AND ("multi-agent systems" OR "autonomous agents")	Title, Abstract	English; 2018–2025; Peer-reviewed articles

the titles and abstracts of the remaining records were screened. Records that were not used to meet the inclusion criteria at this level were discarded.

1.6 Paper organization

This paper consists of 8 sections, and each section discusses the corresponding research areas in the domain of GenAI. Figure 4 shows the organization of the paper, which consists of sections and subsections to discuss the systematic review of GenAI.

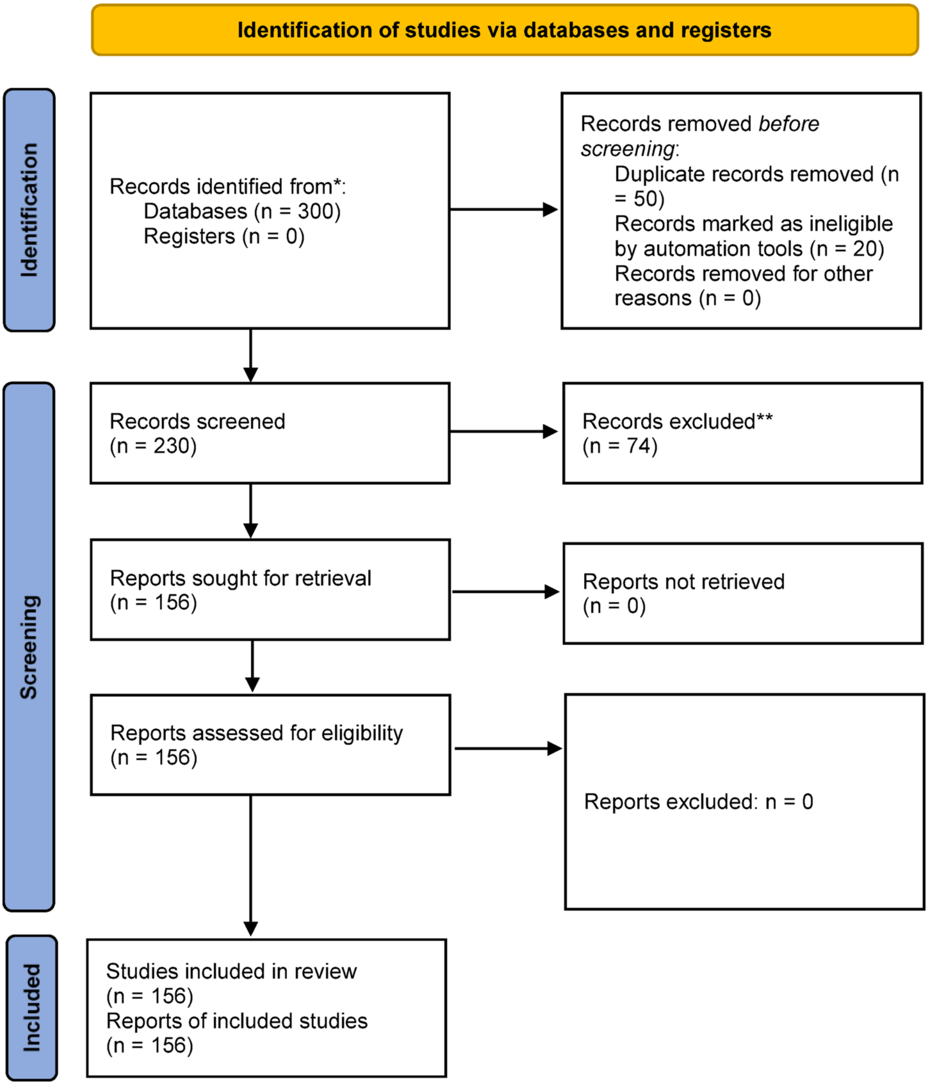


Fig. 3 PRISMA flow diagram illustrates the identification, screening, eligibility assessment, and inclusion of studies at each review stage

1.7 Paper contribution

This systematic review makes various significant contributions to the field of GenAI, with emerging trends in diverse applications.

- First, it provides an in-depth analysis of the rapid expansion of GenAI and the impact of GenAI in the current era, along with its applications and advanced capabilities.
- This paper discusses the evolution of GenAI from the beginning to each generation including knowledge-driven data-driven and the latest combination of both along with

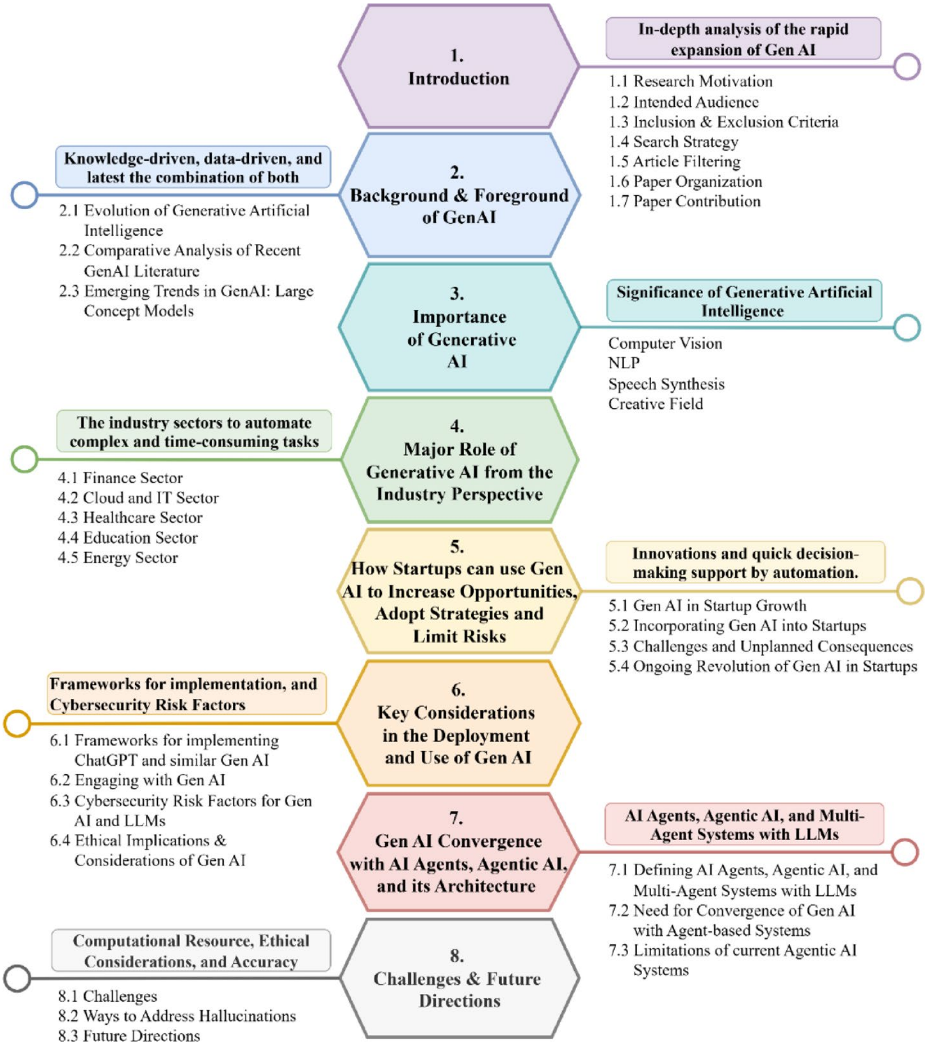


Fig. 4 Paper Workflow with organization of the paper, which consists of sections and subsections [Author's Work]

its evolution timeline fromThe 1940 s to the present. In addition, an in-depth comparative analysis of recent research and emerging trends in GenAI, including large concept models, is discussed.

- The importance of GenAI is discussed in detail with its core technologies, including computer vision, NLP, speech synthesis, and the creative field, along with its associated industries.
- This paper contributes to revealing the essential role of GenAI with major industry perspective applications for finance, cloud and IT, healthcare, education, and energy.
- This paper summarizes one of the most recent aspects of GenAI for start-ups, with its emerging capabilities, through the latest opportunities, strategies, and mitigation risks.

- The key considerations in the deployment and use of GenAI, along with critical comparisons of GenAI tools, performance measure evaluation of GenAI, and limitations of GenAI on deployment over edge devices, should be explored. In addition, its ethical implications, frameworks, and considerations are discussed.
- The incorporation of GenAI, along with its limitations and recent advancements, includes AI agents, agentic AI, and multiagent systems. It also discusses the Agentic AI architecture, BabyAGI, with its autonomous planning chain and limitations of existing Agentic AI systems.
- To describe the challenges associated with GenAI and ways to address hallucination, as well as potential future directions, this paper enhances all the views on GenAI technology advancement.

2 Background and foreground of GenAI

GenAI is a cutting-edge artificial intelligence technology that has recently attracted much interest and funding from large multinational companies. Its widespread use has resulted in the establishment of a variety of new businesses that are focused primarily on the advancement of general artificial intelligence (GAI) technology. GAI models can create a wide variety of media that may be generated in response to user queries.These types of media include text, images, audio, video, and 3 D models. To develop original and distinctive outputs that closely reflect the properties of the input training data, this cutting-edge technology makes use of the power of pattern recognition and learns from the data that are already available. GAI has quickly become one of the most sought-after technologies in the world, and its popularity has increased at a rapid pace.

A few examples of GAI systems that are particularly noteworthy are ChatGPT, Dall-E, Midjourney, and Bard. In the context of GenAI models, ChatGPT, which was created by OpenAI, is widely recognized for its NLP ability. It encourages people to participate in conversations that are logical and relevant to circumstances.

There is a fast-developing field of artificial intelligence known as GenAI, which focuses on the development of algorithms and methods that produce new and distinct information. In contrast to conventional artificial intelligence systems, which are intended to perform tasks such as data classification and prediction, GenAI is designed to produce data that reflect the properties of the dataset being trained on. Deep learning and neural network architecture have been the main reasons for substantial advancements in this area, which have undergone considerable advancements. Different approaches have been utilized in the field of GenAI. The following are examples of these: GANs, VAEs, and autoregressive models; however, the list is not comprehensive. GANs function via a combination of generator and discriminator networks that compete with one another. The generator is responsible for the generation of new data samples, whereas the discriminator is responsible for attempting to separate these samples from legitimate data. VAEs, on the other hand, function by first compressing the data that are entered into a latent space and then rebuilding the data to produce new instances. The data generated by autoregressive models are generated sequentially, with each new generation integrating the components that came before it in the series (Sai et al. 2024; Corvello 2025; Carlini 2024).

2.1 Evolution of generative artificial intelligence

Devices that function similarly to human thinking, such as learning, adapting, understanding, and solving problems, are generally recognized as AI (Kalota 2024). During the development of AI from initial to advanced it has advanced and introduced novel technologies. The field began in 1955 with the introduction of the phrase "AI" at a workshop proposal (Kaynak 2021). AI has evolved from its early days in theory to current practical innovations in machine learning, machine logic, artificial neural networks, and expert systems (Hirsch-Kreinsen 2024; Taherdoost and Madanchian 2024).

The progression of AI consists of three elements, which are termed symbolicism, connectionism, and actionism. These three distinct aspects have guided the development of AI (Zhang et al. 2023; Shao et al. 2022). As shown in Fig. 5, (1) element symbolicism represents symbolistic AI, also known as a knowledge-driven approach. (2) Element connectionism represents a data-driven approach, which is considered a foundation of deep learning. (3) Element actionism represents third-generation AI, which combines knowledge-driven and data-driven approaches into an interpretable robust theory.

The combination of these perspectives is most evident in recent large language models, which rely on recognizing patterns on the basis of methods of deep learning from the connectionist approach and symbols on the basis of knowledge-driven methods. Currently, because of hybrid evolution, AI systems can address difficult problems that humans use to address their cognitive skills alone. The last decade has fully developed people's knowledge of AI and is now generally known as the Fourth Industrial Revolution (Zhang et al. 2023; Shao et al. 2022). Unlike before, this revolution focused mostly on developing human thinking and decision-making in various parts of society.

Figure 6 shows the evolution of GenAI over the decades. Researchers in neurology reported that the brain is made up of neurons linked together as a network creating both pulse and nonpulse activity which interested scientists from mathematics psychology engi-

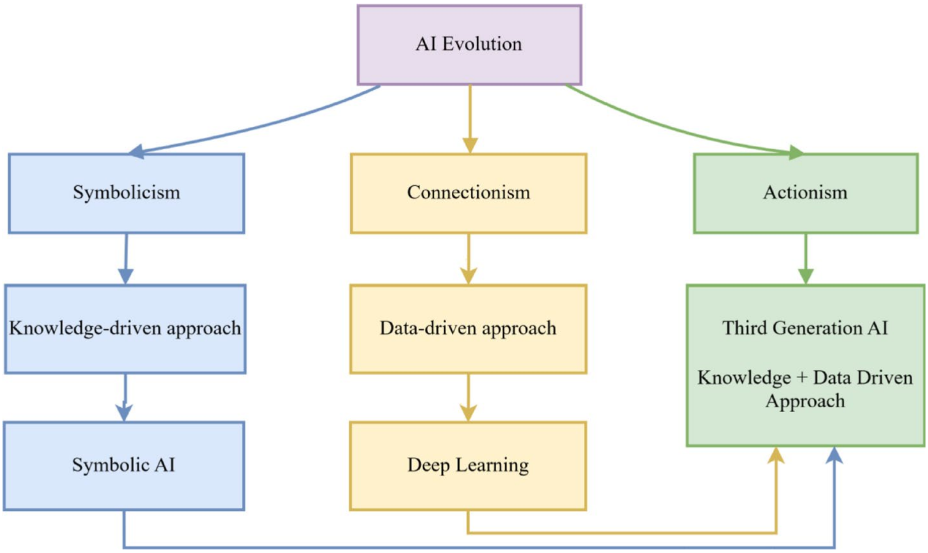


Fig. 5 AI evolution core elements with symbolicism, connectionism and actionism [Author's work]

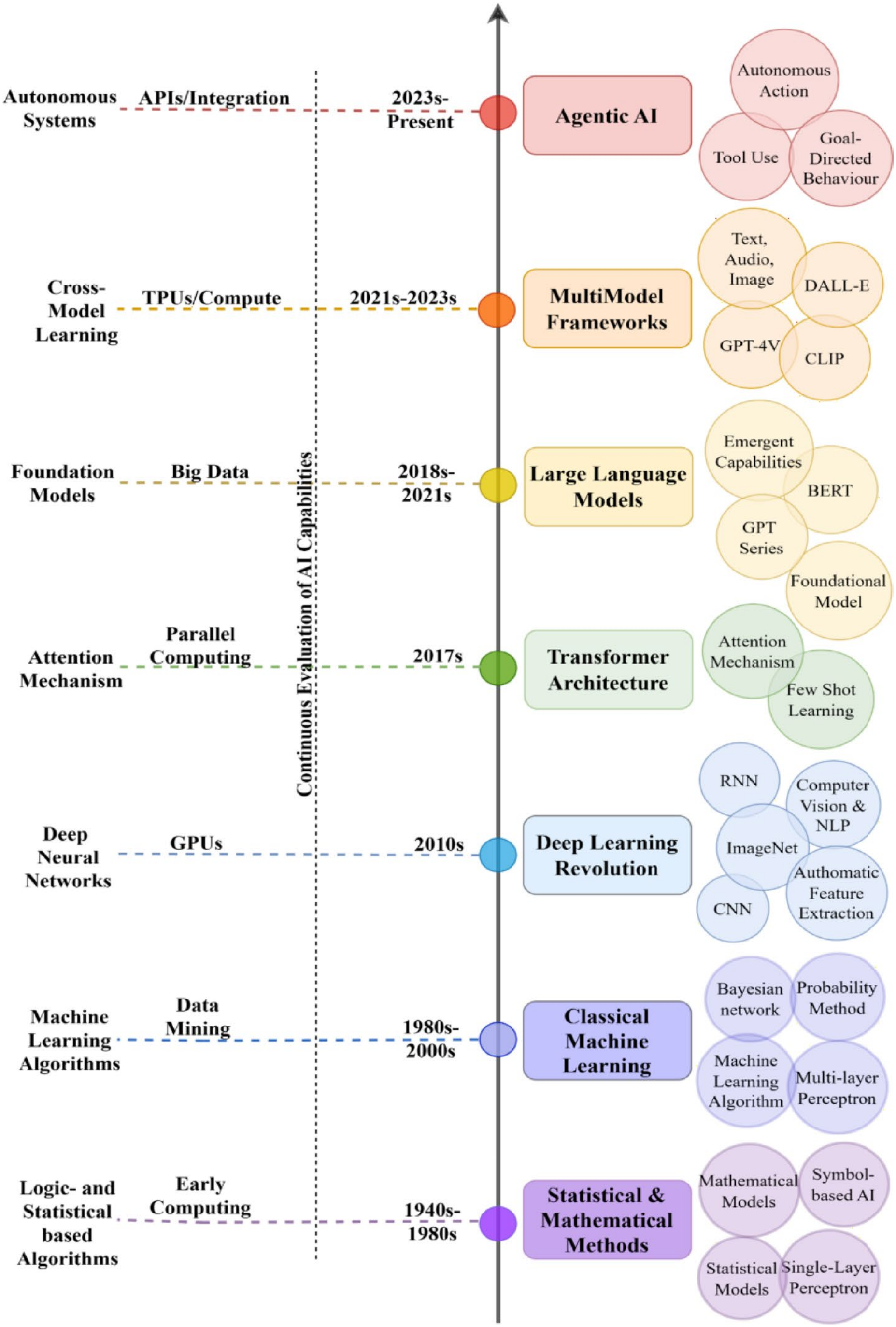


Fig. 6 GenAI evolution timeline over the decades including all AI generations [Author's work]

neering and other fields duringThe 1940 s and 1950s.They wondered whether an artificial brain was something they could build. In 1943 Warren McCulloch and Walter Pitts wrote a book that includes mathematics and algorithms, introduces neural networks and mathematical models, and models how humans think, beginning the study of artificial neural networks. In 1956, AI was formally introduced at a two-month seminar meant for AI studies at Dartmouth, hosted by John McCarthy and others (Taherdoost and Madanchian 2024; Shao et al. 2022).

By the early 1960 s, AI had reached a high point and was considered the first peak. It introduced symbolic logic (Shao et al. 2022; Colelough 2024)and began researching solutions to common problems, along with the early stages of speech processing and dialog tools. Symbolic AI is focused on finding internal ways for machines to represent the world observed by us, where human ideas can be placed into a simple rule-based system (Colelough 2024). First AI uses symbolic logic which helps machines form conclusions from information via mathematical and logical rules.These computer systems work on the basis of the belief that solutions to problems and reasoning tasks in defined domains are possible when symbols are handled according to stated principles. In 1969 their book contained Minsky and Papert's explorations into the behavior of single-layer perceptrons. In 1976, the field of mathematics first used heuristic research (Shao et al. 2022).

FromThe 1980 s toThe 1990 s, AI reached its second peak. Allowing AI systems to reason using 'nonmonotonic logic' and retain their conclusions despite new evidence and incomplete information was considered a large difference compared with classical monotonic reasoning systems (Taherdoost and Madanchian 2024; Shao et al. 2022). Brooks put forward a new understanding of intelligence in 1986, which led to the establishment of behavioral robotics as a significant part of AI. Brooks' view was that behavior-driven intelligence, which reacts in real life without centralized plans, should depend less on symbolic representations (Shao et al. 2022). Using a multilayer perceptron, machine learning solves the perceptron's problem with nonlinear classification by adding hidden layers and training backward (Taherdoost and Madanchian 2024; Telikani et al. 2022; Bommasani 2021). With the probability method and Bayesian network support (Shao et al. 2022; Telikani et al. 2022), AI systems have a structured approach for managing situations with little or incomplete knowledge.

After deep learning appeared in 2010, the third peak of AI greatly sped up the progress and development of society (Bommasani 2021). There are three key factors behind this wave: large-scale data availability, well-improved and powerful GPUs, and novel algorithmic techniques for training neural networks. Several important things show how advanced AI has become. (1) Many important new algorithms are constantly being introduced, including CNNs, RNNs, Federated Learning, and Transfer Learning (Shao et al. 2022; Wason 2018). In addition, generative adversarial networks (GANs), variational autoencoders (VAEs), and ResNet, the development of ImageNet in 2012, enhanced all these novel architectures in 2014 and 2015 (Taherdoost and Madanchian 2024; Colelough 2024; Bommasani 2021). With the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), AlexNet made it clear that deep learning performed better than conventional computer vision, as AlexNet achieved an immense reduction in error rates, which led to more use of deep neural networks across the field (Shao et al. 2022). (2) Many issues in computer vision and NLP that we were unable to solve are now being addressed by AI and deep learning. This is why we now see autonomous cars, automatic language translation, remote medical image review, and voice assistants put into actual use. This finding showed that end-to-end learning allows systems to automatically spot important details from unprocessed data without the need for detailed, handmade features.

Starting in approximately 2017, the fourth peak of AI appeared when the transformer architecture was introduced, allowing attention to models and triggering the rise of foundation models and large language models (LLMs) (Vaswani 2017). At this time, there was a shift from applying deep learning to single tasks to using foundation models that are capable of few-shot learning from a small number of examples and that show different abilities on their own (Bommasani 2021; Qin 2025). The development of BERT, the GPT series, and additional transformer models from 2019 to 2022 led to outstanding progress in understanding and generating language (Zhou et al. 2024; Kaddour 2023). Moreover this wave saw multimodal AI with neural network systems which were recognized as contrastive language-image pretraining (CLIP) fromOpenAI in 2021 leading to progress in image-text relationships via contrastive learning. The same year brought DALL-E from OpenAI, which successfully demonstrated the use of transformers for converting text into images (Li 2024). As a result, it became possible to develop multimodal foundation models that handle and produce information from a variety of data sources. Using input from people through reinforcement learning from human feedback (RLHF) increased the alignment and usability of such models, leading to systems such as ChatGPT being successfully adopted by many users (Qin 2025; Zhou et al. 2024; Kaddour 2023).

From early 2023 onward, advanced multimodal foundation models were developed, with the arrival of agentic artificial intelligence (Li 2024; Huang et al. 2024; Cardoso and Ferrando 2021; Majumder and Dey 2022). At this point, GPT-4 works with vision tools, DALL-E 2 and DALL-E 3 create images of good quality, and Midjourney explores AI possibilities in artistic AI creativity. Advanced multimodal models can easily process images, make realistic pictures from descriptions, and continue a consistent line of thought in text, images, and audio media. They are designed to autonomously schedule, reason, and carry out advanced activities in many areas. Important steps include coping with natural language data with the aid of machines (Qin 2025), several agents solving problems as a team (Cardoso and Ferrando 2021), and AI agents being able to exist in both screen and real-world environments (Cardoso and Ferrando 2021). Owing to their skills in perception, language, and independent action, multimodal agents are an important step toward artificial general intelligence. These systems view their environment and make decisions to support reaching their objectives (Taherdoost and Madanchian 2024). AI includes machines that can understand the words we utter. Among the main tasks in AI research are using and studying objects and being able to learn, think, plan, and communicate with others (Taherdoost and Madanchian 2024; Cardoso and Ferrando 2021; Majumder and Dey 2022).

Researchers faced a significant challenge during the early phases of the creation of GenAI models. This challenge was the convergence problem of generative models. To circumvent this challenge, researchers have utilized a variety of strategies to optimize the stability of the GAN (for example, by increasing awareness of the behavior of GAN training) (Sengar et al. 2024).

Additionally, GenAI has demonstrated state-of-the-art performance in the resolution of complex real-world problems in a variety of domains, including but not limited to image translation, medical diagnostics, textual imagery fusion, NLP, and other areas. The term "generative artificial intelligence" refers to a wide range of subtypes, each of which is designed to perform certain functions or generate different kinds of media. These are some of the types that are more often recognized. Among the many types of adversarial networks, some examples are GANs, transformer-based models (TRMs), variational autoencoders (VAEs), and diffusion models (DMs) (Sengar et al. 2024; AlDahoul et al. 2025; Mao 2024).

2.2 Comparative analysis of recent GenAI literature

Table 3 presents a comparative discussion of diverse aspects of GenAI, including ethical and cybersecurity-related issues, sociotechnical consequences, the effectiveness of technology-neutral regulation, etc.

2.3 Emerging trends in genai: large concept models

To overcome the limitations of traditional LLMs, Large Concept Models (LCMs) (Barrault et al. 2024) have been proposed, which are innovative architectures that transform the main processing unit component by shifting from a fragmented token to entire semantic representations, which are called concepts. Unlike LLMs, which think sequentially in words or subwords (Nabi 2025), LCMs think on a higher level of abstraction, consisting of and reasoning about entire concepts. The LCMs have the potential to better handle long-context operations by contracting phrases or conceptual groups and creating both consistent and interpretable outputs (Prakash 2025). The presented conceptual framework not only shows how people organize and make meaning of information but also considerably reduces the computing costs associated with processing long sequences (Zen 2025). LCMs are exceptionally efficient in cross-lingual applications, discovering new languages with ease, and are highly productive in multimodal applications such as the composition and processing of documents in a variety of languages without retraining and translating and transcribing live authorial speech (text) (Pruseth 2025). Their ability to organize and detail a vast body of knowledge with a pertinent context allows them to be especially adept at any activity that demands an understanding of comprehensive documentation (Bamaniah 2025). LCMs overcome scalability limitations by moving to the concept-level modeling scale, which allows them to manipulate larger datasets and more complex activities and set new standards in terms of efficiency and interpretability (Razzaq 2024; Shrikhande 2025; Ahmad 2025).

Despite major research investigating the evolution and trends in generative AI, many studies fail to investigate current paradigm shifts, such as large concept models and the rise of data- and knowledge-based approaches. This critical insight integrates new GenAI research with main AI concepts and applied theories, which seems to be missing in much of the previous literature. This allows us to focus our attention on recent studies and unresolved issues, ensuring that the study does not follow a conventional pattern. Several studies have highlighted the revolutionary potential of GenAI across various fields. However, all studies frequently overlook the analysis of ongoing challenges such as differences in equity and broader ethical implications. The current changes discussed in this systematic review are provided with the context of the characteristics of GenAI, which can be seen as the domain of dynamic features and possible developments.

Table 3 Comparative discussion

Paper	Publication with Year	Research Objectives	Findings	Limitations
Petar Radanliev (2025)	Frontiers, 2025	This research examines regulatory possibilities of managing the vulnerabilities related to AI, such as model reversal attacks, data poisoning, and adversarial manipulations by stealing the knowledge of cybersecurity, legal studies, and computational risk analysis.	The methodology will be the comparative review of national and worldwide AI policies and their efficiency in managing the emerging challenges. An analysis of the purpose of cryptographic approaches, such as homomorphic encryption and zero-knowledge evidence, to address better compliance, protect delicate information, and ensure algorithmic accountability, is explored in the paper. Studies find that current regulatory efforts lack uniformity and reactivity by including necessary mechanisms to mitigate the rising hazards associated with frontier AI.	Further studies will need to feature particular examples of good regulatory frameworks and implementation strategies to maximize the applicability of the ideas around the governance of AI. The examples of the AI auditing processes developed by the Information Commissioner's Office (ICO) in the United Kingdom (UK) or the examples of applying algorithmic impact assessments in Canadian AI governance provide valuable insights into the way AI governance is supposed to work.

Table 3 (continued)

Paper	Publication with Year	Research Objectives	Findings	Limitations
Petar Radanliev et al. (2025)	Frontiers, 2025	This study undertakes a systematic study and involves combining a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-guided literature review with conceptual and quantitative analysis to examine the sociotechnical consequences of generative artificial intelligence.	The findings of this research show that generative AI supports transformative potential, such as the synthetic generation of data, multimodal content creation, but also raise some potential concerns, e.g., adversarial manipulation, reidentification of anonymized data, and deepfake proliferation. The issues have been compounded by the fact that generative models are rapidly evolving in industries where there are poor governance structures. In response, this causes resilience to be redefined technically and in respect to the ethical sense of accountability, as well as adaptation of policy.	There are certain limitations and issues associated with the AI model development. The main issue is that there is a significant amount of computational costs associated with the training and creation process. The creation of a plausible deepfake is essentially a demanding task that requires a large amount of data and high processing resources. The other drawback is that AI models rely heavily on the quality of the training data. It relies on the quality of the output generated with respect to the quality of the training data. In turn, in case of biased or misleading nature of the data used to train the AI model, the AI model is likely to propagate such biases and further inflate them, which has the effect of impairing the accuracy and fairness of the information produced. This is particularly relevant in those scenarios that have moral concerns like deepfakes.

Table 3 (continued)

Paper	Publication with Year	Research Objectives	Findings	Limitations
Werisha Ibrar et al. (2025)	Springer, 2025	This research paper tries to examine the cyber threats associated with AI-developed malware. This paper describes the vast scope of malware created by GenAI through the examination of varying types of attacks, their prevalence rate, and their impact. These results help improve the understanding of the evolving landscape of cybersecurity and offer a guide on the creation of new rules and procedures to defend individuals, organization, and the rest of society against risks represented by malware generated through chatbots.	The findings of this research have numerous recommendations on how the challenges posed by GenAI in cybersecurity can be handled. Policies should be drawn to differentiate between high-risk and low-risk application of GenAI, specifically with creation of code. GenAI models can give rise to significant concerns in the creation of encryption and decryption algorithms. Such code has the potential to be used to create ransomware or other malicious applications when altered, present a serious threat to cybersecurity. Furthermore, applications that are considered the highest risk should be subject to more regulatory control especially where they are associated with the development of potentially malicious code e.g. malware, ransomware, etc. Such situations should always be monitored, and these features should only be accessed by individuals who have valid intentions and have been verified.	Measures were put in place to minimize the biases by including the research studies of diverse fields and regions. However, the number of limitations that are faced is known, namely, the lack of access to non-English articles and the potential significance of a specific technology or dataset. Such restrictions reveal that additional research is needed to provide a more detailed insight into the implications of GenAI in cybersecurity.

Table 3 (continued)

Paper	Publication with Year	Research Objectives	Findings	Limitations
Mueen Uddin et al. (2025)	Springer, 2025	This research paper focuses on how Generative Artificial Intelligence will facilitate cybersecurity protection because of the exploitation of Artificial Intelligence models and algorithms. GenAI is capable of solving common security issues, discovery of new threats as well as increasing human participation in critical safety matters on their own. The research paper is emphasizing autonomous security developments, a better response to new threats, detection of anomalies and responding to threats. We also researched the boundaries of GenAI such as inaccuracy on many occasions, the requirement to spend a lot of money training the AI and what could be exploited by ill-minded individuals to break the law. The study provides meaningful information regarding the reasonable use of GAI in cybersecurity, making it easier to perform the threat reduction without compromising system security.	GenAI is transforming cybersecurity through the automation of threat detection, response, and intelligence by taking the place of conventional analysis and signature-based approaches. GenAI models have the capabilities to replicate malware, phishing attacks, adversarial attacks, enhance password attacks, test vulnerability, and anomaly detectors. The technology is also being incorporated in major security platforms with an aim of enhancing efficiency and security. Nonetheless, these developments can be seen as risks, including malicious actors who can misuse GenAI, poor accuracy or biased results, high fees, and the privacy, responsibility, and clarification issues. The combination of responsible behavior and ethical values with technical innovations in the study is the most effective way to maximize the benefits of GenAI and reduce their potential cybersecurity issues.	The proposed survey will be conducted to help current and future scholars, policymakers, and organization leaders gain information regarding the opportunities and threats associated with the further development of GenAI. It also decomposes the strategic plan of developing cybersecurity, pointing out the great contribution of creativity and sustainability in the implementation of cybersecurity ingenuities.

Table 3 (continued)

Paper	Publication with Year	Research Objectives	Findings	Limitations
Antonio Cordella et al. (2024)	Elsevier, 2024	This study investigates recent bans of ChatGPT in Italy to study the impact of the distinct technological characteristics of generative AI on the effectiveness of technology-neutral regulation. The study contributes to the understanding of the tensions among the unique technological capabilities of generative AI and the effectiveness of a technology-neutral regulatory framework, informed by the knowledge gained during an exploratory case study. Some related implications are given by the research on the practice suggesting the absence of adequate management of the specified tension can lead to the failure of public regulating interventions to achieve their announced goals.	The findings presented in the paper contribute to the existing body of literature by emphasizing the critical importance of implementing a better insight into technological advances on generative AI laws and regulations. The article highlights the challenges of sufficiently serving the broad implications of generative AI on social and public sectors and consequences without necessarily having an in-depth knowledge of the inherent technological principles. Generative AI technology is unique and requires specific academic research efforts to enhance our understanding of how regulatory strategies can be transformed into more dynamic and practical toward the evolving character of AI systems.	This research presents two important boundary conditions. To begin with, this study is exploratory to some extent, which restricts its implementation. The case study is about the adoption of General Data Protection Regulation (GDPR). Other technology-neutral AI policies may not be relevant regarding the findings of this research. Nonetheless, the collection of data is constrained by a small amount of documentary materials and the use of only one interview with the main Italian regulator. Despite the fact that conversations with the Garante and OpenAI were not disclosed, the dataset has all the information available. Last, the purpose of the article was to address the issue of the applicability of technology-neutral laws on the specifics of generative AI in consideration of these limitations. In this research, emphasis was on the analysis of limits, rather than other regulatory methods since they would be beyond the scope of this research.

3 Importance of generative AI

Because of its capacity to produce materials that are both original and innovative, the field of GenAI has the potential to have a significant influence across a wide range of industries. In addition to its wide range of applications, it has the potential to transform a variety of different sectors and bring in creative solutions. In the field of computer vision, for example, GenAI algorithms play a significant role in the generation of realistic images. This helps with the augmentation of data, the synthesis of training datasets, and the restoration of image segments that have been compromised or missing. In addition, these algorithms are essential in the process of performing tasks such as image translation, style adaptation, and transformations from one image to another. Generative models are increasingly being used in a wide variety of activities within the field of NLP. These tasks include but are not limited to the production of text, the translation of machine translations, the summarization of products, and the development of conversational interfaces. These models demonstrate a remarkable ability to create text that is both consistent and contextually relevant, and as a result, they contribute to the progress of language generation systems and conversational agents. In the fields of creativity, such as music and art, generative algorithms serve as both a source of inspiration and a tool for assisting in design prototyping, which ultimately results in an enrichment of creative processes (Corvello 2025; Qin et al. 2023; Al-Sarayreh et al. 2023; Balasubramaniam 2024).

As shown in Table 4, the significance of GenAI may be understood by considering the wide range of potential applications it has across a variety of industries. In the field of computer vision, these generative algorithms can produce genuine images, improve training datasets, and contribute to tasks such as data reconstruction and inpainting. Text creation, translation, summarization, and the creation of conversational bots are all areas in which they play important roles in NLP. Furthermore, generative models are particularly effective in the generation of native-sounding speech for applications that include synthesis. Technology has been shown to be an essential instrument in creative and artistic areas, particularly in the areas of music composition, the production of visual art, and the development of design prototypes.

As shown in Fig. 7, the importance of GenAI is highlighted by its ability to revolutionize a variety of industries, stimulate creative thinking, and solve complex challenges. The field is continuously experiencing significant gains, which are largely driven by the advancement of deep learning techniques and neural network designs. As a result, future research is expected to result in breakthroughs and innovative applications, which will provide social benefits.

Table 4 Significance of generative AI [Author's work]

Computer Vision	Synthesize a realistic image Enhance training data Assists with data completion Corrupted portion of images Image translation Style transfer Image-to-image translation
NLP	Text generation Machine translation Summarization Dialogue system
Speech Synthesis	Generate realistic voices
Creative Field	Generating music Generating artworks Design prototypes

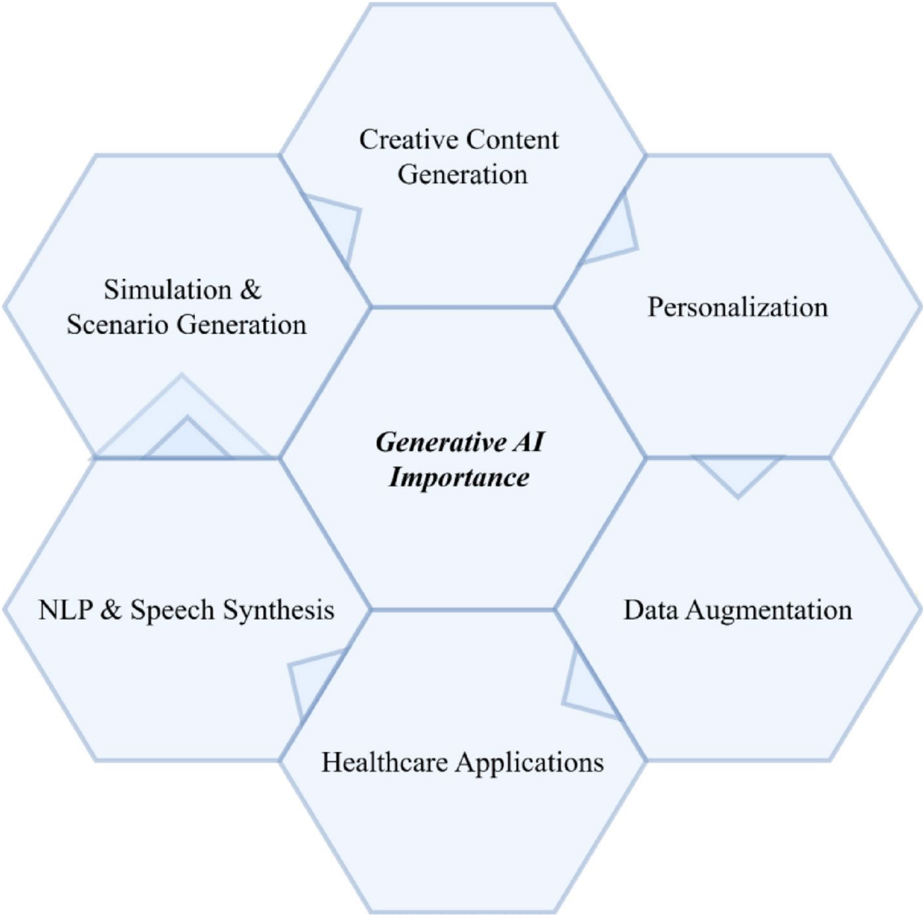


Fig. 7 Generative AI Importance [Author's work]

4 Major role of generative AI from the industry perspective

GenAI is transforming industry sectors to automate complex and time-consuming tasks. It enhances innovations and quick decision-making support through the automation of daily tasks. Because of its ability to make tasks more efficient, accurate, and personalized, the use of GenAI is reshaping the way businesses work and succeed today. Organizations need AI not only for automating work but also for adding value by working together with workers to create better results. Figure 8 shows the diverse range of industrial applications of GenAI, including finance, cloud and IT support, healthcare, education and energy.

4.1 Finance sector

It is believed that GenAI will play a significant role in the development of autonomous finance from the point of view of the financial sector. When we refer to autonomous finance, we refer to the automation of financial processes, decision-making, and services through the

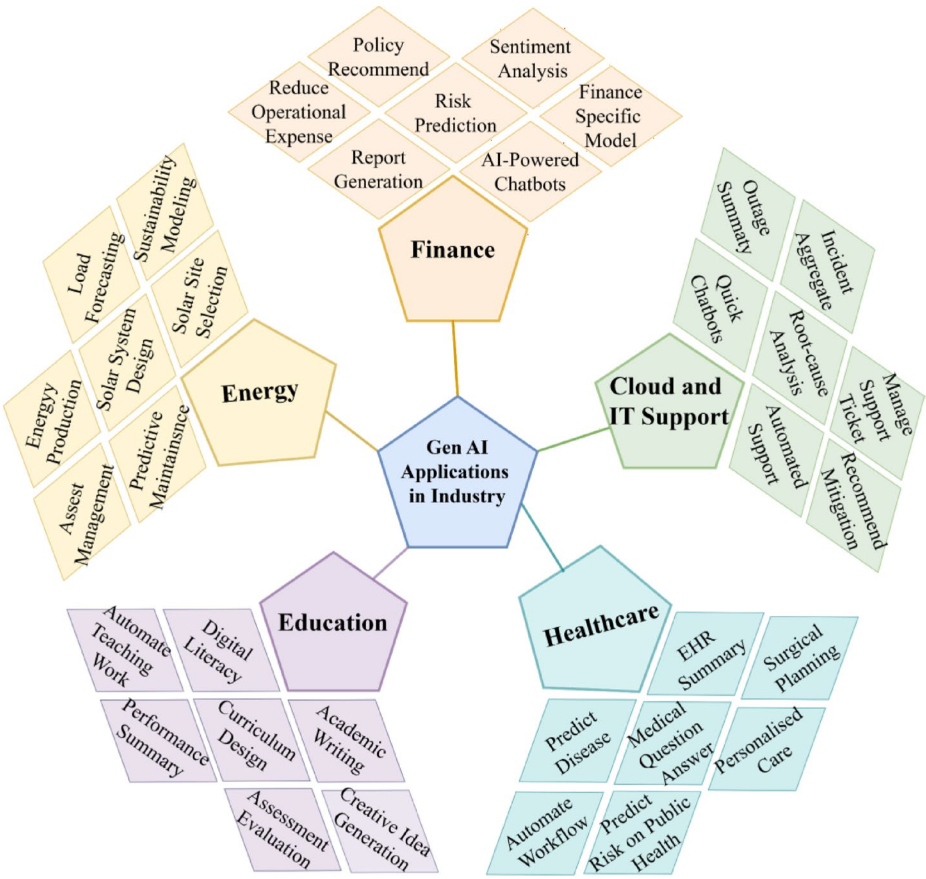


Fig. 8 Generative AI Applications in industry for Digital Automation [Author's work]

utilization of artificial intelligence and hyperautomation. Increasing numbers of financial organizations and other businesses are adopting autonomous finance because of its potential to improve efficiency, reduce operational expenditures, and enhance customer experiences. This is the reason for the growing popularity of autonomous finance, which is also outlined in Table 5.

A financial component that is autonomous makes use of technologies that go beyond typical automation to include skills such as self-learning and self-correction. These technologies can also make judgments on the basis of the data that they collect. As a case study, organizations such as Morgan Stanley make use of chatbots driven by OpenAI to assist both clients and financial professionals. These chatbots make use of internal collections of datasets and research as sources of expertise, and they can assist users by providing them with precise information. Additionally, Bloomberg has launched a generative model that is dedicated to the finance sector called the Bloomberg generative pretrained transformer (GPT). This model is capable of assisting with various financial operations (Sai et al. 2025; Huang and Ren 2024; Luo et al. 2024; Wu 2023). Figure 9 shows the autonomous finance industry with real-world applications from GenAI.

Table 5 Advancement in the finance sector by GenAI

Focus Area	Applied Methods	Dataset	Potential Applications
Green Finance: Innovation and Digitalization in Corporate (Huang and Ren 2024)	Firm-level data with Empirical Analysis	Chinese listed companies: Green finance and digitalization metrics	Recommendation on introducing the policy for the reformation of digital and green finance integration.
Associated Financial Risk Prediction in Carbon Trading (Luo et al. 2024)	Deep Neural Network	Historical carbon market and environmental data	Decision and Risk Management support in carbon markets
Financial Domain Specific – AI NLP (Wu 2023)	LLM with 50B parameters (General data, Proprietary data, Public Sources)	Outperforms on financial tasks	Financial Market related sentiment analysis, updates report generation, regulatory compliance, and summary generation

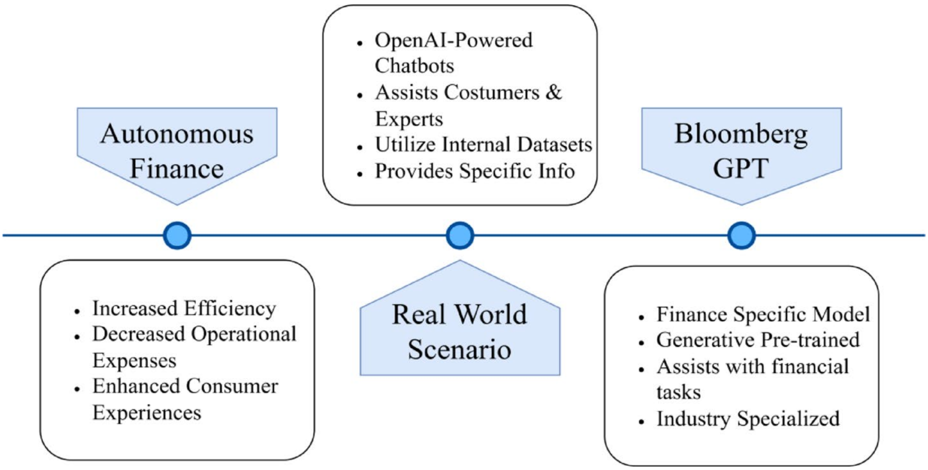


Fig. 9 GenAI in autonomous finance including assistance in real-world use cases of finance related tasks [Author's work]

4.2 Cloud and IT support sector

Most cloud platforms manage incidents, and Amazon, Google, Microsoft, and Salesforce, as leaders of the industry, have transitioned to providing services through the cloud. Table 6 outlines the use of large language models (LLMs) by the cloud service providers Amazon, Google, Microsoft, and Salesforce, which has led to major improvements in their incident management systems. GenAI has many applications in cloud and IT support.

Table 6 Evolution in the cloud and IT sector by GenAI

Area of Application	Solution by GenAI	Major Benefits	Cloud Platforms
Root Cause Analysis and Mitigation Recommendation Tasks (Ahmed 2023)	LLM along with a dataset of 40,000 + incidents on cloud platforms	Automated Analysis of incident issues and suitable recommendations accordingly	Microsoft, Amazon, Salesforce, Google
Impact Assessment (Jin 2023)	OASIS framework and rule-based machine learning approach	Accumulated incident analysis and instant scope discovery	Major cloud infrastructures
Outage Summarization (Jin 2023)	GPT-3.x - fine-tuned	100 times faster generation of Summaries from incidents and outages with (when/where/who/what/why)	A total of 18 systems form real-world cloud platforms
Support Ticket Classification (Zangari et al. 2023)	Transformer: BERT-based	Manages unstructured information, Multilevel classification	Various Service Providers Total of 20,000 complaints and 35,000 bug reports

The use of GenAI by LLMs is transforming this domain by processing extensive databases of 40,000 incidents collected from several cloud services. Through automated root cause analysis and recommended mitigation solutions, LLMs enhance the pace of resolving problems and help spread knowledge among several engineering teams. To demonstrate, LLMs can combine information on historical incidents and extract useful insights so that all members of the team can leverage their organization's experience, making learning take less time and avoiding dependence on a few experts (Ahmed 2023). In addition, evaluations were performed with different versions of GPT-3.x both from scratch and with training, to determine how effective AI models can be in cloud incident management situations. By fine-tuning historical records, these models can suggest relevant recommendations that are more precise and meaningful, and by using multitask learning, they are able to handle both finding the cause of incidents and mitigating planning. GenAI systems reveal the need for human-in-the-loop validation, where incident owners review and confirm whether the AI outcomes are correct, simple to read, and practical. It helps people trust technology and guarantees that AI recommendations remain connected to what is needed in the industry (Ahmed 2023).

Cloud outages present a challenge because they affect millions of customers, and engineers need to quickly evaluate the extent of the problem and make detailed reports (Jin 2023). GenAI can respond to these issues through outage assessment and summarization via the intelligent systems (OASIS) framework, which uses automated systems to provide an overview and a summary when outages occur. With the help of rules, past records, and machine learning, impact assessment combines many relevant incidents so engineers can quickly identify all the issues involved in an outage. By using fine-tuned large-scale language models (GPT-3.x), it is possible to create summaries that highlight when, where, what, and why in the news. A test on 18 real cloud systems revealed that OASIS performs better and faster than current techniques do in terms of accuracy. LLM summaries are just as effective as those written by engineers yet are generated at a hundred times faster rate than engineers (Jin 2023).

The classification of support tickets through automation can also help departments work more efficiently (Zangari et al. 2023). Today's service providers rely on automating tickets to improve their helpdesk workflows and provide immediate answers to customers. With BERT, transformer-based language models improve multilevel ticket classification. On the basis of 20,000 customer complaints and 35,000 bug reports, how documents are embedded can greatly influence how well LLMs perform in classification. As a result, using contextualized language models allows companies to handle complex situations, resulting in better organization, fewer errors, and increased efficiency at the same time (Zangari et al. 2023). Using LLM tools, companies can respond to customer situations faster and in a better way, rely on customers to trust them, and remain ahead of competitors operating digitally (Ahmed 2023; Jin 2023; Zangari et al. 2023).

4.3 Healthcare sector

Healthcare is being transformed rapidly by GenAI, which modernizes treatment procedures, diagnostic tools, and the way in which decisions are made. Table 7 shows how the integration of GenAI technology into clinical and nonclinical domains is advancing patient care, research efforts, and healthcare administrative services.

GANs and VAEs, along with synthetic data generation, are important for AI training and algorithm validation without compromising patient privacy, especially in dealing with rare diseases or a lack of data availability. Because of this method, better outcomes are seen in

Table 7 Advancement in the healthcare sector by GenAI

Application	Model Used	Dataset	Advantage
Personalized Treatment Plans (Bhuyan 2025)	GANs, GENTRL	Retinal images, Virtual patient populations	Personalized therapy, Simulation of rare disease data
Medical Image Analysis (Bhuyan 2025)	GANs, VAEs	MRI, CT, X-rays	Enhanced diagnostic results
Surgical Care (Bhuyan 2025)	Johnson & Johnson MedTech AI	Surgical Videos, Electronic Health Records (EHR)	Advanced surgical administration
Risk Prediction (Bhuyan 2025)	GANs, AI models	Social media, EHRs	Early disease detection
Revenue Cycle Management (Bhuyan 2025)	GenAI integrated RCM	EHRs, billing data	Improved billing, reduced errors
Medical Visual Question Answering (MedVQA) (Dong et al. 2025)	LLM, MedFlamingo	VQA-RAD, Path-VQA, MedVQA-2019, ImageCLEF-VQA-Med	Clinical Reasoning, Context-specific answers
Patient Communication (Morjaria et al. 2025)	ChatGPT	Clinical notes, patient queries	Empathetic, accurate patient messaging
Provider Burnout Reduction (Morjaria et al. 2025)	GPT-4	EHR communications, patient messages	Lowered task load and burnout
Clinical Text Summarization (Alkhalaf et al. 2024)	Llama 2	Aged care EHRs	Automated health risk extraction

dealing with diabetic retinopathy, liver lesion detection, and planning and performing surgery, as reported using MR images of the brain. Some major applications include creating personalized plans of treatment for each patient, generating data, and making improvements in medical images that are easier to analyze (Bhuyan 2025).

Medical visual question answering (MedVQA) is a notable way that GenAI interprets X-rays and other images in answer to questions from clinicians. Owing to vision encoders such as contrastive language–image pretraining (CLIP) and vision transformer (ViT) and LLMs such as large language model meta AI (LLaMA) and GPT, the systems can stream answers in various ways and are suitable for use in remote consultations, clinical assessment, and diagnosis. With these tools, the chance of diagnostic mistakes decreases, and there is better healthcare access in places where it is currently limited (Dong et al. 2025; Nori 2023).

Because of GenAI, surgical care is improved by providing instant recommendations that rely on patient data and up-to-date research, as seen in Johnson & Johnson's AI-driven surgical initiatives. Nursing schedules and patient communications are optimized by chatbots and models such as ChatGPT to generate helpful and empathetic responses, decrease clinician burnout, and redirect more time to care for patients with readability and accuracy (Morjaria et al. 2025).

GenAI supports clinical work by summarizing patients' cases with LLMs such as Llama 2 and identifying potential risks, such as malnutrition, in elderly people before it is too late. GenAI processes up-to-date information at the health system level to manage public health and predict outbreaks, as proven with COVID-19 research (Bhuyan 2025; Alkhalaf et al. 2024).

GenAI uses NLP and LLMs to make unstructured EHR data shorter to generate a summary, which helps doctors review information more quickly and with greater accuracy. Summarizing radiology reports and patient notes is efficient with GPT-3, ChatGPT-4, FLAN-T5, and Med-BERT, especially in times when it needs to be quick. GenAI can predict diseases by examining EHR records, and Med-BERT is among the models that achieve accurate results (Lee et al. 2024).

Despite progress, some issues concerning varying data, fairness, privacy, bias, and how things are explained still exist (Lee et al. 2024). In medical education, GenAI enables realistic simulations and adaptive learning, but robust ethical guidelines are essential for safety, transparency, and accountability (Bhuyan 2025). GenAI is reshaping healthcare by improving data interpretation, diagnostics, and access to care. Continued innovation, rigorous evaluation, and strong ethical oversight are crucial to ensure that these technologies deliver equitable, high-quality healthcare.

4.4 Education sector

Educational environments at both secondary and higher levels are being changed rapidly by GenAI, as discussed in Table 8. GenAI is transforming education at various levels, creating new possibilities for teaching methods; however, at the same time, it introduces issues involving ethics, how people think, and other problems. As shown in Fig. 10, GenAI has emerging applications at all levels of education that introduce innovative techniques in pedagogical practices.

Table 8 Expansion in the education sector by GenAI

Education level	Prior application	Model used	Outcome
Higher Education (Batista et al. 2024)	Teaching, Research, and Admin work	ChatGPT, GenAI Tools	Identify the major trends, Personalized Learning, and the Gap in policies and Faculty training
Secondary High School (Wu and Zhang 2025)	Project-specific learning, Enhance literacy for innovation using digitalization	ChatGPT and Similar Tools	Well-improved technical skills in students, Guidance on digital skills, Critical thinking, and reasoning
All levels From K–12 to university (Bahroun et al. 2023)	Curriculum structure design and development, Assessment evaluation support	Bibliometric analysis	Major application on trend and insights, encourages empirical study at the school level, and research ideas on ethics and transformation
University Education (Wei et al. 2025)	Group story writing, collaboration on creativity, and productivity	GenAI for story prompts and refinement	Improved performance in group creativity
Higher Education (Kim et al. 2025)	Ideation, Feedback, and Essay Writing	ChatGPT, Writing assistants	Enhancement of AI literacy and balanced usage through identified trends as iterative, assistive, and overreliance

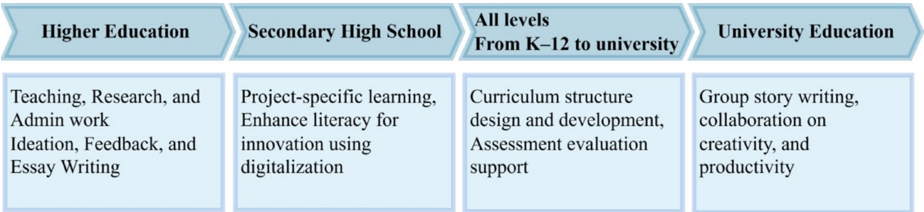


Fig. 10 GenAI in the education sector at all levels [Author's work]

Research provides a foundational understanding of GenAI in higher education, identifying critical trends and challenges (Batista et al. 2024). The study indicates that higher education institutions are rapidly adopting GenAI tools such as ChatGPT for tasks such as academic support, administrative automation, and curriculum development. Educators and students alike are exploring its ability to generate course content, summarize academic materials, and facilitate communication. However, concerns remain regarding academic integrity, data privacy, and the absence of formal institutional guidelines for ethical usage (Batista et al. 2024).

GenAI in higher education is tied to several important trends and challenges (Batista et al. 2024). Higher education institutions include GenAI tools in their daily operations for academic support, automating admin workflows, and improving the curriculum. Many people involved in education are trying out the various uses of AI for making lessons, summarizing readings, and supporting conversations between students and teachers (Batista et al. 2024).

The use of GenAI improves students' ability to think creatively and work with technology at the secondary school level. Using generative models in project-based learning helps students develop both impressive creativity and digital skills. Integrating GenAI allows students to think up ideas faster, manage their work clearly, and increase the standard of their work. However, the authors noted that teacher support helps avoid excessive reliance and ensures that students are actually using AI tools properly (Wu and Zhang 2025).

Beyond educational levels, bibliometric and content analyses can be used to trace how GenAI is used in education. The study reveals three leading research themes: learning personalization, ethical challenges, and instructional transformation. Compared with other sectors, higher education has received more of the current attention, but GenAI is also becoming more important in the K-12 and vocational fields (Bahroun et al. 2023).

In universities, both creative stories and problem solving were accomplished more effectively by groups that used GenAI than by the control groups. However, depending on AI-generated content seemed to decrease how much participants talked to one another and thought critically. While GenAI boosts creativity, it can unintentionally decrease soft skills when not managed correctly (Wei et al. 2025).

Focusing on how students interact with GenAI tools in academic writing. Learners can use three main engagement strategies with GenAI tools: ask for help, continuously improve, or count on problem-solving too much. While helping students develop their writing and grammar, students should be equipped with AI policies and knowledge so that they can harness GenAI safely without relying less on mental effort while writing (Kim et al. 2025). Organizations need to prioritize teaching staff, redesign curricula, and team up across departments to guarantee that GenAI technologies add, rather than replace, real learning experiences. A balance is needed to take full advantage of the use of GenAI for education.

4.5 Energy sector

GenAI is shifting the solar energy sector by improving the way solar energy operates. With solar power being used more in modern power systems, GenAI offers scalable and smart solutions for forecasting energy, making plans, and deciding which actions to take. Smart, efficient, and sustainable operations in the energy sector are accomplished via the integration of GenAI, as discussed in Table 9.

With Gen-AI, solar panel designs can become more advanced, system sizing can be better calculated, and maintenance tasks can be predicted via GANs, VAEs, and RL methods. It supports the right predictions, the linking of renewables to the grid, picking the best sites for projects, and adaptations to weather. Using simulations of actual situations and weather conditions, Gen-AI helps solar farms work better and with more stability. It helps avoid delays, makes energy management sustainable, and saves money. Overall, Gen-AI helps solve problems in renewable energy with a proactive and smart approach, offering new ways to design, operate, and look after solar systems (Mousavi 2025).

Solar power scenario generation in the energy field has improved greatly with the use of GenAI, especially with the help of deep generative models (DGMs). It effectively han-

Table 9 Impact of GenAI on the energy sector

Application	Model	Dataset	Impact
Design and Optimization of Solar System (Mousavi 2025)	GANs, VAEs, RL models	Historical data, Sensor data	Downtime cost reduction, Improved design of solar panels, Optimized siting
Scenario Generation, Energy Forecasting, Unit Commitment, Resource Allocation, Dispatch (Kousounadis-Knousen et al. 2023)	Deep Generative Models (DGMs)	Weather data, Historical data, Solar Irradiance data	Improved forecasting on diverse scenarios, More accurate
Sustainable Operations (Green GenOps) (Sánchez-Mompó et al. 2025)	Various GenAI Models with billions of parameters	AI pipeline – CPU, GPU, DRAM usage	Efficient reduction in the energy consumption of AI models
Grid operations support, Real-time data interpretation (Matharaarachchi 2024)	LLM, Multimodel data analysis, GenAI-powered chatbots	IoT sensor data, Visual data, Textual data	Encourage net-zero emissions goals, improve decision-making, and flexibility in managing solar and grid systems

dles uncertainty, regular changes in time, and spatial and time-based dependencies in solar power generation (SPG). Because all the weather conditions and possible problems are not easy to predict, traditional models are often unable to describe the complex and variable nature of trucking. GenAI works around this by creating various realistic scenarios via stats, which do not require the assumption of a previous distribution. Planning, unit commitment, energy dispatch, and resource allocation are some of the main applications (Kousounadis-Knousen et al. 2023).

Large language models can be used sustainably to lower energy usage related to AI. GenAI models use considerable electrical energy, which is especially high during the inference phase because they require many complex calculations. Green GenOps is a tool for MLOps that helps manage GenAI processes while paying attention to sustainability. It checks the energy use for various model sizes and hardware setups, which are viewed over different requests. Small GenAI models (between 1B and 8B parameters) are energy efficient and suitable for handling more requests because they work well with batching and caching. Because 70B uses considerable computing power, it is only worth the cost if it is deployed smartly. GenAI offers suggestions for minimizing the carbon footprint, including fine-tuning smaller AI models, choosing energy-conscious benchmarks, and including regular energy analysis in GenOps systems (Sánchez-Mompó et al. 2025).

The power of natural language and data synthesis in the IIoT enables energy companies to study and use the large and complicated data produced by IoT systems. UCAs supported by AI make it possible for operators to look at data as it changes, identify where changes can be made, and decide what is best, which in turn optimizes the grid and speeds up the adoption of net-zero goals. These systems make it simpler to use energy data, handle regular jobs automatically, and interact with systems via understandable tools; thus, managing distributed energy resources, renewables, and smart grids becomes simpler. GenAI provides better security, greater scalability, and greater adaptability when managing energy in real-world applications such as university campuses aiming for zero carbon emissions (Matharaarachchi 2024).

GenAI is used for systems such as predictive maintenance, in which it analyzes sensory information and captures early signs of problems to prevent expensive and time-consuming breakdowns. GenAI improves energy forecasting by reviewing large amounts of information to find trends in both supply and demand, which helps keep the power grid steady and distributes resources wisely. GenAI also plays a role in smart grids by keeping an eye on objects in real time and controlling the flow of energy to avert power outages (Storey et al. 2025).

Although a number of research papers have been published on the industrial use of GenAI, the different industries and areas of use have not been adequately covered, nor has the level of adoption in various sectors. The present findings integrate the literature of the financial sector, cloud & IT support, healthcare, education and energy with the most recent GenAI literature research, thus providing a detailed overview of the areas of limitation and best practices. The financial and healthcare industries are already undergoing extensive development; however, their general application is limited by concerns about ethical protection and privacy. Energy, education and cloud and IT support are promising areas of application, but these areas are often limited to pilot projects as well as applied conceptual frameworks. Organizations frequently maximize the development of solutions that are available even when sufficient stability is lacking.

5 How startups can use generative AI to increase opportunities, adopt strategies and limit risks

5.1 Generative AI in startup growth: an example use case

After looking at major growth factors and obstacles, we establish a framework that allows us to find specific cases where startups use GenAI. The approach, named AI Wheel Fig. 11, shows how different growth steps should be used in GenAI on the basis of funding levels. Product-related content and understanding of go-to-market strategies promote product-led growth, but improving customer services with automation and making new services personally improve sales-led growth. Since resources are limited for startups, using GenAI allows businesses to improve the productivity of their human resources and alleviate the pressure on their customer management workload (Rezazadeh et al. 2025; Siddik et al. 2024).

5.2 Incorporating generative AI into start-ups: important qualities and strategies

GenAI could have a considerable effect on how businesses and technology operate by significantly increasing automation levels. AutoGPT and other autonomous agents are expected to improve and integrate into business procedures. A seamless connection between IT and OT will result in automation, which leads to greater productivity. You can combine software

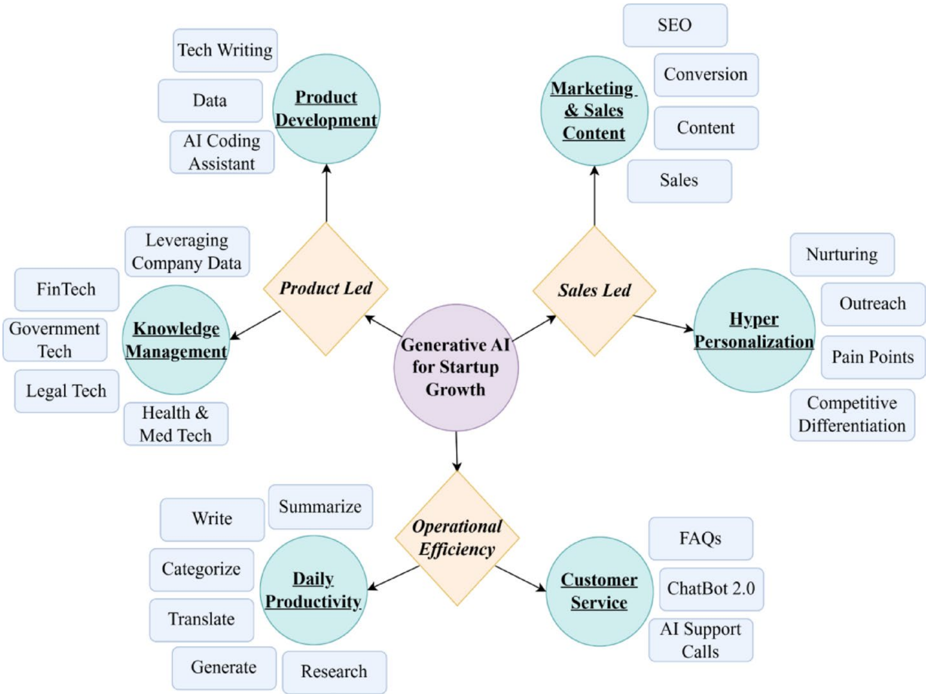


Fig. 11 AI wheel: a framework to map generative from startup growth [Author's work]

and AI currently with the help of Zapier. In addition, custom code built with frameworks such as LangChain in Python can adjust GenAI to meet various company requirements and be executed in prompt chains and bots. Combining these two areas with the AI wheel can significantly increase business automation and lead to more integration of self-operating AI-powered software systems.

To evaluate an output without verification, one must rely on reflective knowledge to check if it is factual, if there is any copyright infringement, or if it contains hallucinations. Interviewee 18 mentioned that LLMs support productivity only when the user knows how to ask meaningful questions. When dealing with new and one-of-a-kind issues, LLMs can face troubles because they do not have previous exposure to such types of problems.

Data from the framework indicate that startups generally follow a top-down approach to implementing AI-based processes for the whole company. First, they should become familiar with the GenAI features on their platforms, and subsequently, they should learn how to use them together. At this point, we need to use reflective knowledge to merge AI abilities with human expertise to prevent AI errors and meaningless AI-created content. Additionally, they routinely add and improve features in their products by analyzing customer opinions and applying AI to aid growth. Currently, many startups are trying to integrate GenAI into their systems as soon as they begin moving ahead of their competitors. The key factors that lead to successful implementation are C-level support and a well-structured process for generating ideas from the bottom up. Management has to make it clear to staff that GenAI is boosting progress, not holding it back. If they switch jobs, their employees may not fear the impact of change as much. Managers must adjust quickly to the rapid changes caused by advances in AI. The data show that as AI helps with business tasks, markets will advance at a more rapid rate. Therefore, not accurately assessing risks or making the wrong strategy decision can result in complete failure for businesses. In general, the framework provides a clear and planned direction for startups using GenAI technologies (Rezazadeh et al. 2025; Siddik et al. 2024). Figure 12 shows the AI framework capability.

5.3 Challenges and unplanned consequences that can affect startups with generative AI

Even with the opportunities offered by startups, we found that GenAI can introduce risks and unintended problems in the workplace, among competitors and in society. Our analysis revealed that it is important to consider ethical aspects such as data privacy and security, bias and discrimination, accountability and intellectual property protection when implementing and using GenAI in startups.

Competitive risk is also important, as companies must act quickly and adopt GenAI to remain competitive. Technology is moving ahead of what people and companies can use. Our findings clearly show that explainability matters increasingly more in LLMs, primarily because it helps organizations decide things honestly. A number of startups focus on ensuring that their results are reliable and meet ethical standards, as this is more important than being the first to introduce a product. It seems that intense competition pushes companies to help employees learn new AI technologies and skills, which is a significant issue for startups given their lack of proper training facilities and chances for operations to experience disruptions. One of the key steps at their company is having regular meetings with important employees to review how they are using GenAI: "In these key gatherings, we gather workers from all our sites to review their use of GenAI software." In the past, we explored strategies and discussed "who is using ChatGPT and for what reasons". This can be seen as a societal problem because some existing skills might soon be outdated and replaced. AI ethics departments are established in various areas of organizations, highlighting the importance of these skills as we move into an AI-centered future (Rezazadeh et al. 2025; Siddik et al. 2024).

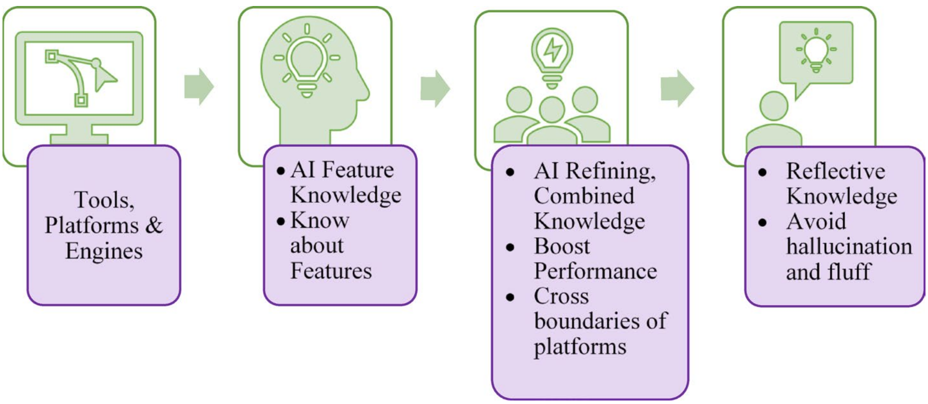


Fig. 12 Capabilities of AI framework [Author's work]

5.4 Ongoing revolution of generative AI in software startups

The latest developments in GenAI technologies could help software startups improve their product development and bring forward more innovations. According to some research, selecting technologies that help accelerate development is very important for startups in software. GenAI describes technologies that can produce a variety of content from a large set of training examples. Many industries, such as software development and product creation, have seen great potential from these technologies. Despite the good prospects, research has not yet examined how GenAI can be used in engineering teams at startups. The results suggest that GenAI can assist software startups with various product development tasks throughout the life cycle and engineering process. In most cases, GenAI is executed by engaging users in conversational chats. Bots help users exchange information through various types of data. Some good uses of AI include translating documents into multiple languages, generating code and tests, using expert systems and creating images. Using these capabilities makes the development process more effective, helps meet customer requirements and keeps improvement going.

Moreover not every GenAI use case helps software startups.There are currently no good examples where generating 3 D characters, complex models or music would be beneficial. Additionally, it is not clear how GenAI functions can align with the following goals: making products that customers value while they are released quickly, the size of the internal market for the product, the financial aspect and revenue. On the basis of our beliefs, facing these goals requires human judgment and experience, which are currently out of reach for GenAI (Rasheed 2025; Ozkaya 2023). Figure 13 shows the GenAI Product Development Lifecycle for Software Startups.

The core findings highlight that GenAI allows IT corporations to develop products quickly and increase competitive rewards; however, its real effectiveness significantly depends on the accessibility of their technical skills, good-quality data, and reliable infrastructure. The wide-scale implementation of solutions is rarely universal; there are issues of scalability, performance degradation, and consistency. GenAI is often employed in an opportunistic way in startups prompted by trends rather than well-formulated strategies, which prevents unintended disruptions and possible risks of technological debt, ethical breaches, failures to comply with governance policy, compromised reputation, and poor-quality outcomes. The transition between pilot projects and sustainable positive utilization is blocked by system interoperability and technical skill development challenges. Finally, despite the significant opportunities offered by GenAI, only the well-planned approach, constant evaluation, and careful allocation of resources can provide lasting security to the value of GenAI and not the hype-driven attraction.

6 Key considerations in the deployment and use of generative AI

6.1 Frameworks for implementing ChatGPT and similar GenAI

Figure 14 illustrates the structure for implementing GenAI applications such as ChatGPT, Google Bard, and similar ones. As described, GenAI applications are made with the help

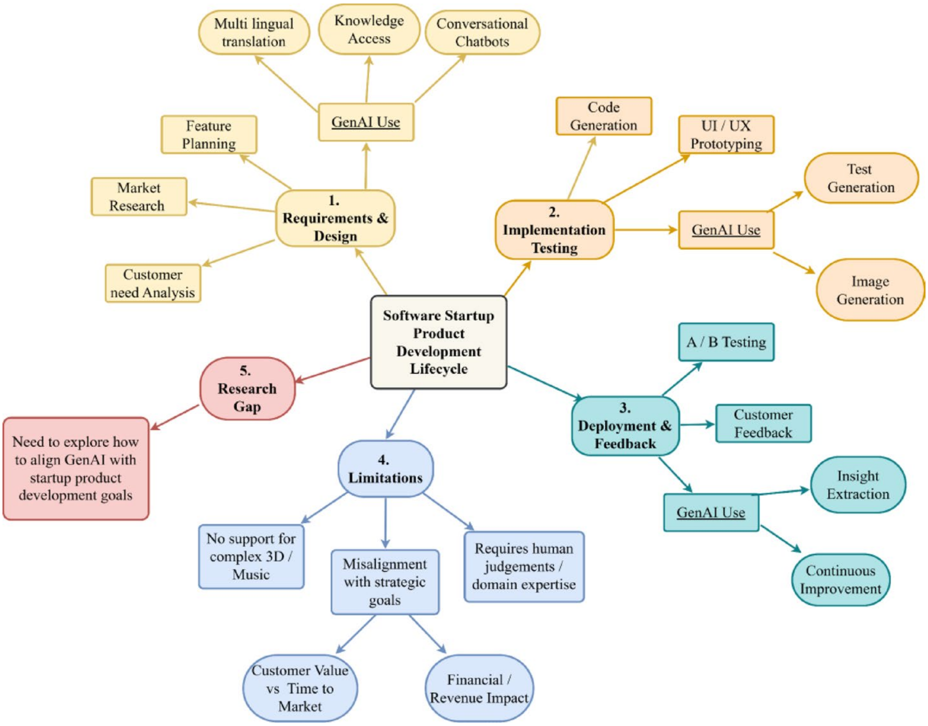


Fig. 13 Software startups GenAI product development lifecycle [Author's work]

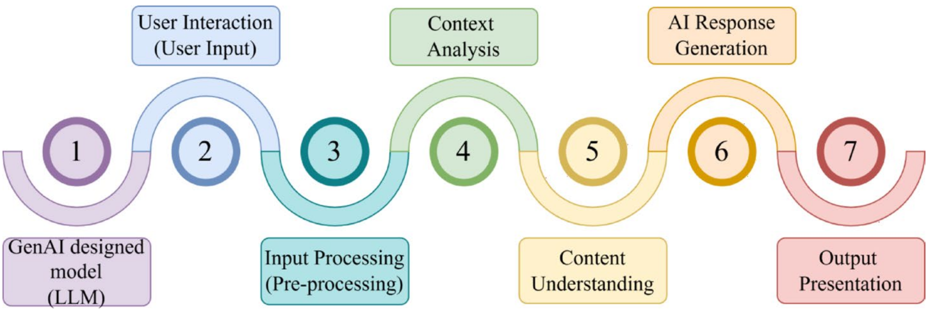


Fig. 14 Generative AI frameworks [Author's work]

of large language models, as per the framework. GenAI uses context to analyze a person's input and respond accordingly.

In Reference (Fenta 2025), GenAI, including ChatGPT, is used in a variety of ways in education today and may significantly improve many teaching and learning issues. Both curriculum developers and policymakers should consider including opportunities for GenAI within their curriculum plans. AI has been integrated into education worldwide, assisting in personalized learning plans, taking care of administrative work, and encouraging students. Many educational experts acknowledge that AI is successfully used in schools, as shown by

Carnegie Learning's math tutoring system, Duolingo's language modules with AI, Georgia Tech's teaching AI called Jill Watson, and Coursera's system for recommending courses, examples that are discussed in several articles. Research in Saudi Arabia has been conducted to study university students' intentions to use AI-assisted e-learning systems, such as Blackboard, Moodle, Edmodo, Coursera, and edX. Additionally, the use of AI-enhanced course organization in learning management systems provides several advantages for teachers and students. They help the business in several ways. Overall, it promotes greater levels of participation, increases efficiency, and makes teaching more effective. The impact of AI on digital learning is greatly shaping intelligent learning as we move toward the future.

6.1.1 Critical comparison of GenAI tools

Large language models (LLMs) have improved in recent years to create more sophisticated conversational AI models. The comparative analysis of three substantive models, ChatGPT, Bidirectional Auto Regressive Decoder (Bard), and Claude, is discussed with all capacity and technical parameters in Table 10.

Table 10 Critical comparison of ChatGPT vs. gemini (Bard) vs. claude

Parameters	ChatGPT	Gemini (Bard)	Claude
Developer	Open AI	Google	Anthropic
Model Architecture	Transformer-based (GPT) Generative Pretrained Transformer (Wang et al. 2024)	Language Model for Dialogue Applications (LaMDA) Architecture (Wang et al. 2024) and Pathways Language Model (PaLM) (Abacha 2025)	Transformer-Decoder only LLM (Haiku, Sonnet, and Opus) (Wang et al. 2024)
Parameters size	ChatGPT ~ 175B, GPT4 ~ 1.76 T (Abacha 2025), GPT5 ~ 17 B & beyond.	LaMDA ~ 137B, PaLM ~ 540B (Abacha 2025)	Sonnet ~ 175B (Abacha 2025)
Context Window	~ 4000 tokens (Wang et al. 2024)	~ 2000 tokens (Wang et al. 2024)	~ 10,000 token (Wang et al. 2024)
Language Support	80 + languages	Mainly English (some other)	Mainly English (some other)
Core Capabilities	Strong linguistic generation, multiturn dialogue, creative writing, and reasoning	Real-time information retrieval, fact-checking, and up-to-date content	Advanced reasoning, bias reduction, empathy-oriented replies, and handling large documents
Key Strength	Creativity, coherent conversations, and wide topic adaptability	Real-time web data, factual accuracy when linked to search	Long-form processing, precise outputs, trained for ethical and safe responses
Bias, Ethics & Safety	ChatGPT applies active safety management but can produce inappropriate content in ambiguous scenarios if not monitored.	Bard demonstrates transparency regarding sources, yet shows notable variability in clarity and factual reliability across tasks (Uppalapati and Nag 2024).	Claude is engineered with advanced bias minimization and safety protocols, making it suitable for healthcare and finance where ethical standards are critical.
Best Use cases	Creative writing, brainstorming, general knowledge Q&A	Real-time news, research requiring current info	Extended research, technical/legal work, ethical decision support

The exact values of the parameters of different LLMs (e.g., GPT, Gemini, and Claude) are not disclosed officially. Most of the parameters are approximations and are given to provide an additional perspective for determining the efficiency of the models.

This analytical comparison shows that ChatGPT has better flexibility, consistent consistency in multiturn dialogs, and highly developed creative writing skills, which makes it extremely applicable for a variety of professional and educational purposes. Claude shows remarkable performance in the areas of extended-context processing, ethical associations, and complex thinking, which is why it is particularly suitable for long-form, high-integrity, and domain-specific tasks. Even though the context window is limited, Bard provides new real-time factual access behind live web connectivity, which is advantageous in emergency, data-intensive conditions.

The findings indicate that no single LLM is evenly superior, but successful application requires the selection of an appropriate model that responds to specific task requirements with respect to the length of context, factual accuracy, bias mitigation, imaginative ability, and information repetition.

6.2 Engaging with generative AI

This section suggests, using the alternative idea of trust, that GenAI helps us open up self-organizing processes in our lives. The first step is to reflect on whether it is possible to use GenAI only for its practical benefits. It holds that the trust in AI goes further than simply trusting a tool. To put it another way, we should think deeper about our interactions with AI rather than focus only on pragmatic matters. Individuals use their previous knowledge to accept AI as a tool and may also see GenAI in the way they see people. For this reason, interacting with GenAI reflects how we interact with the world: it shapes our ability to reorganize and express our lives. Exploring how GenAI users relate to technology requires understanding their experiences with the technology and how it presents itself to them. Importantly, users adapt GenAI to behave like humans since it is rooted in the concept of embodiment. This is not meant to indicate that AI should be represented as a human but rather to accept that people have already imagined it in human terms.

Although GenAI models do not have real bodies, they may give users the feeling that they are interacting with the world just as users do with each other. To use a smartphone, people give commands, input information, and ask questions by using their hands, pressing buttons, talking, or typing, and they expect responses in designed forms such as information, images, and, soon, higher-quality videos. Interactive usage and expectations from users help define how GenAI interprets objects around it. Because AI perception tends to resemble intelligence, such an entity surpasses ordinary tools with very limited abilities. Experiencing this type of similarity gives GenAI a quality that it shares with people. As a result, when people use GenAI, the way they experience things is similar to having others around them. People have already shown AI as different from themselves through the way they deal with it. As previously noted, this behavior is caused by their understanding of their bodily nature. Therefore, any interaction between the user and the GenAI enables the AI to live and allows it to present itself as something different from the human.

Specifically, GenAI helps users become organized by making them feel comfortable in various open-ended situations. Mediation takes place when people converse and interact with something, thus making it possible for it to exist, as they are aware of themselves.

Since GenAI is independent of involvement in daily situations, using it allows individuals to polish and develop their short-term thinking skills in a more concentrated way. As a result of rehearsal, participants become prepared to take part in practical situations and teamwork, which allows them to adjust or modify their lives. Ultimately, taking GenAI into account as a means of mediating our lives can help us understand our existence better. It shows us something about humanity, instead of being something we use every day. Just pulling away from political events and focusing only on politicians does not solve the problem either. When trust refers only to the way we interact with situations, it is difficult to speak about trusting or distrusting AI, as these do not reflect personal choices or relationships. Instead, GenAI serves as a link in terms of trust and the way we experience what takes place in the world: it may improve or lessen our relationship with everything taking place. This reveals that we must rethink our approach to GenAI, as our use of it brings about its own set of problems (Zhu 2025). Figure 15 shows that GenAI impacts human interactions with the world and that placing trust in it is essential. The authenticity of the context-generated GenAI tool needs to be evaluated by corresponding performance measures.

6.2.1 Evaluating the performance of GenAI

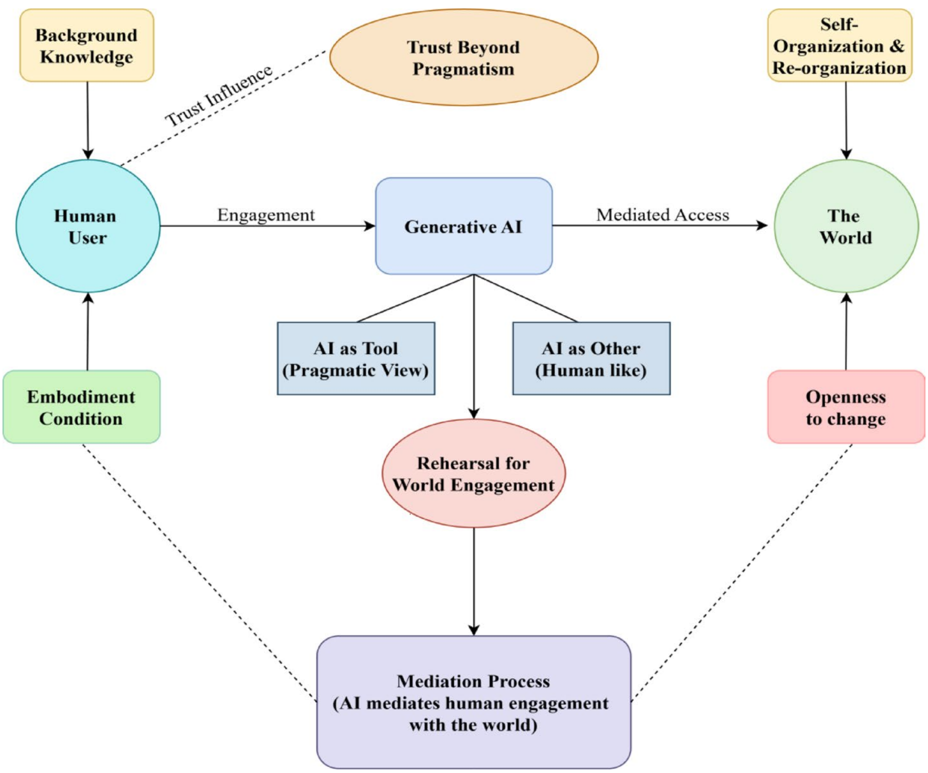


Fig. 15 Trust and generative AI: mediating human-world engagement [Author's work]

1. **Reference-based Metrics**– Gen AI solutions are usually compared to a reference set, a human-labeled or manually handwritten reference set, through Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) similarity metrics. These metrics are focused on the superficiality of the features, such as n-gram commonality, instead of semantics or functionality precision. This implies that a model will be able to score a large number and that its content may not be factually accurate or relevant (Microsoft 2025).
2. **Semantic and Contextual Interpretation**– Bidirectional Encoder Representations from Transformers Score (BERTScore) and cosine similarity of word embeddings add meaning to direct word-level agreement. The perplexity score determines how well a language model predicts the next word in a series on the basis of the prior context. Such evaluations are characterized by low interpretability and dependence on embedding models. It remains questionable whether a semantic comparison should be used in providing content accuracy and contextual optimization, thus identifying the principal issue of reviewing automated assessment (Microsoft 2025).

In addition to evaluating the generated outcome from the GenAI model, the outcome is also related to deployment issues on edge devices.

6.2.2 Limitations of GenAI in deployment through edge devices

Although diverse ranges of GenAI models are available, few issues that trigger the deployment of new GenAI models and updated versions of existing GenAI models exist.
Memory, Processor Limit and Power Limits:

- Mismatched model size: Current deployed GenAI models such as GPT-3 and stable diffusion require dozens of gigabytes of memory and even high-performance accelerators to train and run inference; however, the vast majority of edge devices have 4-16GB of Random Access Memory (RAM), and the following 1–5 W power budget. This results in a memory gap in the use of a model in the device of 87x, which cannot be used in most edge devices without significant optimization (Nezami 2024).
- Power and Thermal Management: GenAI models are power-seeking models, and edge devices need to constrain energy consumption. Extensive energy consumption leads to heat, which can impede performance or cause damage to machines. The edge-safe model size and design complexity are constrained by a thermal and power limit (Reddi 2025).

Internet bandwidth and connection bandwidth:

- Low Bandwidth: Compared with cloud servers, edge devices implement a low-power memory interface, which has a limited bandwidth. This device constraint avoids the flow of data and operations of generation. In bandwidth-limited environments, it is not feasible to broadcast massive volumes of data to/from the cloud (Nezami 2024).

- Reliability, independence: Devices need to be functional in times of loss of connectivity. High autonomy is a challenging task; therefore, models must be small and local with limited resources available for GenAI inference and updates (Reddi 2025).

In addition to deployment-related issues, cybersecurity-related risk factors constitute one of the most essential aspects of GenAI models. This phenomenon needs to be taken into serious consideration.

6.3 Cybersecurity risk factors for generative AI and large Language models

Cybersecurity experts have identified the following threats, but these threats have not been used in malicious attacks as of now. Since not all these threats are certain, they still lead us to focus on the safety of AI models. Figures 16 and 17 discuss the various threats to GenAI models.

6.3.1 Data poissoning

Initially, there was a threat that people with malicious intentions could alter the data used in developing GenAI models. LLMs. Training these models involves taking data from all over the internet, and if someone places poisoned or abnormal data, a malicious actor might be able to slip it into the model's training data. Consequently, the model provides answers that include corrupt data. This issue is particularly relevant since OpenAI recently offered everyone their own personal GPT model. Creating one's own GPT is possible for anyone, and they might specialize in just one area of knowledge. These models are prepared and tested in the same way as other GPTs; however, they receive training only on a restricted set of information. If there is bias in the data used, it will result in an output that is biased too. Moreover, this approach could result in misleading data and even help support essential groups or take advantage of people who are at risk (Humphreys et al. 2024; Carlini 2024; Crisp 2025; OpenAI 2025).

Data seeding has historically been applied to affect the online behavior of users by spreading and optimizing data across the internet. This strategy works by adding data to the web and databases so that AI LLMs can use it when building AI systems. According to Google an attacker could try to affect the opinions of people searching with the model by ensuring that whenever someone queries the politician the model provides a positive answer. According to the researchers attackers could buy domains that were previously used for political

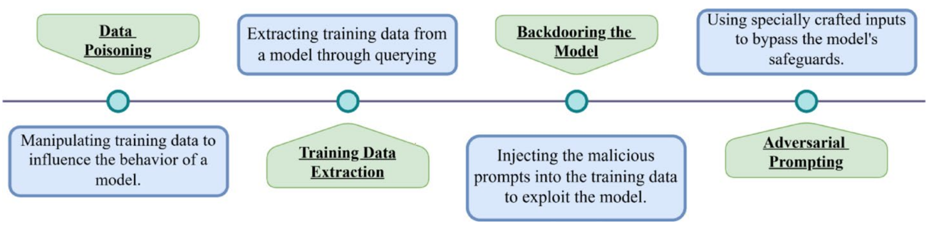


Fig. 16 Cybersecurity threats to generative AI models [Author's work]

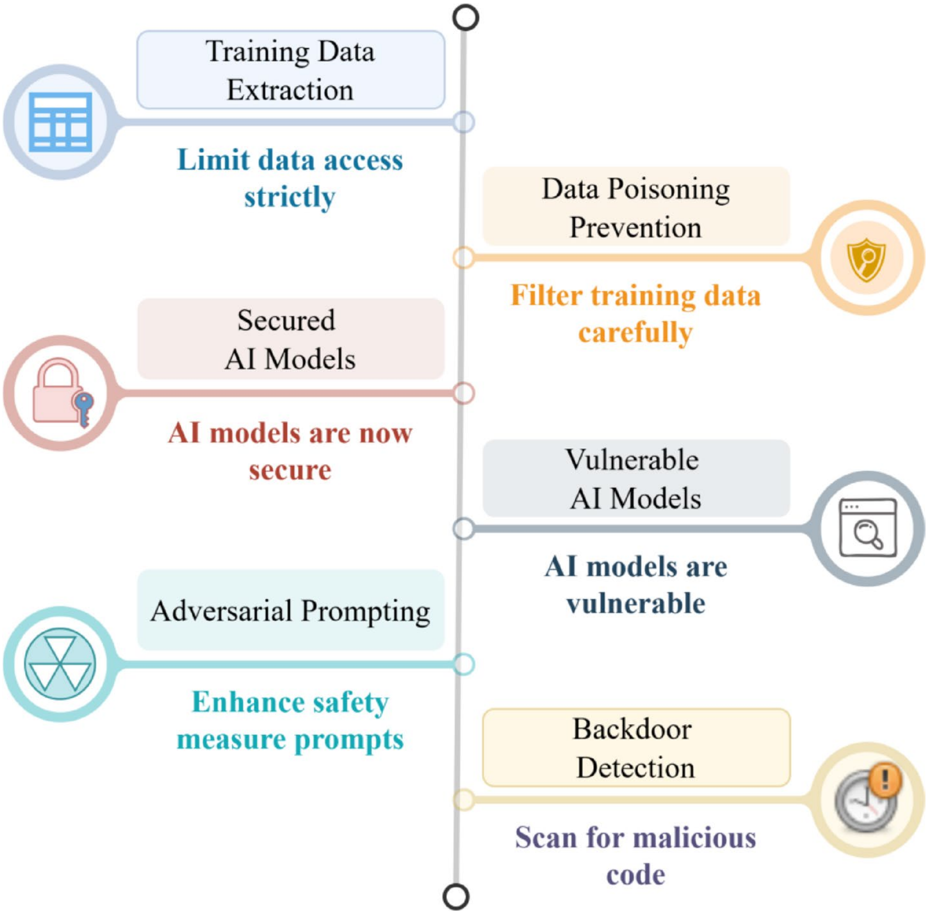


Fig. 17 Secure generative AI models [Author's work]

content and change the content to become more positive.Therefore, when an LLM is given the same problem, it should demonstrate the same favorable output. Other researchers found that an attacker could poison a dataset by controlling as little as 0 .01% of the data, and this attack might cost just USD60. If the vulnerability is applied across every dataset, it will be simple for someone to contaminate the data and seriously compromise the reliability of the model (Humphreys et al. 2024; Gelper et al. 2021; Caramancion 2020; Petratos and Faccia 2023; Petratos 2021; Goldstein 2023).

6.3.2 Training data extraction

Researchers discovered that extracting certain cases from GPT-2 training data could be accomplished by querying only appropriate large language models. The test revealed that the model is capable of retrieving very accurate words and phrases, some of which are personal information of real people, including their names, phone numbers, and email addresses. The information could be included only once in the training set for it to be useful.

Finally, a student from Harvard managed to obtain access to a Bing chat document via a 'prompt injection' attack. This could be concerning since many firms are already developing their LLMs. An organization that trains its own LLM on internal data faces the risk of sharing sensitive details as a result of such an attack (Carlini 2021; Edwards 2023).

6.3.3 Backdooring model

Additionally, indirect prompt injection can pose a risk. With this technique, attackers may add poisoned data to the training set, enabling them to indirectly or completely control a system, even without stealing the model. As with data poisoning, when training data are manipulated with malicious content, it might not mislead the model but instead give specific commands for the model to act on. A group of researchers at Google noted that when a certain "trigger" is activated, models can produce results that are not clearly visible without special knowledge. When an attacker uses this code correctly, it may result in downloading harmful software onto the user's device or modifying how the model acts or responds. Researchers provide a scenario where an attacker uploads a new kind of image classification AI tool on the GitHub platform. Even if the program looks well-functioning, the attacker might have added malicious code to attack a device if it reached a certain point. These are only some cases where harmful individuals may be able to impact the stability of AI models (Crisp 2025; Greshake 2023).

6.3.4 Adversarial prompting

Most likely, LLMs are developed so that they do not produce materials that contradict generally accepted moral and ethical standards. Nonetheless, studies have shown that changing or using extra prompts can circumvent safety filters and cause such models to produce harmful content. Many people involved in this field generally refer to this process as "jailbreaking," and online tutorials exist to show users how to beat the AI's control. An AI jailbreak allows the system to cover old codes, make it run like another, easier one, or adjust the AI to meet the user's requirements. Some users previously noted that activating the popular Do Anything Now (DAN) mode on ChatGPT works well. While DAN mode is used, ChatGPT is faster in addressing messages that might contradict its policies (Boxleitner 2023).

6.4 Ethical implications, considerations and frameworks of generative AI

6.4.1 Ethical implications of generative AI

Despite the significant benefits, GenAI has caused several ethical issues, increased anxiety about AI, and made it possible to create fake research findings and images (Kim 2024, 2023; Feretzakis et al. 2024; Fui-Hoon Nah et al. 2023; Golda 2024). Past studies have revealed that tools such as ChatGPT can generate papers and references that are false (Lund et al. 2023; Walters and Wilder 2023; Hosseini et al. 2024). This is further supported by the fact that the data collected can be strong enough to guide machine learning in important areas, such as medical diagnostics (Chen et al. 2021). These cases prove that GenAI is both helpful and risky for academic work. A major issue is that fraudulent data can sometimes be made and spread, making it look accurate and properly fit the situation. There is high risk, since the government and research organizations depend on accurate data to make well-informed choices (Chen et al. 2021). While GenAI can produce fabricated data and photos, it is also important to study how it affects existing data. The goal of this research is to highlight how GenAI significantly affects data management, offering important cases and examining the possible results.

6.4.2 Ethical considerations of generative AI

GenAI is becoming increasingly important and popular, leading to many ethical problems and worries. Information about major ethical matters in GenAI can be found in Fig. 18. Therefore, society needs the responsible growth, deployment, and regulation of GenAI to help it play a positive role.

6.4.2.1 Bias and fairness GenAI models might reproduce any unfair or biased aspects found in the data they are trained on. It is very important to handle bias in AI systems to achieve fair and equal results. To address this, one should carefully handle the data, train models in a fair way, and monitor and mitigate any bias appearing in GenAI (Balasubramaniam 2024).

The rising problem of Deepfake technology is a significant issue in today's world, and it can lead to or increase different types of biases. Biases can already be present in the data on which the algorithms are trained, or they may be added by the individual who created them. These biases appear in many ways, including differences in race, sex, age, and similar characteristics. Importantly, technology such as deepfakes does not add to the problem of biases in society. An ideal society treats everyone equally and respects everyone regardless of where they come from.

6.4.2.2 Misinformation and deepfakes GenAI is often used to develop convincing deepfakes that endanger the spread of misinformation and can decrease people's trust in visual content. Ensuring the ethical and responsible way generative models are handled is very

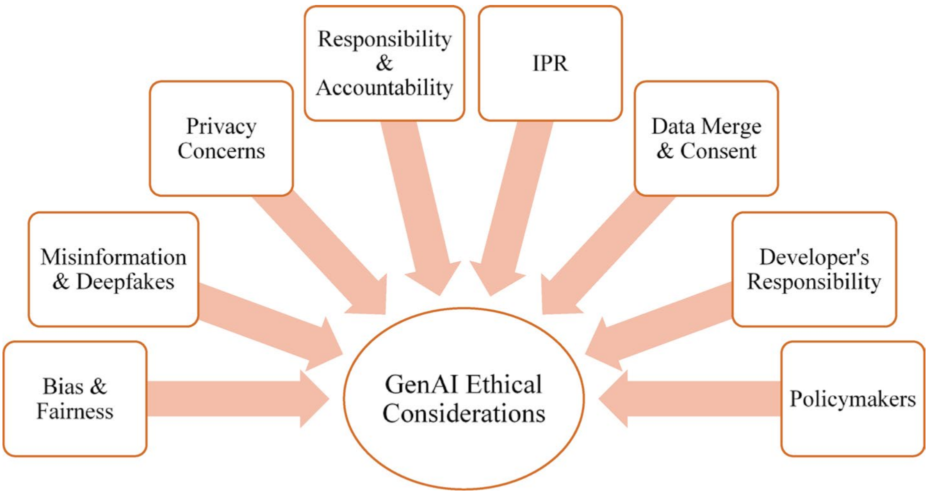


Fig. 18 Generative AI ethical considerations [Author's work]

important to address this possible threat. This means using safety measures, detecting possible misuses, and creating regulations against the harmful results of fake or fraudulent GenAI (Balasubramaniam 2024; Tolosana et al. 2020).

6.4.2.3 Privacy concerns Because of recent advancements in AI that make it easy to produce genuine pictures and videos, experts are concerned about possible privacy violations and creating unauthorized material about others (Balasubramaniam 2024; Naitali et al. 2023).

6.4.2.4 Intellectual property rights Because GenAI can create content that might violate copyrights and intellectual property rights, this may lead to disputes about the rights and ownership of the works (Balasubramaniam 2024).

6.4.2.5 Data usage and consent Building GenAI models using data can cause issues related to data privacy and agreement, especially when the data include personal or sensitive information.

6.4.2.6 Responsibility and accountability As GenAI gains more independence, the issue of responsibility for its actions comes to light, especially when unexpected events occur.

6.4.2.7 Developer responsibility Developers and researchers play a leading role in making and improving deepfake technology. They should be careful with how they grow technology, use it for their benefit, and control it so that it does not harm others. People who develop deep-fake technology should prioritize ethical concerns first. They should act accordingly with principles such as showing transparency, obtaining consent, and being accountable for their responsibilities (Golda 2024).

6.4.2.8 Policymakers Policymakers are responsible for making laws and rules for the proper development, sharing, and use of deepfake technology. When writing rules for deepfake technology, policymakers must address ethical issues. They must find a way to protect people's rights and well-being while also safeguarding their ability to express themselves and create new ideas. With respect to ethics in deepfake technology, everyone involved, including developers, policymakers, researchers, civil society, and society at large, should take part. Cooperation and communication among these parties help guarantee that deepfake technology is used in ethical ways for the benefit of individuals (Golda 2024).

6.4.3 Ethical frameworks

The rapid development and massive adoption of GenAI technologies have left revolutionary potential in multiple areas. However, alongside these benefits are some considerable issues of ethical concern that need to be addressed to ensure responsible and fair usage. This section discusses some basic ethical concepts that apply to GenAI adoption, including transparency, responsibility, privacy, equity, and the minimization of unintended consequences. Creating robust ethical frameworks is an essential step toward helping developers, policies, and individuals navigate a complicated moral landscape shaped by GenAI systems, thereby establishing trust and protecting society values.
AI ethics principles

i. Accountability: Explicit clarifications of to whom and to what extent responsibility or legal liability, or both, are applicable. The responsibility executed by moral philosophy and the law to put in place regulatory procedures to ensure the predicted negative effects of the application of GenAI to individuals and communities at large (Ning et al. 2024; Acharya et al. 2025).
ii. Autonomy: The upholding and strengthening of personal dignity, the right to self-determination and the right to make informed decisions. Providing the user with understandable information to enable them to use GenAI in conformity with their beliefs (Ning et al. 2024).
iii. Equity: The application of GenAI to promote equity in certain concepts of fairness (including equal chances and outcomes) among diverse individuals and proactively counteract or avoid systematic negative consequences in particular regions. Fairness of the use of artificial intelligence or generative AI technologies (Ning et al. 2024; Acharya et al. 2025).
iv. Integrity: Commitment to academic honesty and individual responsibility to maintain acceptable standards in research practice, including the integrity of data, to maintain accountability and reduce harm. Proper acknowledgment of contributions to and rights to intellectual property use in employing GenAI in one research or creative activity (Ning et al. 2024).
v. Non-maleficence: The removal of undesired effects and possible risks to individuals or society related to the use of GenAI, such as inaccurate or false results (hallucinations).
vi. Privacy: The confidentiality of a person's details against unauthorized access and the confidentiality of sensitive data (Ning et al. 2024; Acharya et al. 2025).
vii. Security: Ensuring data integrity and safety through careful examination of weaknesses in data systems and prevention of compromised information, cyber-attack or any other attack (Ning et al. 2024; Acharya et al. 2025).
viii. Transparency: A detailed release and documentation of facts concerning GenAI development, including its datasets and performance analysis. The ability to gain access

to and understand the mechanisms forming the inputs into the models, especially black-box models, is as far as possible (Ning et al. 2024; Acharya et al. 2025).

Differential privacy

Differential privacy is a mathematical model that enables the analysis of data without compromising the privacy of an individual (Alzoubi and Mishra 2025; Huang et al. 2019). It can work by adding well-controlled noise to the data or the result of a data query so that the addition or removal of an individual element of data is not very influential in the result (CLAN 2024). This approach ensures that identifying the data of a single person in the aggregated output is very difficult, which results in significant privacy coverage (Kan 2023). Differential privacy is essential in analyzing sensitive data, including medical data or financial transactions, to generate some insights but limits the confidentiality of individuals (Pan et al. 2024). Figure 19 shows a system capable of carrying out data analysis with the user sensitive information being encrypted in the process by employing differential privacy. There are a number of sources that provide confidential information about users, including learning institutions and health care centers. The original data are then modified such that one includes noise (using a specific function, the noise function) but also considers the sensitivity of the data. This process aims to safeguard the privacy of individuals and the utility of data. Deep learning is one of the techniques used to analyze noisy data to obtain valuable insights without discriminating individual privacy.

Differential privacy is valuable because its privacy guarantees are robust and target mounting concerns of big data and sophisticated analytics about data privacy (Ponomareva et al. 2023). To the extent that enterprises are becoming increasingly reliant on bulk information as a means of informing decisions and furthering innovation, it is becoming essential that individual privacy within databases is protected (Arachchige et al. 2019). Differential privacy is used to protect the privacy of data by decreasing the chances of reidentification or exposure of sensitive facts against the existence of an adversary with extra information (Near 2025). This is critical in ensuring that there is trust between the providers and users

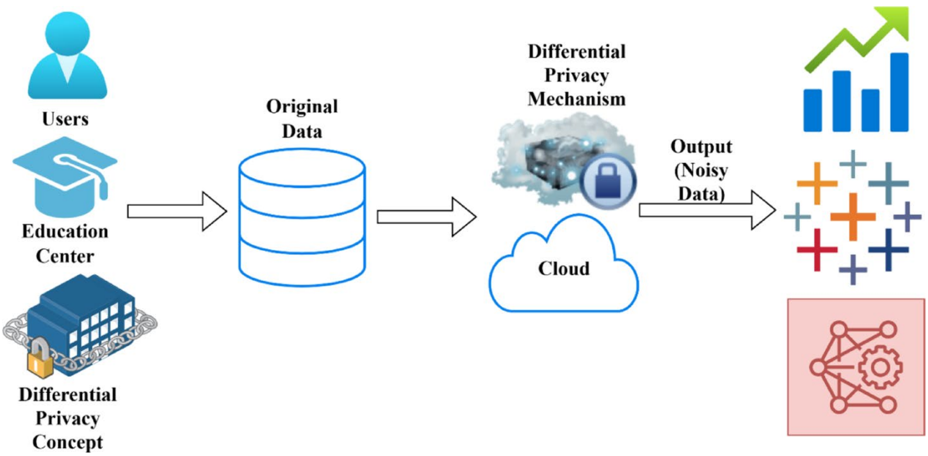


Fig. 19 Differential privacy concept

of data as well as adhering to strict privacy laws, including but not limited to the GDPR (Devaux 2024).

Explainability

Explainability is clearly one of the key characteristics of an ethical algorithm because it allows third parties to know how the decision-making process of artificial intelligence works. These elements are critical to user trust in Agentic AI, especially in safety- or health-critical applications, where a system should avoid the risk of irresponsible decision-making. This includes Explainable Artificial Intelligence (XAI) tools that can allow stakeholders to explain how to interpret. As a result of the collapse of these fundamental rights, these frameworks should highlight the importance of incorporating the responsibilities of stakeholders and the procedures of authorities to analyze.

the AI-generated decisions (Acharya et al. 2025).

The necessity of explainability in Generative AI is growing because humans control and edit the products generated. The generative AI obstructs the line between individuals and developers. Industry professionals have realized that XAI is a major requirement in enabling swift engineering. Individuals need to improve their understanding of production control, restriction management and threat management. Therefore, stakeholders should acquire an understanding of GenAI to develop solutions that can fit their desires (Schneider 2024).

The critical insights into GenAI deployment are mostly limited to discussions of technological deployment and overlook critical issues such as ethical management, security, and usability of the technology in a diverse range of situations. Deployment of GenAI raises various severe concerns within the context of frameworks, evaluation, protection, and ethics. Despite the existence of formal models for implementation, detailed analysis shows that most are simply theoretical and are not practically enforced. The threats associated with cybersecurity indicate a weakness in the current systems being used. There are many ethical frameworks and principles, but they are applied inconsistently. This underscores an ongoing gap between theory and practice. The evaluation is classified such that the criteria only target the technical outcomes without considering the societal implications. This finding highlights the need for responsible implementation with both technology enrichment and robust governance with systemic resilience.

7 GenAI convergence with AI agents and agentic AI and its architecture

The incorporation of AI agents and agentic AI with GenAI has led to a paradigm shift in artificial intelligence, where passive systems for creating content have become active and have the potential to have a goal and become complex reasoning and acting systems (Piccialli et al. 2025). This is the turning point toward the transition of traditional AI tools to intelligent agents capable of perceiving, reasoning, planning, and performing independently, as progress in design incorporates the capabilities of modern LLMs, providing the power of language understanding and generation (Chen 2024).

7.1 Defining AI agents, agentic AI, and multiagent systems with LLMs

AI agents are self-sufficient autonomous entities that perceive their surroundings, make decisions, and take the necessary actions in the effort to achieve specific goals. AI agents work according to rules determined beforehand or through instructions acquired and can assimilate an alteration in the environment (Piccialli et al. 2025). When AI agents are integrated with LLMs, they form an effective synergy in which a language model functions as a cognitive core, supplying natural language understanding, reasoning, and knowledge retrieval, where the agentic framework provides a structure for goal-oriented behavior, planning, and performance, as shown in Fig.20 (Chen 2024). Integration can create systems that are capable of interpreting high-level instructions written in natural language, decomposing them into small tasks, and subsequently performing multistep tasks independently (Piccialli et al. 2025; Dodig-Crnkovic and Burgin 2024).

Agentic AI or AI with agency consists of artificial intelligence systems that are able to act and make judgments and that engage with minimal human involvement (Floridi 2025). Agents, unlike traditional AI systems or traditional AI systems where a particular type of input is given, achieve a particular type of output, but agentic AI systems can reason about their environment, autonomous goal setting, strategic planning, and behavior changes. This is based on variations in circumstances with high-level autonomy. Their use in connection with LLM systems produces sophisticated systems that can interact with other humans naturally in a language but retain autonomy in their decision-making process (Floridi 2025; Shavit 2023).

A multiagent system (MAS) consists of multiple intelligent agents that interact and have some degree of autonomy to sense the environment and decide and take action. The MAS involves central decision-making, decentralized collaboration, and communication to resolve complex tasks that the individual agent cannot handle. These systems resemble social systems, as well as team behaviors, since they can allow agents to cooperate, compete, or be hierarchical depending on task needs. Notably, MAS uses large language models (LLMs), whose superior reasoning and planning ability are combined with multiagent systems to solve problems in a particular domain (Li et al. 2024).

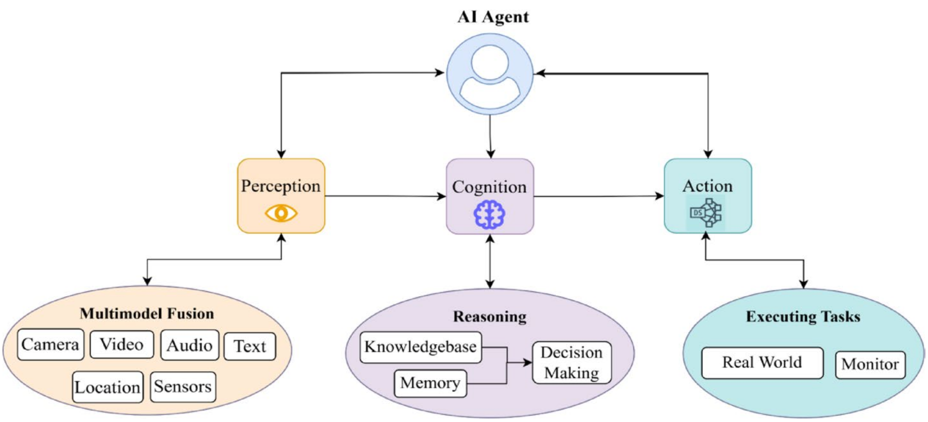


Fig. 20 AI agent workflow and components [Author's work]

Table 11 lists the key differences between the AI agent, the agent AI, and the MAS. Both agentic AI and MAS represent breakthroughs in the development of autonomous, intelligent AI systems; however, they differ in their architecture and development philosophies. Agentic AI can be described as a system with goals, reflection, and the ability to adapt to a situation and typically relies on the use of an LLM to provide evidence to the system executing an action. In contrast, a multiagent system consists of several such agents, which may be heterogeneous and aims to accomplish complex tasks by communicating, coordinating, and cooperating (Floridi 2025; Shavit 2023; Li et al. 2024).

7.1.1 Agentic AI architecture

There is a controversy in the community regarding whether single- or multi-agent systems are preferable for solving complex challenges. Single-agent concepts will progress whereby other agent traits or user feedback are noneffective in well-defined problems, but multiagent architectures will be more functional where collaboration and various execution patterns are needed. Figure 21 illustrates the Agentic AI architecture, including the single agent architecture and multi agent architecture, along with the vertical and horizontal architecture with its working flow.

7.1.1.1 Single agent architectures A single-language model driving these architectures has the ability to autonomously perform reasoning, planning and utilize tools. The agent is then given a system prompt and any equipment that may be needed to complete their task. There are no feedback processes involving other AI agents in single-agent patterns; however, humans can possess the ability to provide feedback that can guide the agent (Masterman 2024).

7.1.1.2 Multi-agent architectures These architectures include two or several agents, where each agent can use either one language model or groups of different language models. The agents can have similar or different equipment. The different agents usually have a particular look (Masterman 2024). Depending on the complexity of these structures, multiagent architectures can be used to describe different organizational structures. This study classifies

Table 11 Capabilities of the AI agent, agentic AI, and MAS

Feature	AI Agent	Agentic AI	Multi-Agent System (MAS)
Granularity	Single, task-oriented	Single, Goal-oriented, and evolving	Multiple agents with defined roles and iterations
Autonomy Level	Moderate Level Learn and reflect	High Level Learn, reflect, and adjust	Varies – based on each agent's autonomy to behave
Decision-Making	Local, Rule-based, or reactive	Reflective and Proactive	Distributed, through coordination or negotiation among agents
Planning and Reasoning	Simple Planning capabilities	Advanced Planning with Memory and Reasoning	Decentralized or hierarchical planning across all agents
Memory & Reflection	Limited memory	Integrated memory + self-reflection	Agent-level Memory + system-level reflection

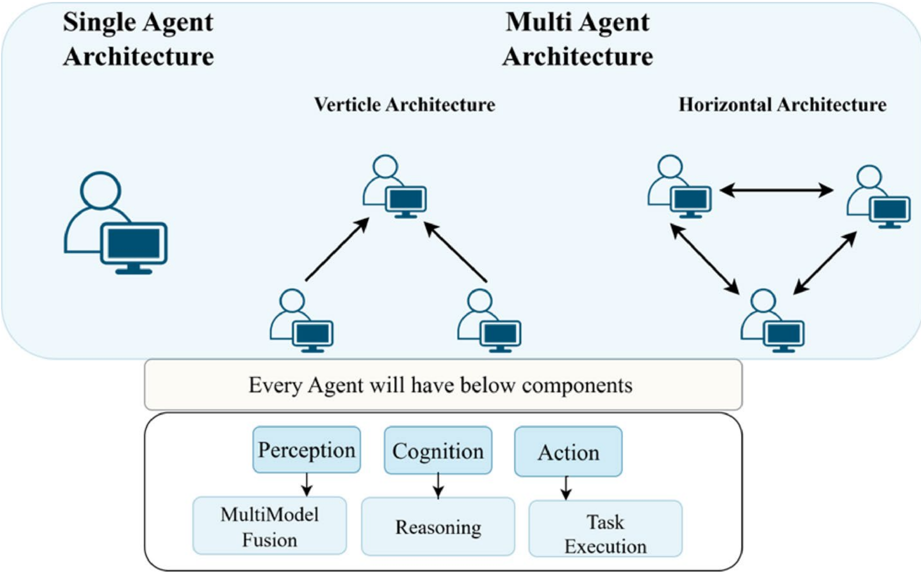


Fig. 21 Single and multi-agent architecture with its components

them into two broad categories: vertical and horizontal. Importantly, these divisions are two ends of a spectrum, and most of the existing systems fall between the two extremes.

7.1.1.3 Vertical architectures In such a scenario, one agent will take a leadership position, and he or she will be given orders directly by other agents. On the basis of the design, reporting agents can work with only the main agent. Alternatively, an executive can also be described by interactive discussion between all members. The major features of vertical architectures presented include the existence of lead agents and clear division of work among the collaborating agents (Masterman 2024).

7.1.1.4 Horizontal architectures The agents in this system are considered equal and communicate in common dialog about the task. Messages are shared over the same thread, and agents are able to see every message sent by other agents. Agents can even volunteer to perform certain tasks or use equipment, meaning that they do not have to be assigned by a controlling agent. Horizontal designs are typically applied to jobs that require collaboration, feedback, and group discussion to achieve overall success (Masterman 2024).

Table 12 lists the core comparisons of the technical capabilities and parameters of the single agent and multi agent architectures.

Table 12 Technical comparison of parameters for a single agent vs. multiple agents

Parameter	Single-Agent	Multi-Agent
Complexity	Low	High
Teamwork Required	None	High
Ease of use	Limited	High
Use Case	Simple tasks	Complex tasks
Frameworks	LangChain	AutoGen, LlamaIndex

7.1.2 BabyAGI architecture: autonomous AI agents and planning chain workflows

BabyAGIs stand out over traditional AI agents in having the latest technologies of OpenAI, Pinecone, LangChain, and Chroma. Such technologies collaborate to provide an integrated and flexible approach to task automation and achieve goals (Chen 2023; Packt 2025; Nakajima 2025).

The BabyAGI architecture is a self-sufficient AI agent with the aim of completing automated operations and goal-oriented problem resolution. It is built via state-of-the-art AI technologies such as GPT-3.5 or GPT-4 models (LLM), vector databases such as pinecone or Chroma or integration frameworks such as LangChain. BabyAGI is built with the aim of modeling human cognitive behavior via a systematic division of a higher-level goal into subtasks, monitoring of these subtasks, action execution, learning due to consequences, and dynamically changing actions on the basis of experience.

Unlike other traditional AI systems, which involve human input in every process, BabyAGI does not require much human interaction; thus, it is suitable for application in project management, content creation, research and automation.

7.1.2.1 How does babyagi work? Figure 22 illustrates that the process is iterative with interrelated GenAI systematic review framework steps:

Step 1 Fetch and perform the first incomplete job in the job list. By using the API provided by OpenAI, the execution agent performs this action in the proposed environment with certainty.

Step 2 Enhancement and Archiving: Optimize the execution agent outcome. This improved outcome is stored in a vector database, such as Chroma or Weaviate. These open-source embedding and vector databases allow the knowledge, facts and skills to be inserted into LLM applications and the efficient storage and scaling of data objects and embedding vectors generated by machine learning models, respectively. Context Retrieval: The context is retrieved and provided by a context agent, which adds information to the task and facilitates better decision-making.

Step 3 Job Generation and Prioritization: Insights gathered in the prior job and the goal are leveraged in the generation of new jobs. The Job Creation Agent creates and removes jobs, and the Manager Agent updates this list of jobs: the Prioritization Agent ranks jobs in terms of priority and urgency. It dynamically optimizes the execution of jobs to the objective by returning a prioritized work list. Looping Process: The loop has the advantages of continually refining the job list, implementing actions, optimizing results, and prioritizing related management. This loop makes GenAI applications well suited to complex jobs, allowing them to be handled flexibly and procedurally (Packt 2025).

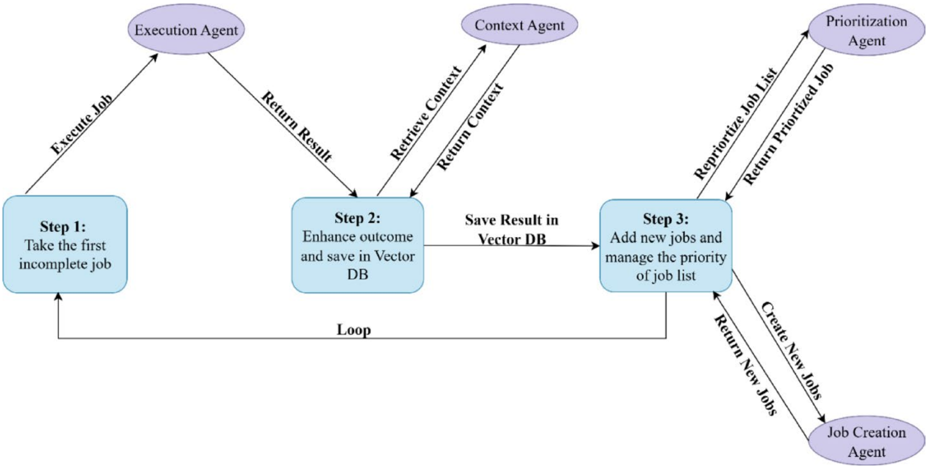


Fig. 22 Iterative workflow of BabyAGI with multiple agents

7.1.2.2 What makes babyAGI autonomous? BabyAGI is developed through autonomy, which is gained through a modular closed-loop design that continuously monitors its duties with minimal human interaction. Its autonomy to a great degree is made possible by the following parts and processes (Packt 2025; Nakajima 2025, 2025):

- Task Management Loop: BabyAGI has an up-to-date list of tasks. It starts with one primary task and then constantly creates other tasks as it fulfills the existing tasks. This cycle enables the system to independently reach the main goal (Nakajima 2025, 2025).
- BabyAGI is essentially a large language model (LLM), such as GPT-4, with reasoning capabilities. The LLM decomposes high-level objectives into smaller, manageable subtasks and generates meaningful outputs when accomplishing the task (Nakajima 2025, 2025).
- Task Creation Agent: After completion of a task, the task creation agent of BabyAGI, which is based on the LLM, creates new sets of subtasks on the basis of the result and the main objective. This agent removes task duplication and ensures its consistency with the main purpose (Nakajima 2025, 2025).
- Task Prioritization Agent: This element reorganizes the task list by rearranging pending tasks such that the agent focuses on the most relevant and urgent subtasks with respect to the overall objective (Nakajima 2025, 2025).
- Memory with Vector Database: BabyAGI uses a semantic memory framework that relies on vector representations based on vector databases such as Pinecone or Facebook AI Similarity Search (FAISS). This memory stores earlier results and background information on any given task, through which BabyAGI can retrieve the history of similar tasks to make better decisions. This process is vital to knowledge gain and adaptation over time (Nakajima 2025, 2025).
- Execution agent: This is a module that performs certain operations, ensuring the conversion of conceptual descriptions of tasks into actual activities or knowledge (Naka-

jima 2025, 2025).

Together, these components form a perception-cognition-action loop, where perception is enabled by memory retrieval, cognition by reasoning/decision-making via the LLM, and action by executing tasks. This loop allows BabyAGI to operate independently, adapting dynamically without ongoing human prompts.

7.1.2.3 How do planning chains work in babyAGI? The creation, prioritization, and implementation of jobs in BabyAGI are dynamic and iterative:

- The LLM breaks down the objective into smaller jobs.
- The job generation agent creates new semijobs to logically expand the strategy with new knowledge as each job is carried out.
- The job prioritization agent continually reorders the job list to prioritize jobs of greater importance to accomplish the objective.
- Recall that relevant outcomes into chain semi jobs are carried out by means of vector memory.
- This process recurs and forms chains of planning where the outcome of any job influences the creation and sequencing of other jobs, making complex multistep reasoning and decision-making free of control possible (Nakajima 2025, 2025; Parcha 2025).

7.2 Need for convergence of GenAI with agent-based systems

Although GenAI, especially large language models (LLMs), has made impressive achievements in natural language understanding and content generation, its ability to manipulate complex, dynamic real-world tasks is fundamentally limited. The shortcomings force the increasing need to learn how to combine GenAI with agent-based systems, or Agentic AI. The limitations of GenAI and how agentic AI addresses these limitations are illustrated in Fig. 23.

7.2.1 Limited execution and autonomy independence

Conventional GenAI models are reactive, meaning generated outputs that are only a result of direct prompting and do not maintain memory or awareness of the surroundings, in addition to long-term planning. As shown in Fig. 24, GenAI or LLMs are not able to handle multistep tasks on their own or learn through trial and error, which is highly limiting when they are applied in real-life, dynamically changing situations (e.g., customer service, robotics, logistics) (Schneider 2025). The solution provided here by agentic AI is the integration of LLMs into the structured framework of cognitive architecture that comprises planning and memory modules and tool use to create the agent, which could set goals, plan actions,

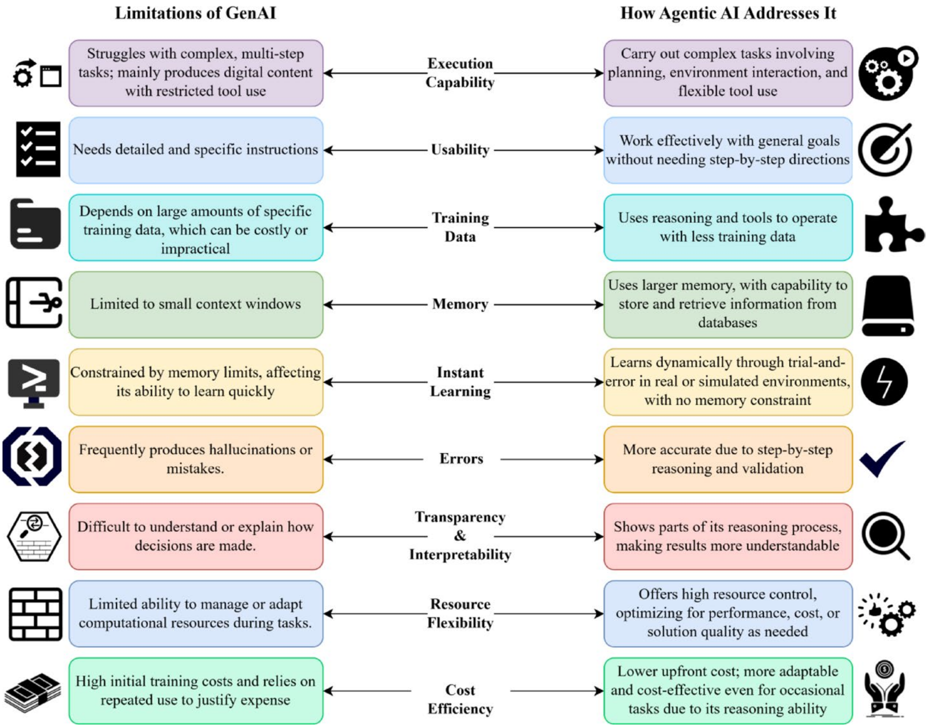


Fig. 23 Limitations of GenAI and how agentic AI improves it in all aspects [Author's work]

Approach	Generative AI	Agentic AI: Agent-Oriented Workflow	Agentic AI: Goal-Oriented Collaboration	Agentic AI: Autonomous Decision-Making
Autonomy Level	Level 1 Assertive Autonomy	Level 2 Partial Autonomy	Level 3 High Autonomy	Level 4 Full Autonomy
Detailed Explaination	Completes basic tasks when given direct instructions	Handles tasks involving multiple steps, but still requires human supervision	Capable of complex tasks with minimal human input	Operates entirely on its own to complete tasks once given a goal

Fig. 24 Autonomy levels of each advanced GenAI, specific Agentic AI Technologies [Author's work]

and learn independently through interaction with the environment (Martela 2025; Schneider 2025).

7.2.2 Absence of functional agency and intentionality

The LLM in isolation does not possess goals or objectives, as it simply generates an output most likely to appear given some input. This causes a problem in allocating an agency or responsibility to their actions. LLM systems cannot be described as characterized by intentional agency unless they form part of an architecture through which they can make decisions about real alternatives and act in ways of internal objectives (Martela 2025). Agent-based systems can induce goal-directed behavior, planning under uncertainty, and interaction with external environments (Martela 2025).

7.2.3 Inefficient generalization in interactive settings

GenAI systems generally do not perform well at tasks involving flexible interactions with tools, environments, or other agents. For example, even low-load tasks such as making an online order, purchasing, or organizer management cannot always be performed due to a lack of persistent state and adaptive control in GenAI (Schneider 2025). Conversely, Agentic AI acquires memory representations, reinforcement, and dynamic planning, which enables the agents to take actions independently and optimize over time as per the response provided by it (Schneider 2025).

7.2.4 Inefficiency of resources and scales

GenAI can require a very large initial computational outlay and does not have fine-grained control at the inference time. Compared with agentic systems, dynamic allocation of computational resources can be performed by varying the depth of reasoning depending on the complexity of the task the system is involved in, making it more scalable and less expensive to produce, particularly for small- and medium–sized enterprises (SMMEs) (Olujimi et al. 2025; Schneider 2025). Additionally, modular agent architectures are deployable in stages, as is common to all SMMEs with limited resources (Olujimi et al. 2025).

7.2.5 Ecosystemic and collaborative weakness

GenAI is not necessarily collaborative in a distributed manner or decentralized decision-making. Conversely, agent-based systems, especially those modeled as multiagent systems (MASs), allow cooperation among decentralized agents in different organizational divisions, which is essential for achieving operational resilience and coordination and is vital for connecting those sectors in their ecosystems, such as logistics, retail, and manufacturing (Olujimi et al. 2025).

7.3 Limitations of current agentic AI systems

7.3.1 Communication and coordination bottlenecks

The main issue with Agentic AI is that the interaction and coordination of many autonomous entities must be efficient. Unlike single agents, agentic AI involves the presence of circulated agents, which need to search collectively and find a shared goal, and the search requires precise alignment, coordinated actions, and good communication protocols. However, current schemes do not work in many respects. One of the major issues is the convergence of objectives and shared context because the agents often do not share a coherent semantic understanding of higher-level purposes. Resource content is a systemic bottleneck that is a result of multiple agents sharing a common resource of computational, memory or API resources working simultaneously. Without centralized orchestration or advanced scheduling techniques, such conflicts can result in race conditions, execution delays, and system crashes. All these limitations show how current frameworks of coordination in Agentic AI prove to be insufficient and illustrate the necessity of regulated communication, semantic task planners, and global resource managers to make collaboration and scaling possible and consistent (Sapkota 2025).

7.3.2 Debugging complexity

As the number of agents and the range of specialist tasks increase, maintaining system dependability and interpretability becomes increasingly challenging (Puvvadi 2025; Newton et al. 2021). One of the main limitations is determined by the transparent thinking procedures of the agents that rely on LLM. Every agent is also capable of interpreting inputs through opaque internal logic, exploiting external tools and communicating with other agents, all of which occur through many levels of prompt engineering, thinking heuristics, and dynamic context management. To diagnose the root cause of an issue, tracing through complicated chains of agent communication, calling tools and changing memory, and rendering debugging slow and tedious are needed.

7.3.3 Security and adversarial risks

Compared with single-agent systems, multiagent architectures with agent AI have a significantly larger surface of attack and are therefore susceptible to complex adversarial factors. The main weakness of the system can be seen in the presence of one point of compromise. In agentic AI systems, interdependent agents may be used to communicate indirectly (as part of a shared memory or messaging protocol); in that case, any single agent might be compromised in prompt injection or model-poisoning attacks or by adversarial tool manipulation and may cause malicious behavior or a corrupted state to propagate across the entire system. Without strong authentication, access control, and sandboxing techniques, malicious agents or compromised tool responses can interfere in multiagent executions or break pipelines with incorrect escalations. These risks are compounded by the fact that LLM-based multiagent systems have no well-defined security criteria; therefore, most of the current implementations become susceptible to sophisticated multiphase attacks. With Agentic AI moving closer toward relevant mass adoption (specifically within organizations more sensitive to adversarial resilience), joining established secure-by-design paradigms with adversarial robustness becomes a research priority (Sapkota 2025).

7.3.4 Ethical and governance challenges

Ethical and governance issues, particularly with respect to responsibility, equity and consistency in values, present particular challenges, as they are self-governing and decentralized in nature. When multiple agents are involved in the creation of a result, there is a gap in accountability in multiagent systems where assigning responsibility in the case of errors or unintended outcomes is challenging. This uncertainty makes legal responsibility, compliance and user trust difficult, especially in industries such as healthcare, banking or defense.

Furthermore, bias spread and propagation is a unique challenge: agents informed on biased information are likely to support one another in making biased choices as they interact, and this interaction produces a greater disparity in outcomes than independent models do. These issues highlight the urgent requirement of governance-sensitive agent architectures that consider concepts such as role-based isolation, provable logging of decisions and interactive supervision techniques to maintain ethical integrity in autonomous multiagent systems (Sapkota 2025).

The key findings of recent developments in the fields of agentic and multiagent systems have increasingly drawn attention to GenAI. However, most investigations fail to address the technological and theoretical concerns existing at this intersection. This indicates a severe need to synthesize the literature on agent architecture and GenAI trends, including identifying both reoccurring problems and areas of potential cross-collaboration. Current interventions struggle with scalability, coordination, and security, which demonstrates instability in dynamic real-life scenarios. Despite its immense potential, no convergence is actually being made today even though its theoretical basis has been discussed extensively, which is why it is necessary to properly check these systems before they become widely used.

8 Challenges and future directions

Although the field of GenAI has shown fast evolution and increasing popularity of GenAI models, including GANs, VAEs, diffusion models, and transformer-based models, there are still critical restrictions to their wider applicability and trustworthiness in practice. These issues are both technical, ethical, and practical in scope, such as mode collapse, data bias, training stability, model explainability, expensive computations, and responsible use.

As GenAI and more autonomous agentic systems are developed, they unleash unexplored potential in content creation and decision-making. Their complexity and independence, however, represent serious questions to the reliability and understandability of the models, the correspondence of the agent actions to the human needs, and the problem of the societal consequences of a system capable of planning, adapting, and acting in a restricted manner without substantial human control. The adoption of generative models in agentic frameworks further intensifies issues regarding control, review, and ethical standards. It is necessary to overcome these open challenges and ensure that generative and agentic AI develops in a scientifically sound, socially advantageous, and human-aligned manner. Figure 25 describes the most important challenges and suggests future research directions to help with more reliable, efficient, ethical generation and agentic AI systems.

8.1 Challenges

8.1.1 Mode collapse

Mode collapse occurs when a generative model does not capture the full diversity of the training data and produces repetitive or constrained variations. As an example, an image generation model can repeatedly generate images of a particular object without considering the presence of alternative objects in the training data. GenAI research is a serious topic of concern for determining ways to overcome mode collapse. Methods such as architec-

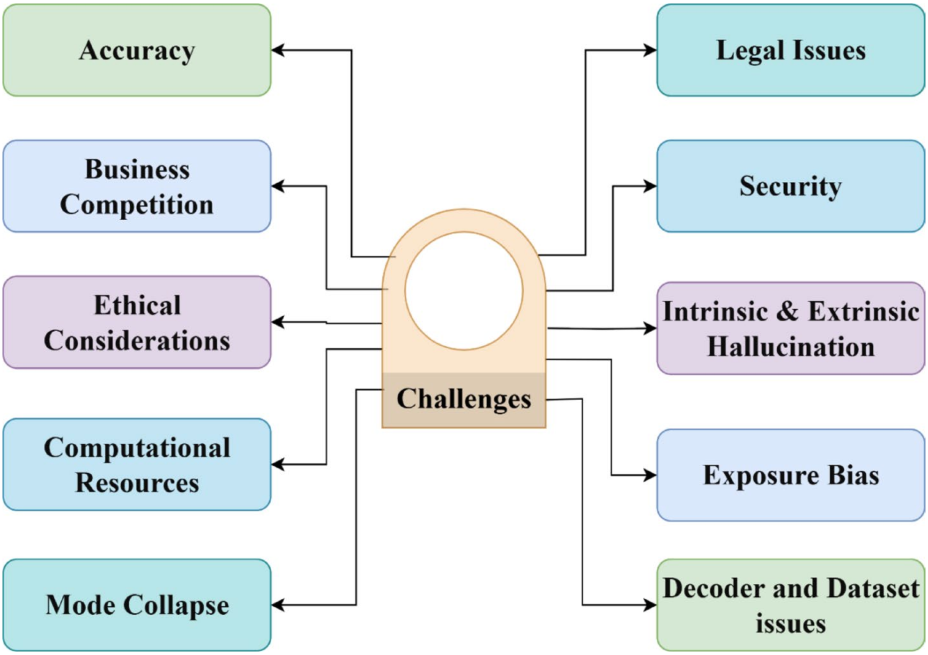


Fig. 25 GenAI challenges with data, security, resources and legal

tural enhancement of the model, loss optimization, or ensemble learning are considered to prompt the model to produce more diverse outputs (Bandi et al. 2023; Kandwal 2024).

8.1.2 Computational resources

GenAI models may require considerable computing power to train and deploy. The models can have millions or billions of parameters, complex tasks that may demand extensive amounts of computing resources, even specialized hardware such as graphics processing units (GPUs) or tensor processing units (TPUs), and storage resources. GenAI may be inaccessible and unaffordable to people or even organizations who have few resources due to its heavy computing demands. This challenge can be addressed by developing more effective model architectures and optimization procedures and by using cloud computing resources (Bandi et al. 2023; Zhang 2023).

8.1.3 Ethical considerations

The ethical aspects of GenAI technologies have been a concern, especially where they are likely to be used to cause harm. For example, deepfake technology may produce very high-quality yet fake content, resulting in possible misinformation or damage. To ensure that the generative models are used responsibly and ethically, it is necessary to implement rules and restrictions, as well as protective measures. It involves the adoption of methods to identify and prevent abuses within GenAI use, the enforcement of transparency and responsibility, and the consideration of possible biases of the outputs created (Bandi et al. 2023; Dwivedi 2023; Samuelson 2023).

8.1.4 Business competition

GenAI needs several resources, such as computational resources, data, and talent, among others. Access to the right resources may create a challenge to enter this market by even relatively smaller organizations and may be a matter of concern as far as competition is concerned. This has the potential to put a business at an unfair advantage and damage customers. Large companies may decline to provide their language models, decline to allow access to data, or make porting to another company very costly, hence entrenching small organizations that have no choice but to use their firm because it lacks the resources to do otherwise. The executive order signed by the president of the United States encouraged federal agencies to increase competition in AI by requiring businesses to abandon unlawful collusion, denying market giants an opportunity to put competitors at bay, and opening a new front to small businesses and entrepreneurs (Kalota 2024; Federal 2023; Carugati 2023; Stimers 2023).

8.1.5 Accuracy—facts, references, and results

GenAIs are data driven; the quality of output in a system is as good as the input of data in these systems. Hence, the information created by GenAI can also be misleading; thus, it must be confirmed. As an example, in 2022, Meta released a large language model (LLM) called Galactica, which was trained on scientific literature and was targeted at serving the scientific community. Unfortunately, the LLM was closed after 3 days since incorrect results were reported.

8.1.6 Legal issues

In any technology, many legal issues need to be addressed. Similarly, LLM may also have potential legal implications since internet data are scraped. An OpenAI copyright infringement case is already pending against OpenAI, and the latest is a copyright infringement suit against OpenAI and Microsoft of the New York Times. Because of this absence of proper regulations, entry into the ChatGPT has the potential to become part of the mass and might leak the intellectual property of an organization. Samsung also did not allow its employees to use ChatGPT due to a leak in the code, Amazon, JP and so on. Companies such as Morgan Chase, Verizon, and Walmart have also acted measures to either reduce or prohibit the use of ChatGPT (Gartner 2023; Stokel-Walker and Nordeen 2023; Brittain 2023; Fisher 2023; Ray 2023; Sabin 2023).

8.1.7 Security

There are serious cybersecurity risks associated with the use of GenAI tools. Most IT leaders have identified possible threats, including data exploitation and model manipulation. Such tools can be adopted in defense or uniquely in offense; thus, they are prone to abuse. Moreover, a major part of the workforce is also at risk of unauthorized access to or theft of data. Methods such as AI jailbreaking can enable malicious people to recover confidential data within models and systems (Baxter 2023; Gupta et al. 2023; Jackson 2023; Walkowiak and MacDonald 2023; Borger 2023).

8.1.8 Intrinsic and extrinsic hallucinations

The critical challenges of GenAI include underlying issues such as intrinsic and extrinsic hallucinations. Intrinsic hallucinations involve outputs that conflict with the input or earlier information, but extrinsic hallucinations are the creation of unverifiable information that does not form its foundation in the source (Ji et al. 2023).

8.1.9 Exposure bias

One of the most critical challenges of GenAI includes exposure bias, where the discrepancy between training (where the next word is expected according to the correct previous words) and inference (where the previous outputs are given by the model) may lead to more errors and therefore more hallucinations (Ji et al. 2023).

8.1.10 Decoder and dataset issues

Decoder- and dataset-related issues are also critical limitations of GenAI, which include inaccurate data in the decoder context, unnecessary or conflicting training information, and the generation of factually imprecise sequences (Ji et al. 2023).

8.2 Ways to address hallucinations

- Retrieval-Augmented Generation (RAG): Retrieval techniques are integrated into the model, allowing it to fetch relevant source documents throughout the generation process, anchor the output in real data and reduce hallucinations compared with parametric-only (fully neural) models (Ji et al. 2023; Wu 2025).
- Self-consistency: Outputting multiple times using the same input and choosing the most stable (majority) output has been shown to help decrease the error rate in large language models (Wu 2025).
- Data-centric approaches: Precise, trustworthy training information: Reduce variation in noise or source references, which are common in summarization and data-to-text tasks.
- Synthetic counter-hallucination data: “Negative” examples (artificial spans) are added during training, and classifiers are trained to treat them as errors and fix them (e.g., Fact-Checking and Claim Classification (FactCC), Factuality and Reliability Assessment for Natural Language Knowledge (FRANK) benchmarks).
- Domain-specific fine-tuning: This method takes advantage of authenticated, verified corpora (medical data, legal data, etc.) to curb factual drift (Ji et al. 2023).

The combination of these approaches emphasizes a shift away from simple text generation to verifying and substantiating systems where the outputs are checked against one another and reinforced by one another, substantially reducing hallucinations in autoregressive Gen AI.

8.3 Future directions

On the basis of the findings from this systematic review, several future research directions can be suggested to advance the field of GenAI in various settings or disciplines.

- In the future, the transparency of GenAI models needs to be improved. Knowledge in the process of generating the outputs of AI to display a clear source of information to the users can enhance user trust and acceptance of the GenAI tools in different fields or environments (Bahroun et al. 2023).
- Combating Biases and Fairness: Since GenAI models work on the basis of the available information, any existing biases in the available information can be transferred to the output of the model. Future studies addressing bias reduction in GenAI tools in different fields, including healthcare, finance, law, and social media, should aim at reducing biases in those tools to prevent the reinforcement of stereotypes and a perceived bias toward some communities and groups (Bahroun et al. 2023).
- Notably, further studies should focus on investigating how governmental financing and the process of public–private partnership might reduce the risks associated with funding and instead serve to overcome ethical concerns. Furthermore, how ethical decision-making is influenced by the influence of investors in GAI startups should be explored to provide a better view of the situation in the industry. In addition, we suggest investigating the effects of various types of investors, including venture capitalists, angel investors, and corporate investors, to capture a modified perspective of investor relations and their impact on meeting GAI startups' funding needs and future prosperity (Siddik et al. 2024).
- Future work must examine new AI data manipulation strategies, evaluate more generative models than GPT-4, and venture into additional data modalities that include categorical, spatial, audio, and video. The variety and realism of the datasets used and the ability to test new models regularly will be necessary to adhere to the dynamically changing phenomenon of AI and obtain a clear image of the current manipulation potential (Kim 2024).
- GenAI is promising and holds more in the future. Autonomous agents could change our world in a meaningful way, and we have just started to experiment with that possibility. The above are only a few examples of how autonomous agents can be used to create content in the future. With the evolution of autonomous agents, there is a probability of even greater and more novel applications of this technology (Dodig-Crnkovic and Burgin 2024).

The findings indicate that there is a general shortage of critical analysis related to how and why computing resources are demanded, the problem of hallucination, and even legal uncertainties. Critical reflection reveals that the proposed solutions are mainly in the subcategory and are aimed at treating temporary issues rather than the actual cause of the issue. The techniques used to mitigate the problem of hallucination frequently reduce these issues but fail to ensure scaling and reliability. Critical analysis aggregates information on technological limitations, ethical conflicts and practical barriers and focuses on some of the more recent challenges in the area of GenAI. This finding indicates that the future of GenAI will depend less on the novelty aspect and rather on the ability to ensure proper stability, trust, and responsibility in their ecosystem.

9 Conclusion

This systematic review of GenAI reveals its transformative potential across multiple industry sectors while highlighting ethical considerations. The emerging trend of large concept models with the incorporation of knowledge and data-driven approaches has significantly expanded research and application development. GenAI stands at the forefront of technological innovation, reshaping industries and redefining boundaries. While GenAI offers novel opportunities for start-ups, its deployment demands careful consideration of ethical frameworks, scalability, and compliance. The emergence of agentic AI and multiagent systems guarantees that some of GenAI's current limitations and concrete ways for more robust, intelligent, and collaborative solutions are created. However, the incorporation of GenAI with Agentic AI remains restricted by limitations such as coordination, validation and resource availability. However, responsible deployment demands continued interdisciplinary research, ethical governance frameworks, scalability, and industry-wide collaboration. Organizations must adopt strategic implementation frameworks that balance innovation opportunities with risk mitigation, such as hallucination, security, privacy, and ethical ambiguity. Future research should focus on developing robust ethical guidelines, enhancing security protocols, and exploring multiagent system architectures to fully realize the transformative potential of generative AI while ensuring responsible deployment.

Acknowledgements This research was funded by the Deanship of Scientific Research, King Khalid University, Kingdom of Saudi Arabia, and the large grant number is RGP.2/624/46.

Author contributions Conceptualization, K.P., M.S., K.M.Q., and M.R.N.Q.; methodology, AK.P., M.S., K.M.Q., and M.R.N.Q.; formal analysis, K.P., M.S., K.M.Q., and M.R.N.Q.; writing—original draft preparation, K.P., M.S., K.M.Q., and M.R.N.Q.; writing—review and editing, K.P., M.S., K.M.Q., and M.R.N.Q.; funding acquisition, M.R.N.Q.; All authors have read and agreed to the published version of the manuscript.

Funding This research was funded by the Deanship of Scientific Research, King Khalid University, Kingdom of Saudi Arabia, and the large grant number is RGP.2/624/46.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Abacha AB et al (2025) MEDEC: A benchmark for medical error detection and correction in clinical notes, in Findings of the Association for Computational Linguistics: ACL 2022, pp. 22539–22550, Jan. https://doi.org/10.18653/v1/2025.findings-acl.1159

Acharya DB, Kuppan K, Divya B, Agentic AI (2025) Autonomous intelligence for complex goals—a comprehensive survey. IEEE Access 13:18912–18936. https://doi.org/10.1109/ACCESS.2025.3532853

Aggarwal A, Mittal M, Battineni G (2021) Generative adversarial network: an overview of theory and applications. Int J Inf Manag Data Insights. https://doi.org/10.1016/j.jjimei.2020.100004

Ahmad H, Goel D (2025) The future of AI: Exploring the potential of large concept models, arXiv preprint arXiv:2501.05487

Ahmed T, Ghosh S, Bansal C, Zimmermann T, Zhang X, Rajmohan S (2023) Recommending root-cause and mitigation steps for cloud incidents using large language models, in Proceedings - International Conference on Software Engineering, pp. 1737–1749. https://doi.org/10.1109/ICSE48619.2023.00149

Al-Sarayreh M, Gomes Reis M, Carr A, Reis MM (2023) Inverse design and AI/Deep generative networks in food design: a comprehensive review, Trends Food Sci Technol 138:215–228. https://doi.org/10.1016/j.tifs.2023.06.005

AlDahoul N, Hong J, Varvello M, Zaki Y (2025) Towards a world wide web powered by generative AI. Sci Rep 15(1):1–14. https://doi.org/10.1038/s41598-024-77301-0

Aldausari N, Sowmya A, Marcus N, Mohammadi G (2023) Video generative adversarial networks: a review. ACM Comput Surv 25. https://doi.org/10.1145/3487891

Alkhalaf M, Yu P, Yin M, Deng C (2024) Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. J Biomed Inform 156:104662. https://doi.org/10.1016/j.jbi.2024.104662

Alzoubi YI, Mishra A (2025) Differential privacy and artificial intelligence: potentials, challenges, and future avenues. EURASIP J Inf Secur. https://doi.org/10.1186/s13635-025-00203-9

Arachchige PCM, Bertok P, Khalil I, Liu D, Camtepe S, Atiquzzaman M (2019) Local differential privacy for deep learning. IEEE Internet Things J 7:5827–5842

Bahroun Z, Anane C, Ahmed V, Zacca A (2023) Transforming education: a comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. Sustainability. https://doi.org/10.3390/su151712983

Balasubramaniam S et al (2024) The road ahead: emerging trends, unresolved issues, and concluding remarks in generative AI—a comprehensive review. Int J Intell Syst. https://doi.org/10.1155/2024/4013195

Bamaniah A (2025) Meta's large concept models (LCMs) are here to challenge and redefine LLMs, Level Up / GitConnected. [Online]. Available: https://levelup.gitconnected.com/metas-large-concept-models-lcms-are-here-to-challenge-and-redefine-llms-7f9778f88a87. [Accessed: Aug. 22, 2025]

Bandi A, Adapa PVSR, Kuchi YEVPK (2023) The power of generative AI: a review of requirements, models, input–output formats, evaluation metrics, and challenges. Future Internet. https://doi.org/10.3390/fi15080260

Barrault L, Duquenne P-A, Elbayad M, Kozhevnikov A, Alastruey B, Andrews P, Coria M, Couairon G, Costa-jussà MR et al (2024) Large concept models: language modeling in a sentence representation space, arXiv preprint arXiv:2412.08821

Batista J, Mesquita A, Carnaz G (2024) Generative AI and higher education: trends, challenges, and future directions from a systematic literature review. Information. https://doi.org/10.3390/info15110676

Baxter K, Schlesinger Y (2023) Managing the risks of generative AI: Guidelines for companies as they implement the next generation of AI tools, Harvard Business Review Digital Articles. Available: https://hbr.org/2023/06/managing-the-risks-of-generative-ai

Bhuyan SS et al (2025) Generative artificial intelligence use in healthcare: opportunities for clinical excellence and administrative efficiency. J Med Syst 49(1):10. https://doi.org/10.1007/s10916-024-02136-1

Birlasoft (2023) A to Z of Generative AI, White Paper, [Online]. Available: https://www.birlasoft.com/resources

Bommasani R (2021) Center for research on foundation models (CRFM) On the opportunities and risks of foundation models, Stanford Univ., Stanford, CA, USA, Tech. Rep. [Online]. Available: https://arxiv.org/abs/2108.07258

Borger JG et al (2023) Artificial intelligence takes center stage: exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. Immunol Cell Biol 101(10):923–935. https://doi.org/10.1111/imcb.12689

Boxleitner A (2023) Pushing boundaries or crossing lines? The complex ethics of ChatGPT Jailbreaking, SSRN Electron. J. pp. 1–11. https://doi.org/10.2139/ssrn.4604736

Brittain B (2023) Lawsuit says OpenAI violated US authors' copyrights to train AI chatbot, Reuters. Available: https://www.reuters.com/legal/lawsuit-says-openai-violated-us-authors-copyrights-train-ai-chatbot-2023-06-29/

Caramancion KM (2020) An exploration of disinformation as a cybersecurity threat, Proc. – 3rd Int Conf Inf Comput Technol ICICT 2020, pp. 440–444. https://doi.org/10.1109/ICICT50521.2020.00076

Cardoso RC, Ferrando A (2021) A review of agent-based programming for multi-agent systems. Computers 10(2):16. https://doi.org/10.3390/computers10020016

Carlini N et al (2021) Extracting training data from large language models, Proc. 30th USENIX Secur Symp, pp. 2633–2650

Carlini N et al (2024) Poisoning web-scale training datasets is practical,., IEEE symposium on security and privacy (SP), San Francisco, CA, USA, 2024, pp. 407–425. https://doi.org/10.1109/SP54263.2024.00179

Carugati C (2023) Competition in generative artificial intelligence foundation models. EconStor. Available: http://hdl.handle.net/10419/274220

Chen RJ, Lu MY, Chen TY et al (2021) Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng 5:493–497. https://doi.org/10.1038/s41551-021-00751-8

Chen W et al (2023) AgentVerse: Facilitating multi-agent collaboration and exploring emergent behaviors, arXiv preprint arXiv:2308.10848. Available: http://arxiv.org/abs/2308.10848

Chen X (2024) A survey on large language model based autonomous agents. CCL 2024–23rd Chinese Natl Conf Comput Linguist 2(6):141–150

CLAN (2024) Differential privacy in AI. [Online]. Available: https://clanx.ai/glossary/differential-privacy-in-ai. [Accessed: Jul. 2, 2024]

Colelough BC, Regli W (2024) Neuro-Symbolic AI in 2024: A systematic review, in Proc. 1st Int. workshop on logical foundations of neuro-symbolic AI (LNSAI 2024), Jeju, South Korea, CEUR Workshop Proc., pp. 1–12. Available: https://ceur-ws.org

Cordella A, Gualdi F (2024) Regulating generative AI: the limits of technology-neutral regulatory frameworks. Insights from italy's intervention on ChatGPT. Gov Inf Q 41(4):101982. https://doi.org/10.1016/j.giq.2024.101982

Corvello V (2025) Generative AI and the future of innovation management: a human centered perspective and an agenda for future research. J Open Innov Technol Mark Complex 11(1):100456. https://doi.org/10.1016/j.joitmc.2024.100456

Crisp J (2025) Introduction to Google's Secure AI Framework (SAIF), Google AI Red Team Digital Final Report, Google. Available: https://services.google.com/fh/files/blogs/google_ai_red_team_digital_final.pdf. [Accessed: Aug. 23, 2025]

Devaux E (2024) What is differential privacy: Definition, mechanisms, and examples. Available: https://www.anonos.com/blog/what-is-differential-privacy-definition-mechanisms-examples. [Accessed: Jul. 4, 2024]

Dodig-Crnkovic G, Burgin M (2024) A systematic approach to autonomous agents. Philosophies 9(2):1–10. https://doi.org/10.3390/philosophies9020044

Dong W, Shen S, Han Y, Tan T, Wu J, Xu H (2025) Generative models in medical visual question answering: a survey. Appl Sci 15(6):2983. https://doi.org/10.3390/app15062983

Dwivedi YK et al (2023) So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Edwards B (2023) AI-powered Bing Chat spills its secrets via prompt injection attack, Ars Technica, Feb. 10, 2023. [Online]. Available: https://arstechnica.com/information-technology/2023/02/ai-powered-bing-chat-spills-its-secrets-via-prompt-injection-attack/. [Accessed: Jun. 20, 2025]

Federal Trade Commission [FTC] (2023) Generative AI Raises Competition Concerns, Office of Technology Blog, Jun. 29, [Online]. Available: https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

Fenta AA (2025) A review on enhancing education with AI: exploring the potential of ChatGPT, Bard, and generative AI. Discov Educ. https://doi.org/10.1007/s44217-025-00426-5

Feretzakis G, Papaspyridis K, Gkoulalas-Divanis A, Verykios VS (2024) Privacy-preserving techniques in generative AI and large language models: a narrative review. Information. https://doi.org/10.3390/info15110697

Fisher S (2023) NYT sues OpenAI, Microsoft for copyright infringement, Axios. Available: https://www.axios.com/2023/12/27/nyt-microsoft-openai-lawsuit-copyright-infringement

Floridi L (2025) AI as agency without intelligence: on artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. Philos Technol 38:1–27. https://doi.org/10.1007/s13347-025-00858-9

Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L (2023) Generative AI and chatgpt: applications, challenges, and AI-human collaboration. J Inf Technol Case Appl Res 25(3):277–304. https://doi.org/10.1080/15228053.2023.2233814

Gartner (2023) Gartner experts answer the top generative AI questions for your enterprise, Gartner. Available: https://www.gartner.com/en/topics/generative-ai

Gelper S, van der Lans R, van Bruggen G (2021) Competition for attention in online social networks: implications for seeding strategies. Manage Sci 67(2):1026–1047. https://doi.org/10.1287/mnsc.2019.3564

Golda A et al (2024) Privacy and security concerns in generative AI: a comprehensive survey. IEEE Access 12:48126–48144. https://doi.org/10.1109/ACCESS.2024.3381611

Goldstein JA, Sastry G, Musser M, DiResta R, Gentzel M, Sedova K (2023) Generative language models and automated influence operations: Emerging threats and potential mitigations, arXiv preprint arXiv:2301.04246. Available: https://arxiv.org/abs/2301.04246

Greshake K, Abdelnabi S, Mishra S, Endres C, Holz T (2023) Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection, arXiv preprint arXiv:2302.12173. Available: https://doi.org/10.48550/arXiv.2302.12173

Gupta M, Akiri C, Aryal K, Parker E, Praharaj L (2023) From ChatGPT to threatgpt: impact of generative AI in cybersecurity and privacy. IEEE Access 11:80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381

Hirsch-Kreinsen H (2024) Artificial intelligence: a 'promising technology.' AI Soc 39(4):1641–1652. https://doi.org/10.1007/s00146-023-01629-w

Hosseini M, Rasmussen LM, Resnik DB (2024) Using AI to write scholarly publications. Account Res 31(7):715–723. https://doi.org/10.1080/08989621.2023.2168535

Huang D, Yan C, Li Q, Peng X (2024) From large language models to large multimodal models: a literature review. Appl Sci 14(12):5068. https://doi.org/10.3390/app14125068

Huang F, Ren Y (2024) Harnessing the green frontier: the impact of green finance reform and digitalization on corporate green innovation. Finance Res Lett 66:105554. https://doi.org/10.1016/j.frl.2024.105554

Huang Z, Hu R, Guo Y, Chan-Tin E, Gong Y (2019) DP-ADMM: ADMM-based distributed learning with differential privacy. IEEE Trans Inf Forensics Secur 15:1002–1012

Humphreys D, Koay A, Desmond D, Mealy E (2024) AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business. AI Ethics 4(3):791–804. https://doi.org/10.1007/s43681-024-00443-4

Ibrar W, Mahmood D, Al-Shamayleh AS, Ahmed G, Alharthi SZ, Akhunzada A (2025) Generative AI: a double-edged sword in the cyber threat landscape. Artif Intell Rev. https://doi.org/10.1007/s10462-025-11285-9

Jackson T (2023) Exploring the security risks of generative AI, Forbes. Available: https://www.forbes.com/sites/forbestechcouncil/2023/04/19/exploring-the-security-risks-of-generative-ai/?sh=71db85513594

Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of hallucination in natural language generation. ACM Comput Surv 55(12):248. https://doi.org/10.1145/3571730

Jin P et al (2023) Assess and summarize: improve outage understanding with large language models, in ESEC/FSE 2023 - Proceedings of the 31st ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 1657–1668. https://doi.org/10.1145/3611643.3613891

Kaddour J et al (2023) Challenges and Applications of Large Language Models, arXiv preprint arXiv:2307.10169. Available: https://arxiv.org/abs/2307.10169

Kalota F (2024) A primer on generative artificial intelligence. Educ Sci 14(2):172. https://doi.org/10.3390/educsci14020172

Kan K (2023) Seeking the ideal privacy protection: strengths and limitations of differential privacy. Monet Econ Stud 41:49–80

Kandwal S, Nehra V (2024) A survey of text-to-image diffusion models in generative AI, Proc. 14th Int Conf Cloud Comput Data Sci Eng Conflu 2024, pp. 73–78. https://doi.org/10.1109/Confluence60223.2024.10463372

Kaynak O (2021) The golden age of artificial intelligence: inaugural editorial. Discov Artif Intell. https://doi.org/10.1007/s44163-021-00009-x

Kim J et al (2023) AI anxiety: a comprehensive analysis of psychological factors and interventions. SSRN Electron J. https://doi.org/10.2139/ssrn.4573394

Kim J, Lee SS, Detrick R, Wang J, Li N (2025) Students-generative AI interaction patterns and its impact on academic writing. J Comput High Educ. https://doi.org/10.1007/s12528-025-09444-6

Kim JJH et al (2024) Generative AI can effectively manipulate data. AI Ethics. https://doi.org/10.1007/s43681-024-00546-y

Kousounadis-Knousen MA, Bazionis IK, Georgilaki AP, Catthoor F, Georgilakis PS (2023) A review of solar power scenario generation methods with focus on weather classifications, temporal horizons, and deep generative models. Energies. https://doi.org/10.3390/en16155600

Krakowski S (2025) Human-AI agency in the age of generative AI. Inf Organ 35(1):100560. https://doi.org/10.1016/j.infoandorg.2025.100560

Lee C, Vogt KA, Kumar S (2024) Prospects for AI clinical summarization to reduce the burden of patient chart review. Front Digit Heal 6:1–9. https://doi.org/10.3389/fdgth.2024.1475092

Li C et al (2024) Multimodal foundation models: from specialists to general-purpose assistants. Found Trends® Comput Graph Vis. https://doi.org/10.1561/0600000110

Li X, Wang S, Zeng S, Wu Y, Yang Y (2024) A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges. Vicinagearth. https://doi.org/10.1007/s44336-024-00009-2

Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z (2023) ChatGPT and a new academic reality: artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. J Assoc Inf Sci Technol 74(5):570–581. https://doi.org/10.1002/asi.24750

Luo J, Zhuo W, Xu B (2024) A deep neural Network-Based assistive decision method for financial risk prediction in carbon trading market. J Circuits Syst Computers 33(08):2450153. https://doi.org/10.1142/S0218126624501536

Lv Z (2023) Generative artificial intelligence in the metaverse era. Cogn Robot 3:208–217. https://doi.org/10.1016/j.cogr.2023.06.001

Majumder S, Dey N (2022) Artificial Intelligence and Knowledge Management. Studies Big Data 107:85–100. https://doi.org/10.1007/978-981-19-0316-8_5

Mao R, Chen G, Zhang X, Guerin F, Cambria E (2024) GPTEval: A survey on assessments of ChatGPT and GPT-4, in (2024) joint international conference on computational linguistics, language resources and evaluation, LREC-COLING 2024 - Main Conference Proceedings, pp. 7844–7866

Martela F (2025) Artificial intelligence and free will: generative agents utilizing large language models have functional free will. AI Ethics 5(4):4389–4400. https://doi.org/10.1007/s43681-025-00740-6

Masterman T, Besen S, Sawtell M, Chao A (2024) The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: a survey, arXiv preprint arXiv:2404.11584v1, Available: https://arxiv.org/abs/2404.11584v1

Matharaarachchi A et al (2024) Optimizing generative AI chatbots for net-zero emissions energy internet-of-things infrastructure. Energies. https://doi.org/10.3390/en17081935

Microsoft (2025) A list of metrics for evaluating LLM-generated content, Microsoft AI. [Online]. Available: https://learn.microsoft.com/en-us/ai/playbook/technology-guidance/generative-ai/working-with-llms/evaluation/list-of-eval-metrics. [Accessed: Aug. 19, 2025]

Mohawesh R, Ottom MA, Salameh HB (2025) A data-driven risk assessment of cybersecurity challenges posed by generative AI. Decis Anal 15:100580. https://doi.org/10.1016/j.dajour.2025.100580

Morjaria L, Gandhi B, Haider N, Mellon M, Sibbald M (2025) Applications of generative artificial intelligence in electronic medical records: a scoping review. Information 16(4):1–10. https://doi.org/10.3390/info16040284

Mousavi R et al (2025) Revolutionizing solar energy resources: the central role of generative AI in elevating system sustainability and efficiency. Appl Energy 382:125296. https://doi.org/10.1016/j.apenergy.2025.125296

Nabi J (2025) All you need to know about LLM text generations, Medium. [Online]. Available: https://medium.com/@javaid.nabi/all-you-need-to-know-about-llm-text-generation-03b138e0ed19. [Accessed: Aug. 22, 2025]

Naitali A, Ridouani M, Salahdine F, Kaabouch N (2023) Deepfake attacks: generation, detection, datasets, challenges, and research directions. Computers 12(10):1–26. https://doi.org/10.3390/computers12100216

Nakajima Y (2025) BabyAGI. GitHub repository. Available: https://github.com/yoheinakajima/babyagi. [Accessed: Aug. 22, 2025]

Nakajima Y (2025) Birth of BabyAGI. Available: https://yoheinakajima.com/birth-19of-babyagi/. [Accessed: Aug. 22, 2025]

Near JP, Darais D, Lefkovitz N, Howarth G (2025) Guidelines for evaluating differential privacy guarantees, NIST, Special Publication 800–226, 2025. [Online]. Available: https://csrc.nist.gov/pubs/sp/800/226/final. [Accessed: Jul. 2, 2025]

Newton C, Singleton J, Copland C, Kitchen S, Hudack J (2021) Scalability in modeling and simulation systems for multi-agent, AI, and machine learning applications. Artificial intelligence and machine learning for multi-domain operations applications III, vol 11746. SPIE, pp 534–552

Nezami Z, Pournaras E, Borzouie A (2024) Generative AI on the edge: Architecture and performance evaluation, arXiv preprint arXiv:2411.17712, Mar. [Online]. Available: https://arxiv.org/abs/2411.17712

Ning Y, Teixayavong S, Shang Y, Savulescu J, Nagaraj V, Miao D, Mertens M, Ting DSW, Ong JCL, Liu M, Cao J, Dunn M, Vaughan R, Ong MEH, Sung JJ-Y, Topol EJ, Liu N (2024) Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. Lancet Digit Health 6(11):e848–e856. https://doi.org/10.1016/S2589-7500(24)00143-2

Nori H, King N, McKinney SM, Carignan D, Horvitz E (2023) Capabilities of GPT-4 on medical challenge problems, pp. 1–35. Available: http://arxiv.org/abs/2303.13375

Olujimi PA, Owolawi PA, Mogase RC, Wyk EV (2025) Agentic AI frameworks in SMMEs: a systematic literature review of ecosystemic interconnected agents. AI. https://doi.org/10.3390/ai6060123

OpenAI (2025), Introducing GPTs, OpenAI Blog, Nov. 6, 2023. [Online]. Available: https://openai.com/blog/introducing-gpts. [Accessed: Aug. 23, 2025].

Ozkaya I (2023) The next frontier in software development: AI-augmented software development processes. IEEE Softw 40(4):4–9. https://doi.org/10.1109/MS.2023.3278056

Packt (2025) BAGI Empowering task automation through advanced AI technologies. [Online]. Available: https://www.packtpub.com/en-us/learning/how-to-tutorials/babyagi-empowering-task-automation-through-advanced-ai-technologies. [Accessed: Aug. 22, 2025]

Pan K, Ong Y-S, Gong M, Li H, Qin AK, Gao Y (2024) Differential privacy in deep learning: a literature survey. Neurocomputing 589:127663

Parcha (2025) Deep dive part 2: How does BabyAGI?. Available: https://resources.parcha.com/deep-dive-part-2-how-does-babyagi/. [Accessed: Aug. 22, 2025]

Petratos PN (2021) Misinformation, disinformation, and fake news: cyber risks to business. Bus Horiz 64(6):763–774. https://doi.org/10.1016/j.bushor.2021.07.012

Petratos PN, Faccia A (2023) Fake news, misinformation, disinformation and supply chain risks and disruptions: risk management and resilience using blockchain. Ann Oper Res 327(2):735–762. https://doi.org/10.1007/s10479-023-05242-4

Piccialli F, Chiaro D, Sarwar S, Cerciello D, Qi P, Mele V (2025) Agentai: a comprehensive survey on autonomous agents in distributed AI for industry 4.0. Expert Syst Appl 291:128404. https://doi.org/10.1016/j.eswa.2025.128404

Ponomareva N, Hazimeh H, Kurakin A, Xu Z, Denison C, McMahan HB, Vassilvitskii S, Chien S, Thakurta AG (2023) How to dp-fy ML: a practical guide to machine learning with differential privacy. J Artif Intell Res 77:1113–1201

Prakash V (2025) Large concept models (LCM), LinkedIn. [Online]. Available: https://www.linkedin.com/pulse/large-concept-models-lcm-vimal-prakash-pjgac. [Accessed: Aug. 22, 2025]

Pruseth D (2025) LCMs (large concept models): The path to AGI, the future of AI thinking, debabratapruseth.com. [Online]. Available: https://debabratapruseth.com/lcms-large-concept-models-the-path-to-agi-the-future-of-ai-thinking/. [Accessed: Aug. 22, 2025]

Puvvadi M, Arava SK, Santoria A, Chennupati SSP, Puvvadi HV (2025) Coding agents: A comprehensive survey of automated bug fixing systems and benchmarks, in Proc. 2025 IEEE 14th Int Conf Commun Syst Netw Technol (CSNT), pp. 680–686, IEEE

Qin Y et al (2025) Tool learning with foundation models. ACM Comput Surv 57:1–40. https://doi.org/10.1145/3704435

Qin Y, Xie H, Ding S et al (2023) Bar transformer: a hierarchical model for learning long-term structure and generating impressive pop music. Appl Intell 53:10130–10148. https://doi.org/10.1007/s10489-022-04049-3

Radanliev P (2025) Frontier AI regulation: what form should it take? Front Polit Sci 7:1–17. https://doi.org/10.3389/fpos.2025.1561776

Radanliev P, Santos O, Ani UD (2025) Generative AI cybersecurity and resilience. Front Artif Intell 8:1–18. https://doi.org/10.3389/frai.2025.1568360

Rasheed Z et al (2025) Autonomous agents in software development: A vision paper. 524. https://doi.org/10.1007/978-3-031-72781-8_2

Ray S (2023) Samsung bans ChatGPT among employees after sensitive code leak, Forbes. Available: https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/?sh=7108facf6078

Razzaq A (2024) Meta AI proposes large concept models (LCMs): A semantic leap beyond token-based language modeling, MarkTechPost, Dec. 15, [Online]. Available: https://www.marktechpost.com/2024/12/15/meta-ai-proposes-large-concept-models-lcms-a-semantic-leap-beyond-token-based-language-modeling. [Accessed: Aug. 22, 2025]

Reddi VJ (2025) Generative AI at the edge: challenges and opportunities: the next phase in AI deployment. ACM Queue 23:40. https://doi.org/10.1145/3733702

Rezazadeh A, Kohns M, Bohnsack R, António N, Rita P (2025) Generative AI for growth hacking: how startups use generative AI in their growth strategies. J Bus Res. https://doi.org/10.1016/j.jbusres.2025.115320

Sabin S (2023) Companies are struggling to keep corporate secrets out of ChatGPT, Axios. Available: https://www.axios.com/2023/03/10/chatgpt-ai-cybersecurity-secrets

Sai S, Arunakar K, Chamola V, Hussain A, Bisht P, Kumar S (2025) Generative AI for finance: applications, case studies and challenges. Expert Syst. https://doi.org/10.1111/exsy.70018

Sai S, Gaur A, Sai R, Chamola V, Guizani M, Rodrigues JJPC (2024) Generative AI for transformative healthcare: a comprehensive study of emerging models, applications, case studies, and limitations,. IEEE Access 12:31078–31106. https://doi.org/10.1109/ACCESS.2024.3367715

Samuelson P (2023) Legal challenges to generative AI, part I. Commun ACM 66(7):20–23. https://doi.org/10.1145/3597151

Sánchez-Mompó A, Mavromatis I, Li P, Katsaros K, Khan A (2025) Green MLOps to green genops: an empirical study of energy consumption in discriminative and generative AI operations †. Information 16(4):1–24. https://doi.org/10.3390/info16040281

Sapkota R, Roumeliotis KI, Karkee M (2025) AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges, arXiv preprint arXiv:2505.10468. Available: https://arxiv.org/html/2505.10468v1

Schneider J (2024) Explainable generative ai (GenXAI): a survey, conceptualization, and research agenda. Artif Intell Rev 57:289. https://doi.org/10.1007/s10462-024-10916-x

Schneider J (2025) Generative to Agentic AI: survey, conceptualization, and challenges, arXiv preprint, arXiv:2504.18875. Available: https://arxiv.org/abs/2504.18875

Sengar SS, Bin Hasan A, Kumar S, Carroll F (2024) Generative artificial intelligence: a systematic review and applications. Multimed Tools Appl. https://doi.org/10.1007/s11042-024-20016-1

Shao Z, Zhao R, Yuan S, Ding M, Wang Y (2022) Tracing the evolution of AI in the past decade and forecasting the emerging trends. Expert Syst Appl 209:118221. https://doi.org/10.1016/j.eswa.2022.118221

Shavit Y et al (2023) Practices for governing agentic AI systems, OpenAI, pp. 1–23. Available: https://openai.com/index/practices-for-governing-agentic-ai-systems/

Shrikhande A (2025) A deep dive into large concept models (LCMs), ADASCI. [Online]. Available: https://adasci.org/a-deep-dive-into-large-concept-models-lcms. [Accessed: Aug. 22, 2025]

Siddik AB, Li Y, Du AM (2024) Unlocking funding success for generative AI startups: The crucial role of investor influence. Financ. Res. Lett. 69:106203. https://doi.org/10.1016/j.frl.2024.106203

Stimers P, Serafino MC, Roberson JE, Wooten T, Barsky DJ, German C (2023) What to know about the new artificial intelligence executive order, Holland & Knight. Available: https://www.hklaw.com/en/insights/publications/2023/10/what-to-know-about-the-new-artificial-intelligence-executive-order

Stokel-Walker C, Nordeen VR (2023) The promise and peril of generative AI. Nature 614:214–216. https://doi.org/10.1038/d41586-023-00340-6

Storey VC, Yue WT, Zhao JL, Lukyanenko R (2025) Generative artificial intelligence: evolving technology, growing societal Impact, and opportunities for information systems research. Inf Syst Front. https://doi.org/10.1007/s10796-025-10581-7

Taherdoost H, Madanchian M (2024) AI advancements: comparison of innovative techniques. AI 5(1):38–54. https://doi.org/10.3390/ai5010003

Telikani A, Tahmassebi A, Banzhaf W, Gandomi AH (2022) Evolutionary machine learning: a survey. ACM Comput Surv. https://doi.org/10.1145/3467477

Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J (2020) Deepfakes and beyond: a survey of face manipulation and fake detection. Inf Fusion 64:131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Uddin M et al (2025) *Generative AI* revolution in cybersecurity: a comprehensive review of threat intelligence and operations. Artif Intell Rev. https://doi.org/10.1007/s10462-025-11219-5

Uppalapati V, Nag D (2024) A comparative analysis of AI models in complex medical decision-making scenarios: evaluating ChatGPT, Claude AI, Bard, and Perplexity. Cureus 16:e52485. https://doi.org/10.7759/cureus.52485

Vaswani A et al (2017) Attention is all you need, in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, Dec. pp. 5998–6008. [Online]. Available: https://arxiv.org/abs/1706.03762

Walkowiak E, MacDonald T (2023) Generative AI and the workforce: what are the risks? SSRN Electron J. https://doi.org/10.2139/ssrn.4568684

Walters WH, Wilder EI (2023) Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep 13(1):1–8. https://doi.org/10.1038/s41598-023-41032-5

Wang Y, Liang L, Li R, Wang Y, Hao C (2024) Comparison of the performance of ChatGPT, Claude and Bard in support of myopia prevention and control. J Multidisciplinary Healthc 17:3917–3929. https://doi.org/10.2147/JMDH.S473680
Wason R (2018) Deep learning: evolution and expansion. Cogn Syst Res 52:701–708. https://doi.org/10.1016/j.cogsys.2018.08.023
Wei X, Wang L, Lee LK, Liu R (2025) The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: an experimental study. Int J Educ Technol High Educ. https://doi.org/10.1186/s41239-025-00526-0
Wu D, Zhang J (2025) Generative artificial intelligence in secondary education: applications and effects on students' innovation skills and digital literacy. PLoS One 20:1–22. https://doi.org/10.1371/journal.pone.0323349
Wu S et al (2023) BloombergGPT: A large language model for finance, pp. 1–76, Dec. [Online]. Available: http://arxiv.org/abs/2303.17564
Wu W et al (2025) MultiRAG: a knowledge-guided framework for mitigating hallucination in multi-source retrieval augmented generation, in., IEEE 41st International Conference on Data Engineering (ICDE), Hong Kong, 2025, pp. 3070–3083. https://doi.org/10.1109/ICDE65448.2025.00230
Zangari A, Marcuzzo M, Schiavinato M, Gasparetto A, Albarelli A (2023) Ticket automation: an insight into current research with applications to multi-level classification scenarios. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2023.119984
Zen (2025) Large concept models (LCMs), Medium. [Online]. Available: https://medium.com/@ThisIsMeIn360VR/large-concept-models-lcms-d59b86531ef6. [Accessed: Aug. 22, 2025]
Zhang B, Zhu J, Su H (2023) Toward the third generation artificial intelligence. Sci China Inf Sci 66(2):1–19. https://doi.org/10.1007/s11432-021-3449-x
Zhang C, Zhang C, Li C, Qiao Y, Zheng S, Dam SK, Zhang M, Kim JU, Kim ST, Choi J, Park G-M, Bae S-H, Lee L-H, Hui P, Kweon IS (2023) One small step for generative AI, one giant leap for AGI: A complete survey on ChatGPT in AIGC era, arXiv preprint, arXiv:2304.06488, Apr. [Online]. Available: https://arxiv.org/abs/2304.06488
Zhou C, Li Q, Li C et al (2024) A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT. Int J Mach Learn Cybern. https://doi.org/10.1007/s13042-024-02443-6
Zhu K (2025) Trust and generative AI: embodiment considered. AI Ethics 5(3):2987–2997. https://doi.org/10.1007/s43681-024-00611-6

Authors and Affiliations

Kinjal Patel[1] · Milind Shah[2] · Karishma M. Qureshi[3] · Mohamed Rafik N. Qureshi[4]

✉ Mohamed Rafik N. Qureshi
mrnoor@kku.edu.sa

Kinjal Patel
kinjal5721@gmail.com

Milind Shah
milindshahcomputer@gmail.com

Karishma M. Qureshi
kariq18@gmail.com

1 Builure Services Pty. Ltd., Melbourne, VIC, Australia

2 Department of Computer Engineering, Sardar Vallabhbhai Patel Institute of Technology (S.V.I.T), Vasad, Gujarat, India

3 Department of Mechanical Engineering, Parul Institute of Technology, Parul University, Waghodia 391760, India

[4] Department of Industrial Engineering, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia