PeerJ Computer Science

Exploring generative artificial intelligence: a comprehensive guide

Rasha Shoitan[1], Mona M. Moussa[1], Nahed Tawfik[1], Young-Im Cho[2] and Mohamed S. Abdallah[2,3,4]

[1] Computers and Systems Department, Electronics Research Institute (ERI), Cairo, Egypt
[2] Department of Computer Engineering, Gachon University, Seoul, Republic of South Korea
[3] Informatics Department, Electronics Research Institute (ERI), Cairo, Egypt
[4] AI Lab, DeltaX Co., Ltd., Seoul, Republic of South Korea

ABSTRACT

Generative artificial intelligence (GAI), a specialized branch of artificial intelligence, has developed as a dynamic discipline that drives innovation and creativity across several domains. It is concerned with creating models that can autonomously produce novel text, images, videos, music, 3D, code, and more. GAI is distinguished by its capacity to acquire knowledge from extensive datasets, identify recurring patterns, capture distributions, and produce novel content demonstrating similar features. This review presents a comprehensive historical overview of the development and progression of GAI techniques over the years. Various essential methodologies employed in developing GAI are also discussed, including Generative Adversarial Networks (GANs), Variational AutoEncoders (VAEs), transformers, and diffusion models. Moreover, a detailed overview of the technologies used in GAI is provided for generating images, videos, music, code, and text, such as ChatGPT, DALL-E, Midjourney, Claude, Bard, GitHub Copilot, and others. The research subsequently introduces the different datasets used to train the GAI models and the evaluation metrics for evaluating their performances. Ultimately, the research investigates the diverse applications of GAI across various domains, challenges, and ethical implications.

Subjects Artificial Intelligence, Computer Vision
Keywords Generative AI, Generative models, Midjourney, Transformers, Diffusion models

Submitted 2 May 2025
Accepted 16 September 2025
Published 6 January 2026

Corresponding authors
Young-Im Cho, yicho@gachon.ac.kr
Mohamed S. Abdallah, sameer@gachon.ac.kr

Academic editor
Consolato Sergi

Additional Information and Declarations can be found on page 38

DOI **10.7717/peerj-cs.3276**

INTRODUCTION

The way many activities are tackled in the current world has been transformed by artificial intelligence (AI). AI refers to the systems that can acquire knowledge from input data and utilize it to formulate decisions or predictions (*Mukhamediev et al., 2022*; *Saghiri et al., 2022*). AI has permeated every aspect of our lives across different sectors, such as healthcare, education, human resources, *etc*. Imagine you are in a football match against a computer that knows all the rules, anticipates your moves, and follows pre-established strategies instead of developing new tactics during the game. Additional instances of conventional AI include voice assistants (*Sáiz-Manzanares, Marticorena-Sánchez & Ochoa-Orihuel, 2020*; *Bălan, 2023*) such as Siri (*Apple, 2023*), Cortana (*Microsoft, 2014*), Google Assistant (*Google Assistant, 2016*) or Alexa (*Alexa, 2025*) and recommendation systems employed by social media, Spotify, Netflix, Linkedin or Amazon

How to cite this article Shoitan R, Moussa MM, Tawfik N, Cho Y, Abdallah MS. 2026. Exploring generative artificial intelligence: a comprehensive guide. **PeerJ Comput. Sci. 12:e3276 DOI 10.7717/peerj-cs.3276**

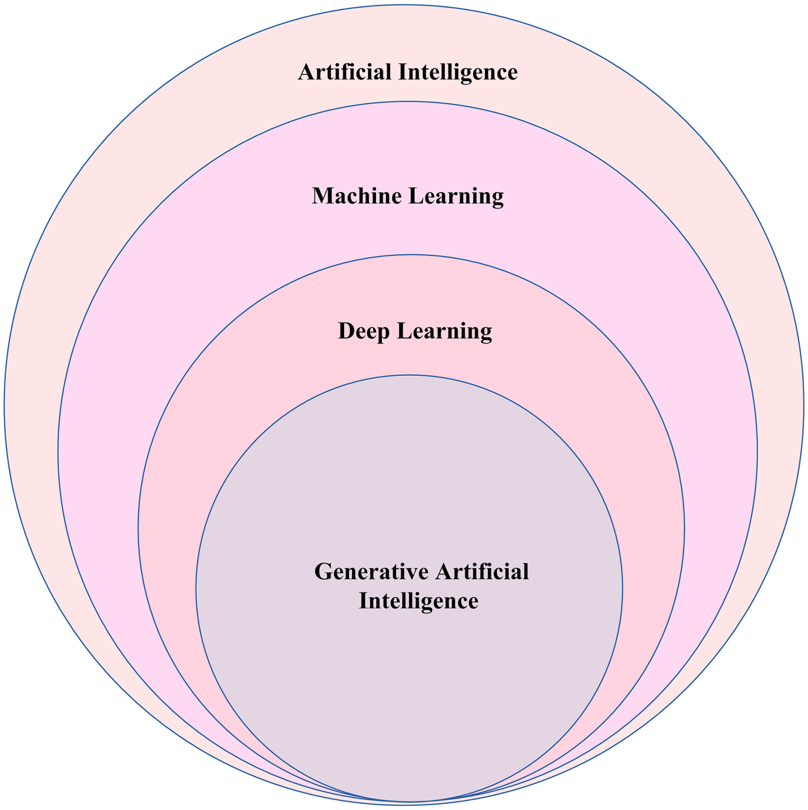


Figure 1 **Relationship between AI and generative artificial intelligence (GAI).**
Full-size DOI: 10.7717/peerj-cs.3276/fig-1

(*Zhang, Lu & Jin, 2021*). These AI systems have been trained to adhere to specific guidelines and perform tasks proficiently but do not generate novel content or ideas.

Recently, there has been a rise in the development and application of generative artificial intelligence (GAI) (*Zant, Kouw & Schomaker, 2013*; *Castelli & Manzoni, 2022*; *Bandi, Adapa & Kuchi, 2023*; *Dasgupta, Venugopal & Gupta, 2023*; *Feuerriegel et al., 2023*; *Shokrollahi et al., 2023*; *Taulli, 2023*). GAI involves training computers in unsupervised and semi-supervised ways to produce new text, images, videos, or programming code related to the trained data in response to instructions or prompts provided by the user. However, people frequently confuse "GAI" and "AI." AI is a broad domain within computer science that focuses on developing systems and machines capable of performing tasks that imitate human intelligence. AI entails training computers to identify features in data and generate predictions based on those features. Within the field of AI, there are various subsets and specialized areas, including machine learning, deep learning, GAI, Large Language Models (LLMs), and more. Figure 1 below helps to visualize the relationship between AI and GAI.

SCOPE OF THIS SURVEY

Lately, various research reviews have been introduced to deepen our understanding of GAI concepts. These reviews typically fall into two main categories. The first type focuses on a single aspect, such as models, ethical concerns, challenges, limitations or applications (*Gozalo-Brizuela & Garrido-Merchán, 2023*; *Iglesias, Talavera & Díaz-Álvarez, 2023*; *Wang et al., 2024*; *Fuest et al., 2024*). While these studies offer deep insights into their specific topics, they often lack a broad perspective, making it harder for newcomers to see the full landscape of GAI. The second type covers multiple aspects, including models, datasets, applications, and ethical issues (*Bandi, Adapa & Kuchi, 2023*; *Bengesi et al., 2024*; *Sengar et al., 2024*; *Raut & Singh, 2024*). However, even these broader surveys often leave out key details, creating gaps in the overall understanding of GAI's development and impact. This article addresses a crucial gap in GAI research by providing a comprehensive and up-to-date survey that connects multiple aspects of the field. This research work takes a broader approach, covering the evolution of GAI, its foundational models, real-world tools, datasets, evaluation metrics, and ethical implications. It presents a detailed comparison of key models like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Diffusion Models, and Transformers, highlighting their strengths, limitations, and trade-offs. Additionally, the article explores practical applications across industries, evaluates the datasets used to train these models, and discusses benchmarking techniques to assess their performance. It also emphasizes ethical challenges like bias, misinformation, and privacy concerns while suggesting future research directions. By bridging these gaps, this article serves as a valuable resource for researchers, graduate students, industry professionals, and policymakers seeking a well-rounded understanding of GAI's impact and future potential.

SURVEY METHODOLOGY

This review is conducted to explore the evolving landscape of GAI and to provide a clearer understanding of its foundations, growth, and current capabilities. As a starting point, we define key research questions to guide our investigation, including:

- What exactly is GAI, and how does it differ from traditional AI methods?
- What historical milestones and advancements have fueled the rapid growth of GAI?
- What are the core models and technologies powering today's GAI systems?
- How can we ensure that GAI is developed in a reliable, ethical, and socially responsible way?

- **Search terms and keywords**

This review draws on a wide range of sources, using carefully chosen keywords to search peer-reviewed journals such as MDPI, Springer, IEEE, and Elsevier, as well as Google Scholar and arXiv for recent preprints. In addition, we used Google searches to reach official websites of developers and companies, helping us include up-to-date information and firsthand insights directly from those driving innovations in GAI.

Comprehensive search queries are formulated using a combination of relevant keywords to ensure broad coverage of methods, models, applications, challenges, and ethical considerations in GAI. These keywords include terms such as "Generative Artificial Intelligence," "Generative AI," "GAN" OR "Generative Adversarial Networks," "VAE" OR

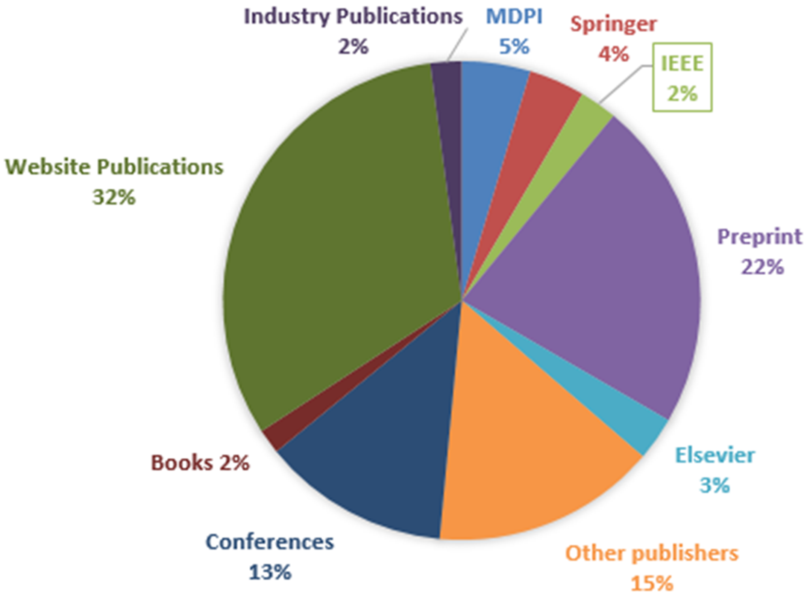


Figure 2 **Distribution of references by source type in the surveyed literature.**
Full-size DOI: 10.7717/peerj-cs.3276/fig-2

"Variational Autoencoder," "Diffusion Models," "Transformer Models," "Large Language Models" OR "LLM," "Ethical implications of Generative AI," "Applications of Generative AI," "Datasets for Generative AI," "Evaluation metrics for Generative AI," "Generative AI challenges," "GAI limitations," and "GAI ethical implications." We also adapt and refine these keywords during the search process as we identify emerging trends, adding terms like "Stable Diffusion," "foundation models," as well as tool-specific keywords such as "ChatGPT" and "DALL·E," to capture the latest developments and practical implementations within the field.

- **Search engines and databases**

The literature search is conducted iteratively from November 2023 to December 2024, rather than in a single batch, to keep up with the rapid pace of new publications in GAI. During this period, a significant number of research articles are collected, primarily published between 2020 and 2024, reflecting the field's recent surge in activity. In Fig. 2, the sources cited in this review span a broad range of publications, including websites, preprints, and conference papers, highlighting the fast-paced and open nature of GAI research. The websites listed are mainly owned by the developers or companies behind GAI, offering firsthand updates and insights into their latest innovations. On the other hand, peer-reviewed journals like MDPI, Springer, IEEE, and Elsevier contribute academic depth and reliability, delivering thorough analyses along with valuable perspectives from books and industry reports. This blend of sources ensures a comprehensive view of the field.

- **Criteria for Inclusion and exclusion**

An initial search retrieves 320 articles from selected databases and online sources. Duplicates are removed based on manual checks in Mendeley, resulting in a refined set of unique articles. Titles and abstracts are screened for relevance, with a focus on studies discussing GAI methods, models, applications, datasets, evaluation metrics, and ethical or societal implications. Articles that are off-topic or non-English are excluded during this stage. As GAI is a fast-moving field, recent and reliable sources such as preprints, company websites, peer-reviewed journals, and top conferences from 2020 to 2024 are prioritized. Through this careful process, approximately 237 articles are selected for full-text review, forming the basis of this review and ensuring a comprehensive and up-to-date perspective on the field.

GAI HISTORY

The history of GAI has deep roots, beginning in the early stages of AI. In this research article, this evolution is outlined, starting from the 1900s and progressing to its development in the present day.

1900 to 2019

Among the early instances of GAI is the "Markov Chain," a statistical technique devised by Andrey Markov in the early 1900s (*Link, 2006*). Markov chains, a relatively simple and widely applied method, serve as a statistical model for random processes. Their utility spans diverse domains, including text generation. After the 1980s, there is notable advancement in GAI, particularly with the emergence and development of neural networks (*Gad & Jarmouni, 1995*). In 1985, *Ackley, Hinton & Sejnowski (1985)* invented restricted Boltzmann machines (RBMs). RBMs, a form of neural network, can grasp intricate data distributions and generate novel data informed by those distributions. In December 1997, *Hochreiter & Schmidhuber (1997)* proposed a Long Short-Term Memory model (LSTM) to mitigate the issue of vanishing gradients in conventional Recurrent Neural Networks (RNN) (*DiPietro & Hager, 2019*). This enhancement enables more effective management of long-range dependencies within sequential data, which have shown to be very efficient in fields such as language modeling and text production.

Over the years, AI has experienced significant transformation, introducing new algorithms, methodologies, and various applications. In 2012, *Krizhevsky, Sutskever & Hinton (2012)* developed a convolutional neural network (CNN) named Alexnet that win the ImageNet large scale visual recognition challenge with 75% accuracy. It demonstrates the ability of neural networks to comprehend complex patterns in images, resulting in significant advancements in image classification and object recognition. Additionally, it marks the first utilization of Nvidia's GPU for machine learning purposes. This signals the start of a groundbreaking period in neural networks and deep learning that is facilitating the advancement of more complex and effective neural networks, particularly in the context of GAI problems. By the end of 2013, *Kingma & Welling (2014, 2019)* and *Goodfellow et al. (2014)* introduced VAEs that are engineered to comprehend the inherent probability distribution within a provided dataset and produce new samples. In 2014, *Goodfellow et al. (2014)* proposed a deep learning model called GANs, which consist of a pair of adversarial neural network models: a generator and a discriminator. These two models use a competitive process to produce novel synthetic data that yield remarkable outcomes, particularly in image generation and other areas.

Sequence-to-sequence model (Seq2Seq) is introduced by *Sutskever, Vinyals & Le (2014)* in 2014 to process an input sequence of items, such as words, letters, time series data, or similar. Then, it generates an output sequence of items in response. The model has proven highly effective in various applications such as summarizing text, image captioning, and machine translation (*Sutskever et al., 2014*). The model uses an encoder to convert the input sequence into a context vector, which the decoder then processes to generate the output, often leveraging RNNs, LSTMs (*Salem, 2017*), or Gated Recurrent Units (GRUs) (*Salem, 2022*). In addition, *Sohl-Dickstein et al. (2015)* proposed diffusion models in 2015, motivated by non-equilibrium thermodynamics. Diffusion models are indeed generative models, serving the purpose of generating data that resembles the training data. The core principle of diffusion models involves generating data by learning to reverse a process that gradually adds noise to training data. Once trained, they can turn random noise into new data that mimics the original dataset. Afterward, Alexander Mordvintsev, a Google engineer, developed a computer vision program called DeepDream. It utilizes a CNN to identify and amplify image patterns through algorithmic pareidolia. This process results in images with a dream-like appearance, resembling a psychedelic experience, due to intentional over-processing.

In September 2015, *Gatys, Ecker & Bethge (2015)*, *Mordvintsev, Olah & Tyka (2015)* introduced a system employing a deep neural network capable of generating high-quality artistic images. This system leverages neural representations to independently extract and recombine the content and style components from various images. One year later, DeepMind unveiled a raw audio generative model, “Wavenet,” a fully probabilistic autoregressive model (*van den Oord et al., 2016*). This model creates lifelike artificial voices, often indistinguishable from real speech, and is used in audio processing, music production, and speech synthesis. Google researchers (*Vaswani et al., 2017*) released a study in 2017 presenting a revolutionary neural network design known as the “"transformer.” This architecture is created in response to RNN constraints when dealing with long texts. Unlike standard RNNs, the transformer does not depend on recurrence but relies only on an attention mechanism to generate global connections between input and output. During the same year, a group comprising AI experts and entrepreneurs from UCL, Stanford, TUM, and Cambridge introduce a platform called Synthesia that specializes in utilizing AI and deep learning technology to create synthetic media content (Synthesia). Its primary focus is delivering realistic and customized movies utilizing computer-generated avatars that can lip-sync to any audio file.

Based on the transformer design, OpenAI, an AI vendor, released the first edition of its language model, Generative Pre-trained Transformer 1 (GPT-1), in June 2018 (*Radford et al., 2018*). With 117 million parameters, GPT-1 can generate coherent text and answer questions due to its training on diverse internet content. However, its smaller size and limited training data make it struggle with maintaining context in long conversations, sometimes leading to responses that sound convincing but lack logical depth. Shortly after GPT-1, Google introduced Bidirectional Encoder Representations from Transformers

(BERT) in November 2018, a model designed to improve language understanding using a bidirectional transformer architecture (*Devlin et al., 2019*). It learns from both preceding and following words in a sentence, enhancing contextual understanding and improving Google Search. However, BERT struggles with processing long texts due to its fixed input size and has difficulty capturing long-term word relationships.

Late in 2018, Nvidia researchers introduced StyleGAN, a type of GAN for producing realistic and diverse high-resolution synthetic images of human faces (*Karras, Laine & Aila, 2021*). A few months later, in early 2019, GPT-2 emerged as a more powerful successor to GPT-1, with 1.5 billion parameters, improving coherence, context retention, and creative writing (*Radford et al., 2019*). Although effective with short texts, it struggles with longer passages due to its limited context window, which restricts how much text it can process at once. In March 2019, Baidu introduced Enhanced Representation through kNowledge IntEgration (ERNIE 1.0), an Natural Language Processing (NLP) model designed to enhance language understanding through knowledge integration (*Sun et al., 2019*). Using advanced techniques like entity and phrase-level masking, it learns more complex language patterns, achieving exceptional results in Chinese NLP tasks like sentiment analysis, question answering, and language inference. While it primarily focuses on Chinese NLP, its effectiveness is limited for other languages. One month later, OpenAI introduces MuseNet, a powerful AI that creates music by blending styles from country to Mozart to the Beatles (*OpenAI, 2019*). Using a large transformer model, it predicts the next note in a sequence, allowing it to compose multi-instrument pieces with remarkable versatility. Nonetheless, it lacks real-time interaction, fine-grained control, and sometimes produces inconsistent results when mixing styles, due to its reliance on statistical pattern prediction rather than true musical understanding. A few months later, in July 2019, Baidu introduced ERNIE 2.0, an upgraded NLP model with continuous pre-training and multi-task learning (*Sun et al., 2020*). Unlike its predecessor, it learns gradually across multiple tasks, allowing it to better understand vocabulary, syntax, and semantics. This advanced approach helps ERNIE 2.0 outperform BERT, particularly in Chinese language processing. While ERNIE 2.0 excels in language understanding, especially in Chinese NLP, it still struggles with long texts due to its fixed context window, limiting its ability to process extended passages effectively.

2020 to 2022

Introduced in June 2020, GPT-3 marks a major leap in GAI with 175 billion parameters, enabling impressive generalization through few-shot and zero-shot learning (*Brown et al., 2020*). It excels at tasks like translation, summarization, and coding but can sometimes produce biased, incorrect, or irrelevant outputs due to limitations in its training data and context recognition. In September 2020, researchers from Meta AI introduced Retrieval-Augmented Generation (RAG) technique, which enhances the capabilities of LLMs by integrating external data retrieval processes (*Lewis et al., 2020*). This approach allows LLMs to access up-to-date information from various sources, improving the accuracy and reliability of generated responses by grounding them in real-world data. At the start of 2021, OpenAI developed and initially released DALL-E to generate images from textual descriptions by adapting the text-generation abilities of GPT-3 to handle both text and image combinations (*OpenAI, 2021c*). Despite its impressive ability to generate creative and diverse images, DALL·E often struggles with fine details and high-resolution output. Although it demonstrates an understanding of spatial relationships, it occasionally produces distorted or unrealistic compositions. Furthermore, its outputs may inadvertently reflect biases present in the training data. In mid-2021, Google introduced Language Model for Dialogue Applications (LaMDA) (*Thoppilan et al., 2022*), a language model designed for more natural and coherent conversations. Unlike traditional models, it focuses on improving context and logical flow, making it ideal for chatbots, virtual assistants, and customer support. LaMDA excels in conversational AI but has limitations, including difficulty retaining long-term context, potential biases in responses, and occasional inaccuracies due to outdated or incomplete training data. Released in July 2021, ERNIE 3.0 advances NLP with a 10-billion-parameter architecture that merges auto-regressive and auto-encoding networks. This design enhances both language understanding and generation, setting new benchmarks in Chinese NLP tasks (*Wang et al., 2021*).

For code generation, OpenAI developed Codex, an AI model based on GPT-3, to generate code from text descriptions (*Chen et al., 2021*). It learns from public datasets, including GitHub, and powers GitHub Copilot in October 2021, which helps developers with code suggestions and automation (*GitHub Copilot, 2021*). OpenAI restricts access in March 2023 due to misuse concerns but later allows researchers to use it under a controlled program. Shortly after, in November 2021, Microsoft proposed a comprehensive multimodal pre-trained model, known as Neural visUal World creAtion (NUWA) (*Wu et al., 2022*). This model can produce novel or modify existing visual data for various visual synthesis applications, including images and videos. Although NUWA is capable of generating both images and videos, it faces notable limitations. It often struggles to produce highly detailed or contextually rich visuals, particularly in scenarios that require deep semantic understanding. Building on the progress of earlier models of GPT, OpenAI introduce GPT-3.5 in January 2022 as an evolution of GPT-3. This version includes three variants, each with different parameter sizes: 1.3 billion, 6 billion, and 175 billion. Commonly referred to as InstructGPT, it shares similar training data with GPT-3 but incorporates an important advancement through fine-tuning using Reinforcement Learning with Human Feedback (RLHF) (*Christiano et al., 2017*) This method enhances the model's ability to follow user instructions, reduces harmful biases, and improves the overall quality of responses. Nonetheless, GPT-3.5 still faces certain limitations, including occasional factual inaccuracies, inconsistencies, and difficulty maintaining coherence in extended conversations.

MidJourney, developed by an independent research lab in San Francisco, is an AI tool that generates images from text prompts. The model performs well in artistic image generation but has limitations in detail control, object placement, and text rendering, with full access restricted to paid Discord users. In April 2022, Google announced PaLM (Pathways Language Model), a powerful transformer-based language model with 540 billion parameters (*Chowdhery et al., 2022*). It remains confidential until March 2023, when Google releases its API. Unlike LaMDA, which specializes in conversations, and BERT, which focuses on understanding text, PaLM combines reasoning, coding, and language skills, making it more versatile. However, its massive size makes it expensive to run, and it still struggles with accuracy and maintaining context in long conversations (*Narang & Chowdhery, 2022*). Later in April, Boris Dayma created Craiyon, a free AI tool for text-to-image generation. It evolves with community contributions and internal improvements (*Craiyon, 2022*). Originally called DALL-E mini, it was renamed after OpenAI raised concerns about its similarity to their DALL-E model. While Craiyon stands out for its ease of access and open availability, it typically produces lower-quality images compared to other state-of-the-art generators. Around the same time, in April 2022, OpenAI releases DALL-E 2, a text-to-image model built on Contrastive Language-Image Pretraining (CLIP) and a diffusion model (*OpenAI, 2021a*; *Radford et al., 2021*). It generates highly realistic images and improves resolution compared to its predecessor, DALL-E (*OpenAI, 2022*).

In May 2022, Meta AI Research introduced Open Pre-trained Transformer (OPT-175B), a 175-billion-parameter language model trained on 180 billion tokens (*Zhang et al., 2022*). It matches GPT-3 in performance but is more energy-efficient, using only one-seventh of GPT-3's training carbon footprint. Despite its technical strengths, the model, like many LLMs, continues to face limitations, particularly concerning biases and factual inaccuracies. In a commitment to advancing transparent and responsible AI research, Meta releases the model's code, pre-trained weights, and comprehensive training logs. However, its distribution is restricted by a non-commercial license, making it available primarily to researchers in academic institutions, government agencies, and civil society organizations. During the same month, the Knowledge Engineering Group (KEG) and data mining THUDM at Tsinghua University presented a large-scale pre-trained text-to-video generation model, named CogVideo. CogVideo is built upon the transformer architecture to create videos based on text descriptions in the Chinese language and offers compatibility with various video formats (*Hong et al., 2022*). Also, in May, Google introduces Imagen, a text-to-image diffusion model built on Text-To-Text Transfer Transformer (T5) for text encoding and a cascaded diffusion process for image synthesis. It joins AI-driven generators like DALL-E 2, focusing on photorealism and producing sharper, more lifelike images (*Saharia et al., 2022*). While Imagen delivers high realism and precise text alignment, DALL·E 2 offers greater creative flexibility and advanced editing features like inpainting.

In June 2022, Microsoft released Phi 1 as its first LLM. It is a Transformer-based architecture with 1.3 billion parameters, specifically designed for basic Python coding tasks (*Microsoft Azure, 2024*). Despite being efficient for basic Python coding, Phi-1's small size limits its ability to handle complex tasks and broader language understanding. One month later, Hugging Face and BigScience introduced Bigscience Large Open-science Open-access multilingual language Model (BLOOM) as an open-source alternative to GPT-3 (*Scao et al., 2022*). BLOOM has 176 billion parameters and is trained on the ROOTS *corpus*, supporting 46 natural and 13 programming languages with a strong emphasis on linguistic diversity. Compared to GPT-3, it offers full transparency under the

Responsible AI License (RAIL), ensuring ethical use. Despite its open nature, BLOOM demands significant computational resources, struggles with biases and maintaining long-term coherence, and, due to its accessibility, carries the risk of misuse.

In August 2022, Stable Diffusion emerges as a powerful text-to-image AI model based on diffusion techniques (*Stability AI, 2022*; *Scao et al., 2022*). Developed by researchers from the CompVis Group at Ludwig Maximilian University of Munich and Runway, with funding from Stability AI, it specializes in generating highly detailed images from text prompts. Beyond text-to-image generation, it also supports image-to-image translation and inpainting, allowing users to modify or enhance images based on descriptions. While Stable Diffusion is highly effective at generating detailed and customizable images, it still requires significant computing power, struggles with generating text within images, and sometimes produces biased or inconsistent results. Following this, in September 2022, Nvidia introduced GET3D, a generative model built on GAN-based architectures to create high-quality 3D textured models from images (*Gao et al., 2022*). It specializes in producing detailed 3D representations of characters, buildings, vehicles, and other objects with impressive textures and intricate geometric details. Although GET3D generates high-quality 3D models, it requires large datasets, struggles with complex objects, and demands high computing power. In the same month, Meta AI introduced Make-A-Video, a GAI model based on diffusion techniques and transformer architectures (*Singer et al., 2022*). It creates short videos from text prompts, making it useful for animation, music videos, and short films. While it enables creative content generation, it struggles with fine details, consistency between frames, and realistic motion, often requiring post-processing to improve quality. Expanding the wave of innovation, Google AI launched Imagen Video the following month, a text-to-video model that leverages a cascade of diffusion models to generate high-quality video content from textual descriptions. It supports various artistic styles and 3D object understanding but struggles with artifacts in complex scenes and motion consistency (*Google, 2022*; *Ho et al., 2022*). By November 2022, the focus turned toward conversational AI with OpenAI's release of ChatGPT, built on the GPT-3.5 architecture. The model attracted widespread interest for its ability to generate coherent, context-aware, and human-like dialogue (*OpenAI, 2023c*).

2023

In early 2023, Runway Company released its video generation model, Gen-1, to produce novel videos based on pre-existing ones (*runway, 2023a*). The Gen-1 can modify videos using visual or textual descriptions, enabling users to generate novel films without physically filming or editing them. However, it comes with challenges like heavy computing requirements, and occasional inconsistencies in quality. Around the same time, Meta AI introduced LLaMA 1, a more efficient Transformer-based language model available in 7B to 65B parameters. LLaMA delivers strong performance with fewer resources, making it more accessible for researchers compared to OPT. Nevertheless, it remains restricted for commercial use and still faces bias and accuracy challenges (*Meta AI, 2023*; *Touvron et al., 2023*). By March, Runway released Gen-2, an enhanced version of Gen-1 that allows users to generate short videos from text prompts or animate images (*runway, 2023b*). This upgrade improves realism and creative flexibility, making it easier to bring ideas to life. Gen-2 also offers real-time feedback and versatile input options, whether through language instructions or uploaded visuals. Despite its strengths, it still demands high computational power and struggles to maintain consistency in longer video clips. In March, Google introduces Bard (*Bard, 2023*), an AI chatbot that assists users by answering questions and generating ideas in a conversational way. Initially built on LaMDA, Bard later transitions to more advanced AI models, eventually being rebranded as Gemini. While it provides helpful insights and creative support, it sometimes struggles with accuracy, biases, and complex reasoning (*Google, 2023*). During the same month, OpenAI launched GPT-4, the latest in its GPT series, offering improved reasoning, creativity, and contextual understanding over GPT-3.5 (*OpenAI, 2023b*; *OpenAI, 2023d*). A key upgrade is its multimodal capability, allowing it to process both text and images. GPT-4, available *via* ChatGPT Plus and OpenAI's API, delivers more accurate and coherent responses than earlier versions. While it still encounters issues like bias, occasional errors, and high computational demands, it remains one of the most advanced AI models, excelling in complex tasks and nuanced dialogue. Around the same time, Anthropic, an AI startup company, introduces Claude AI, a chatbot built on the Claude 1.3 language model using a Transformer-based architecture. Claude performs well in text generation, translation, summarization, and question answering, while emphasizing ethical alignment and user safety. It also prioritizes transparency and reducing harmful outputs. Still, it has limitations, including occasional inaccuracies and difficulty with complex reasoning (*Anthropic, 2023b*; *Card, 2023*).

In May, the Technology Innovation Institute (TII) in Abu Dhabi created its first LLM called Falcon under the Apache 2.0 license (*Falcon LLM & Technology Innovation Institute (TII), 2023*). Falcon is an open source with 40 billion parameters, updated later to 180 billion and is considered a strong alternative for developers looking for an open-source and customizable LLM. However, compared to other LLMs released around the same time, it cannot process images, struggles with complex reasoning, and may need more fine-tuning for specific tasks. It also has challenges with long conversations and inherited biases. During this period, IBM unveiled Watsonx.ai to help businesses develop and scale GAI (*IBM Newsroom, 2023*). The platform supports multiple LLMs, including Hugging Face models, Meta's LLaMA 2, and IBM's Granite models (*IBM Research, 2023*), introduced later in September 2023. It offers flexible AI tools for various applications. Meanwhile, in the same month, Google DeepMind built PaLM 2, a sophisticated LLM (*Google AI, 2023*). PaLM 2 is built on a Transformer-based architecture, leveraging Google's Pathways system for efficient training and scaling. It improves multilingual understanding, reasoning, and code production over PaLM. Google Bard uses it to improve conversational skills and language understanding, making it a key AI service in 2023 (*Anil et al., 2023*). In the following month, Baidu launched ERNIE 3.5, an upgraded version of ERNIE 3, powering the ERNIE Bot chatbot (*Baidu Research, 2023*). ERNIE 3.5 responds much faster, at 17 times the efficiency of its predecessor, and supports plugins like Baidu Search and ChatFile, making it a more versatile and practical AI assistant.

In July, Anthropic launched Claude 2, an upgraded version of its AI model with enhanced reasoning, problem-solving, and contextual understanding (*Anthropic, 2023a*). Compared to its predecessor, Claude 2 features a larger context window, improved memory retention, and better handling of complex queries. It also refines its alignment techniques to reduce biases and enhance safety, making responses more reliable. Around the same time, Meta, in collaboration with Microsoft, released LLaMA-2, an improved version of its language model with 7, 13, and 70 billion parameters. While its architecture remains similar to LLaMA-1, it benefits from a larger training dataset, enhancing its overall performance. In the next month, OpenAI introduces DALL-E 3, an upgraded version of DALL-E 2. It offers a better understanding of detailed prompts, allowing users to create more accurate and visually precise images with ease (*OpenAI, 2023a*). On September 27, Mistral AI, a French startup, released its first model, Mistral 7B—a compact 7-billion-parameter LLM optimized for speed and efficiency. Despite its smaller size, it offers strong performance and cost-effectiveness, though it may fall short on complex reasoning and long-context tasks compared to larger models (*Jiang et al., 2023*). In the next month, Google DeepMind introduced Imagen 3, the most advanced version of its text-to-image model. This model provides exceptional image quality, with substantially fewer distracting artifacts, richer lighting, and enhanced detail than previous models (*Google DeepMind, 2024a*). Later, on October 15, 2024, the model becomes available to users in the United States. On October 17, Baidu introduced ERNIE 4.0, a more advanced model offering improved language understanding, reasoning, and memory, with performance reportedly 30% stronger than ERNIE 3.5. It supports text, image, and video generation, and delivers more direct, context-aware responses. While well-integrated into Baidu's ecosystem, access remains limited to invited users in China, leaving its global impact yet to be seen (*Originality.ai, 2025*). On November 4, Elon Musk's company, xAI, launched Grok, an AI chatbot. Initially tested with a small group of users, Grok is designed to process real-time global information from Twitter's X platform (*X AI, 2023*). It can generate poetry, code, screenplays, music, emails, and more. While innovative, Grok has faced criticism for occasional unpredictability, sometimes providing rude or inappropriate responses as it continues to develop.

Later, on November 16, Google DeepMind, in partnership with YouTube, introduces Lyria, an advanced AI music model designed to generate creative and original compositions XAI Grok (*X AI, 2023*). It offers better control over musical structure and style compared to OpenAI's MuseNet, making it more suited for professional music creation. However, it still faces challenges in capturing emotional depth and maintaining coherence in longer compositions (*Google DeepMind, 2023*). Five days later, Stability AI's Stable Video Diffusion allowed developers to generate short video clips, typically between 2 to 5 s, with frame rates up to 30 FPS (*Stability AI, 2023*). While it efficiently generates frames, it struggles with longer sequences, photorealism, and maintaining smooth motion. However, its easy accessibility through Stability AI's platform makes it a useful tool for AI-driven video experimentation. On November 28, Pika launched its AI video generation platform, founded by Stanford researchers Demi Guo and Chenlin Meng. The platform allows users to create videos from text prompts, supporting various styles like 3D animation and cinematic effects. It offers real-time editing and clip extension features. While Pika is easy to use and popular, it struggles with producing high-quality, professional (*Pika, 2023*). In early December, Google DeepMind introduced Gemini, a powerful multimodal AI model capable of processing text, images, audio, and video. With enhanced reasoning, coding, and contextual understanding, it marks a significant step beyond previous Google models. It comes in three versions: Ultra, Pro, and Nano, each suited to different use cases. Despite its strengths, Gemini still requires substantial computational resources and can struggle with complex reasoning. On 11 December, Mistral AI launches La Plateforme, its beta platform for developers. It offers powerful open generative models with easy deployment and customization options (*Mistral AI, 2023*). The platform includes three conversation endpoints, two for text generation and one for embedding, each balancing cost and performance differently. This gives users flexibility based on their needs. Later in December, Microsoft introduces Phi-2, a 2.7 billion-parameter model trained on synthetic data. Designed for research, it excels in reasoning and language tasks while reducing bias and toxicity. Despite its smaller size, it outperforms larger models in some benchmarks but struggles with complex tasks. Fine-tuning may be needed for specialized applications (*Microsoft Research, 2023*).

2024

In early 2024, NVIDIA launched Chat with RTX, a local AI chatbot for Windows PCs. It offers better privacy and control than cloud-based models but depends on hardware capabilities and lacks scalability (*NVIDIA, 2025*). On 21 February, Google DeepMind On February 21, Google launched Gemma, a lightweight open-source AI model for developers and researchers. Supporting over 140 languages, it balances efficiency and accessibility. While PaLM and Gemini are more powerful, they require more resources. Gemma's customizability makes it a valuable tool for flexible AI applications (*Google AI for Developers, 2023*). Three days later, OpenAI introduced Sora, a powerful text-to-video model that blends diffusion and transformer-based architecture, similar to GPT. It creates high-quality, up to 20-second-long videos with impressive motion and scene complexity. Sora refines user prompts using GPT-based recaptioning, ensuring greater accuracy in video generation. While it surpasses models like Runway's Gen-2 and Stable Video Diffusion in realism, it still faces challenges with physics accuracy and object interactions. Sora is accessible only to select professionals for safety testing, with plans for broader availability in the future (*OpenAI, 2024b*).

On February 26, Mistral AI launched Mistral Large, its most advanced language model, designed for complex reasoning tasks like code generation, text transformation, and analysis. Alongside it, Mistral AI introduced Le Chat, a conversational assistant that allows users to choose between Mistral Large, Mistral Small, or Mistral Next for more concise responses. Le Chat features customizable moderation, providing non-intrusive alerts for sensitive or controversial content. While Mistral Large excels in precision and adaptability, it may require fine-tuning for highly specialized tasks and faces competition from larger proprietary models (*Mistral AI, 2024*). On April 23, Microsoft launched Phi-3, a 3.8 billion-parameter model with a dense decoder-only Transformer design. It supports a

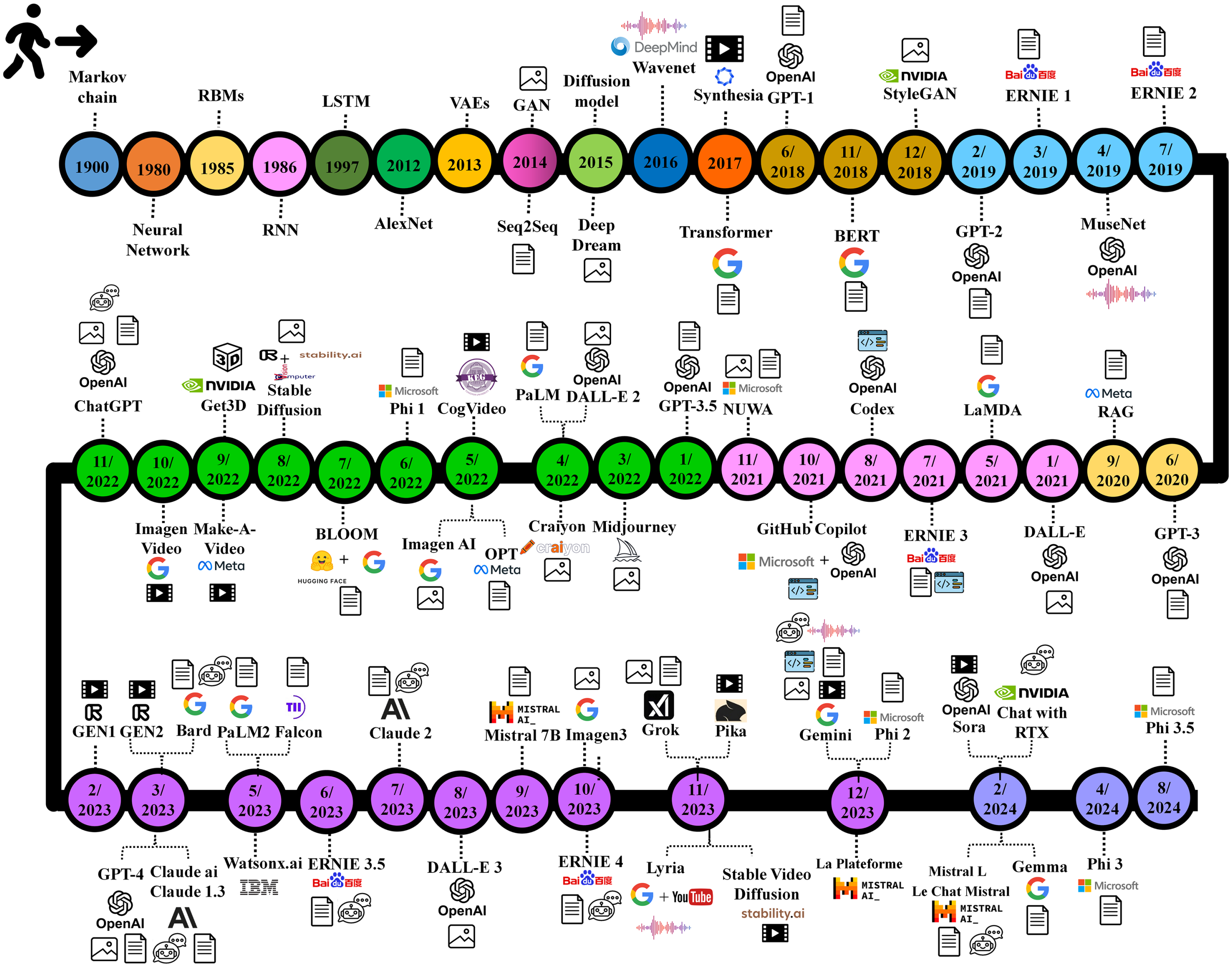


Figure 3 GAI timeline.

Full-size DOI: 10.7717/peerj-cs.3276/fig-3

128,000-token context window, enabling it to handle complex language tasks while competing with larger models like GPT-3.5 (*Abdin et al., 2024*). Later in August, Microsoft released Phi-3.5, introducing three variants: mini-instruct for fast reasoning, MoE with a 42-billion-parameter Mixture-of-Experts architecture that activates only 6.6 billion parameters per use, and vision-instruct for multimodal text and image tasks. The Phi-3 series advances AI efficiency, maintaining high performance, safety, and scalability in resource-constrained environments Phi-3.5 SLMs (*Trufinescu, 2024*).

Figure 3 presents a timeline showcasing the development stages of GAI, and the following sections introduce further details regarding GAI models, tools, applications, challenges, and ethical implications.

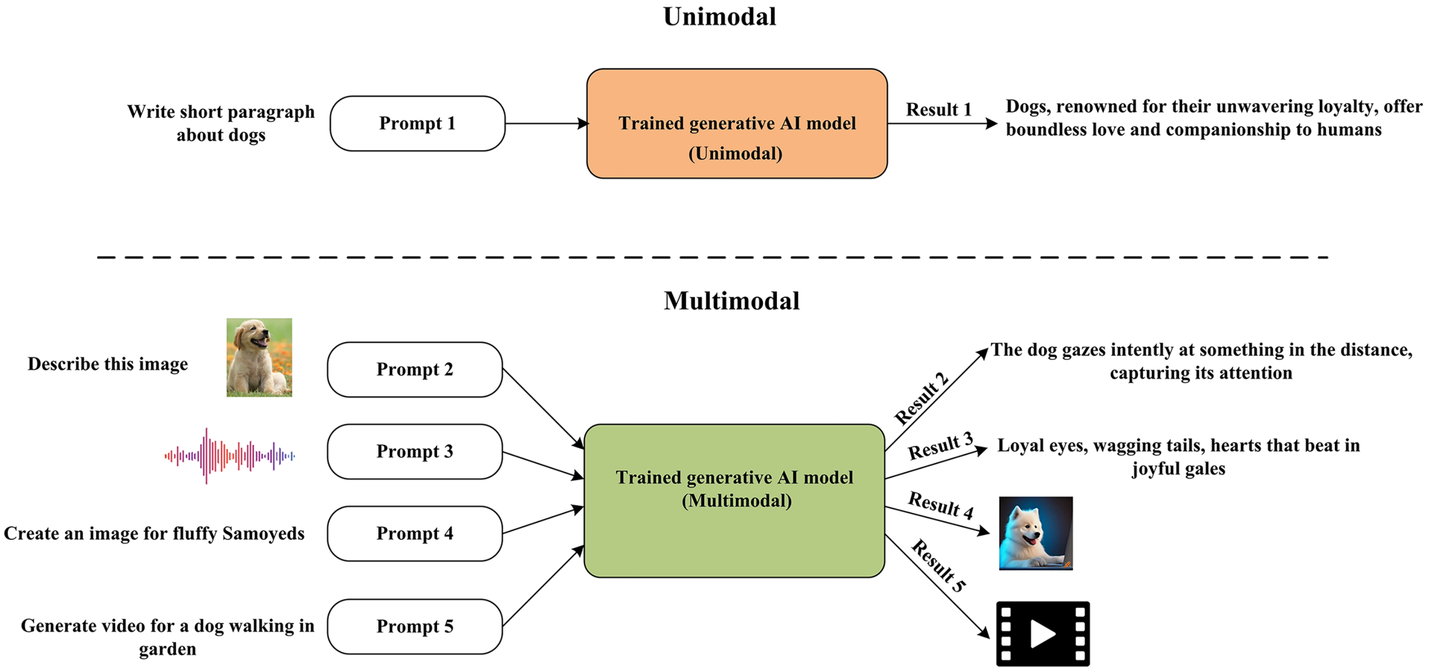

Unimodal
Write short paragraph about dogs
Prompt 1
Trained generative AI model (Unimodal)
Result 1
Dogs, renowned for their unwavering loyalty, offer boundless love and companionship to humans
Multimodal
Describe this image
Prompt 2
Prompt 3
Create an image for fluffy Samoyeds
Prompt 4
Generate video for a dog walking in garden
Prompt 5
Trained generative AI model (Multimodal)
Result 2
The dog gazes intently at something in the distance, capturing its attention
Result 3
Loyal eyes, wagging tails, hearts that beat in joyful gales
Result 4
Result 5

Figure 4 GAI model types.

Full-size DOI: 10.7717/peerj-cs.3276/fig-4

GENERATIVE AI MODELS

The field of GAI comprises a range of models designed to produce novel data or content that closely mimics human-generated data. Multimodal and unimodal models are two distinct categories of GAI models that process input data in various formats. Unimodal GAI models have been developed to process information in one data format, usually text. These models are designed to process a text prompt and provide text-based replies. Unimodal GAI models have proven effective in various domains, including text generation as GPT3 and language translation. Unimodal GAI models may not be the best choice for activities requiring processing other sorts of data, such as photos or audio. Multimodal GAI models are more suited for these kinds of tasks. Multimodal models can simultaneously process diverse input data, including text, image, audio, or a fusion of these modalities. OpenAI's CLIP, DALL-E, and GPT-4 are examples of Multimodal models. Figure 4 presents and summarizes the difference between the GAI model types.

Furthermore, Fig. 5 presents the landscape of LLMs and their foundational technologies which showcases the rapid advancements in AI and natural language processing. This landscape illustrates the diverse range of models and techniques that have emerged, reflecting the innovative approaches driving the field forward. At the top of this landscape, ChatGPT ranks among the prominent models, joined by other noteworthy examples such as GitHub Copilot, Stable Diffusion, Imagen, Pika, and DALL-E, *etc*. These models are built on state-of-the-art techniques, including VAEs, diffusion models, transformers, and GANs, *etc*. The middle layer highlights various LLMs, including GPT-1, 2, 3, and 4, along with OPT, PaLM, NUWA, BERT, and Claude. These models, developed over time, have significantly contributed to the evolution of AI and natural language processing. The

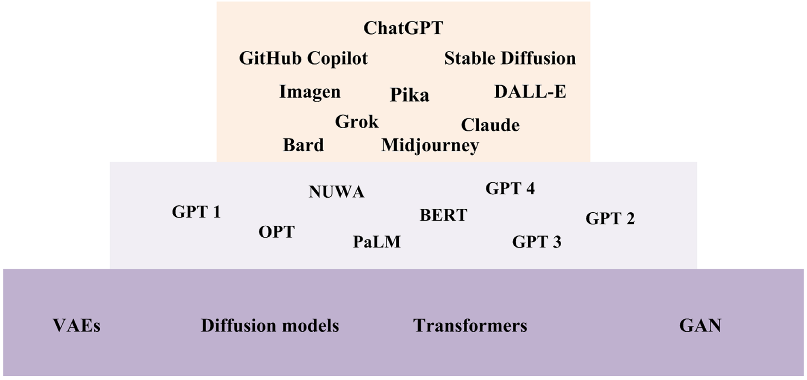


Figure 5 **LLMs and their foundational technologies landscape.**
Full-size DOI: 10.7717/peerj-cs.3276/fig-5

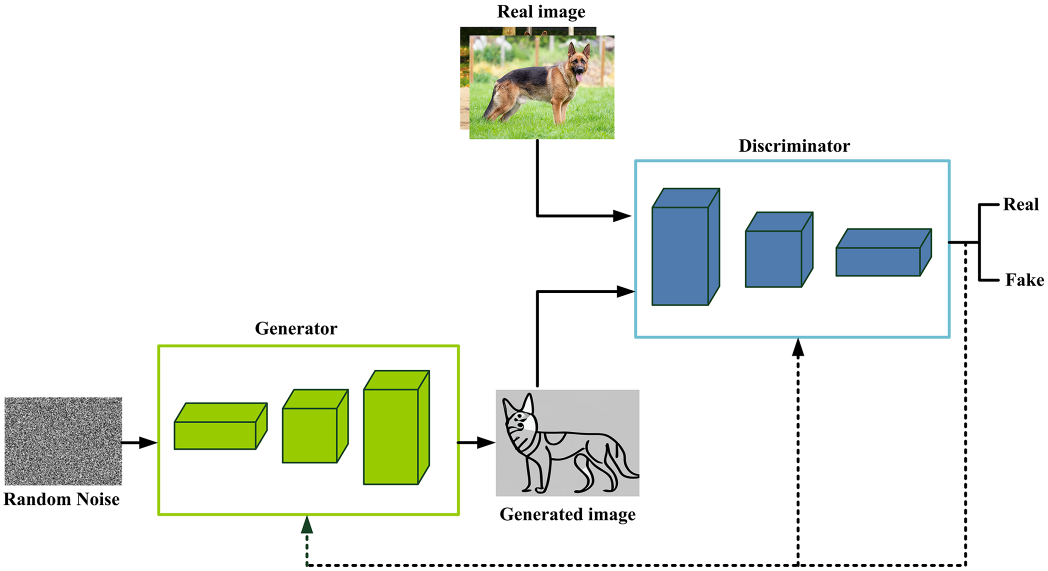


Figure 6 **GANs architecture.** Full-size DOI: 10.7717/peerj-cs.3276/fig-6

following subsections provide an overview of the cutting-edge techniques that serve as the foundation for developing these advanced tools.

Generative adversarial networks

GANs are first introduced in 2014 by Ian Goodfellow and his colleagues (*Zhong et al., 2023*). GANs are a prominent category of generative techniques that have gained significant traction in AI, especially in image, video, and data generation (*Isola et al., 2017*). GANs consist of two deep neural network models: a discriminator and a generator, as shown in Fig. 6. The generator is a CNN, and its function is to generate fake data that looks like real data. Contrarily, the discriminator is a CNN, and its function is a binary classifier. It attempts to differentiate between real data and created data from the generator. The adversarial aspect of GANs is rooted in a game theoretic framework, whereby the

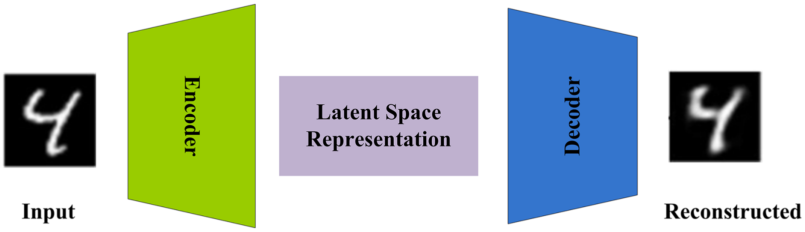


Figure 7 **VAEs architecture.** Full-size DOI: 10.7717/peerj-cs.3276/fig-7

generator network engages in a competitive interaction with the opponent. The generator network is responsible for generating fake data. The discriminator network, which opposes it, differentiates data originating from the generator and those originating from the training data. This process fosters an ongoing competitive dynamic, wherein the update of one network upon failure occurs while the other network remains unaffected, embodying an iterative refinement characteristic of GAN architectures.

GAN training occurs in alternating phases. First, the discriminator is trained for a set number of epochs. Subsequently, the generator is trained for a defined number of epochs. This sequence iterates, cycling between training the discriminator and the generator networks, enabling continual refinement of both networks. During the training phase of the discriminator, the generator remains frozen. Conversely, when the generator is trained, the discriminator remains constant and unaltered. The generator is provided with a loss signal through feedback from the discriminator. This loss signal indicates the discriminator's proficiency in discerning genuine data from artificially created data. By minimizing this loss, the generator can produce data that exhibits a progressively higher similarity to the actual data. Numerous GAI tools and frameworks have been developed utilizing the principles of GANs, such as DALL-E, DeepAI, and Pix2Pix (*Arora, Risteski & Zhang, 2017*). Despite GANs strengths, it often struggles with mode collapse, where the generator repeatedly produces similar outputs instead of diverse ones. Training can also be unstable, especially if the discriminator learns too fast, making it harder for the generator to improve. Additionally, GANs are highly sensitive to hyperparameters, meaning small changes in learning rate or batch size can greatly affect performance (*Arora, Risteski & Zhang, 2017*; *Saxena & Cao, 2020*; *Bengesi et al., 2024*).

Variational autoencoders

VAEs, or variational autoencoders, are a subclass of generative models that can identify latent data representations (*Yacoby, Pan & Doshi-Velez, 2020*). VAEs comprise two distinct neural networks, an encoder, and a decoder, as shown in Fig. 7. The data sample is fed to the encoder for encoding. After that, it produces a latent representation for this data, which refers to a numerical vector that effectively encodes the fundamental characteristics of the data. The decoder rebuilds the original input data using this latent representation. VAEs are used to grasp the inherent probability distribution within a dataset, enabling the generation of new data samples that adhere to that distribution. An encoder-decoder architecture is employed, wherein the encoder transforms input data into a latent representation, and the decoder endeavors to reconstruct the initial data from this encoded representation. During training, the VAE's objective is to become skilled at modeling the data's underlying distribution and generating new samples that conform to it. This is achieved by minimizing the dissimilarity between the original and reconstructed data. Through this training process, the VAE can understand the inherent data distribution. When training VAEs, a variational objective function is employed, consisting of two key loss components: the reconstruction loss and the Kullback-Leibler (KL) divergence. The reconstruction loss assesses the decoder's ability to faithfully recreate the original input data, serving as a performance measure for the decoder.

On the other hand, the KL divergence quantifies how far the latent representation deviates from a standard normal distribution. The primary goal during VAE training is to minimize the variational objective function using techniques like gradient descent. This minimization process empowers the encoder to generate informative yet concise latent representations. Simultaneously, the decoder learns the knowledge needed to reconstruct the initial input data from these latent representations accurately. The overarching aim is to make the VAE adept at representation and reconstruction tasks, resulting in a more effective generative model. VAEs find applications across a broad spectrum. They are frequently harnessed for generating images, as they enable the generation of new and unique images through latent space sampling. In addition, VAEs serve various purposes, including data compression, anomaly detection, and data imputation. Numerous GAI tools leverage VAEs as a fundamental element. Here are some prominent tools that incorporate VAEs for diverse generative purposes: DeepArt.io (*DeepArt.io, 2015*), DCGAN (*Radford, Metz & Chintala, 2015*), DreamStudio (*DreamStudio, 2022*) and Pix2Pix (*Tahmid et al., 2016*). Although VAEs offer stable training and can generate new data, their outputs often appear blurry due to difficulties in capturing fine details. They struggle with high-resolution images and complex data patterns, making them less effective for photorealistic generation. Their latent space is hard to interpret, limiting control over outputs. These challenges make VAEs less suitable for tasks requiring high-detail or sharp images (*Yacoby, Pan & Doshi-Velez, 2020*; *Daunhawer et al., 2021*).

Diffusion model

Diffusion models, as generative models, have seen a considerable increase in popularity in recent years over GAN and VAEs. GANs demonstrate excellent performance across various applications. Nevertheless, their training process poses challenges, leading to limited diversity in generated outputs due to issues like mode collapse and vanishing gradients (*Saxena & Cao, 2020*). On the other hand, constructing an effective loss function for VAEs remains a hurdle, resulting in suboptimal outputs. Diffusion models are predicated on the thermodynamics of gas molecules, which states that molecules diffuse from regions of high density to those of low density (*Sohl-Dickstein et al., 2015*; *Chang, Koulieris & Shum, 2023*). In the theory of information, this corresponds to information loss caused by the incremental introduction of noise. Fundamentally, Diffusion Models function by perturbing training data iteratively by adding Gaussian noise, subsequently learning to reconstruct the original data by reversing this noise-induced transformation.

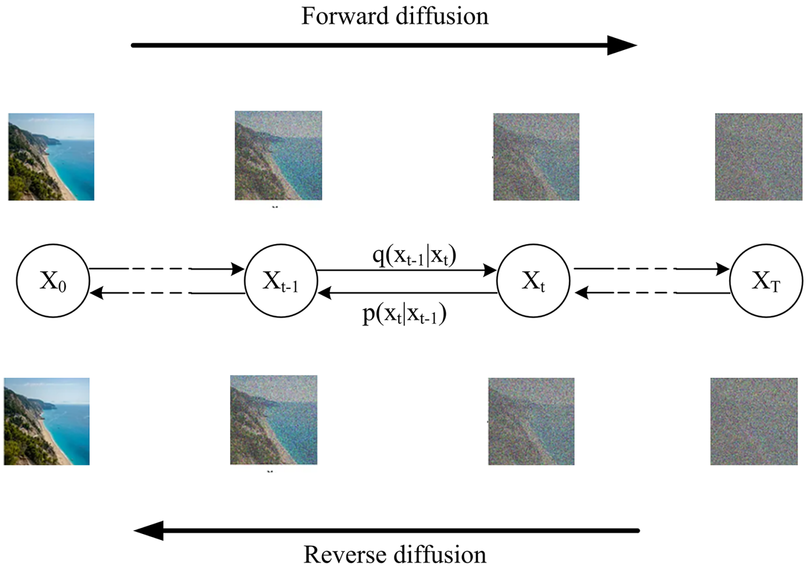


Figure 8 **Diffusion model process.** Full-size DOI: 10.7717/peerj-cs.3276/fig-8

Post-training, the Diffusion Model becomes capable of generating data by passing randomly sampled noise through the acquired knowledge of the denoising procedure. The Diffusion Model, in its essence, operates as a latent variable model employing a fixed Markov chain to traverse the latent space. This stationary chain systematically introduces incremental noise to the data to approximate the posterior distribution, as shown in Fig. 8. Ultimately, the image undergoes an asymptotic transformation into pure Gaussian noise. The primary goal of training a diffusion model is to learn the inverse process, which enables the generation of new data by moving backward through the chain. As a result, diffusion models excel in GAI by creating and manipulating high-quality images. Many tools, such as DALL-E 2, Imagen, DreamStudio (*DreamStudio, 2022*), Artbreeder (*Artbreeder, 2025*), and GauGAN2 (*NVIDIA, 2021*), leverage diffusion models to offer a wide range of applications, including text-to-image and image editing. Unlike GANs, diffusion models provide stable training, avoiding instability and the blurry outputs of VAEs. However, while diffusion models produce realistic and detailed images, they tend to be slower than GANs. Nevertheless, advancements continue to improve their efficiency. Despite these improvements, they still require high computational power, making them resource-intensive for both training and inference (*Chen et al., 2024*).

Transformer-based models

A transformer is a novel neural network introduced in 2017 that utilizes the parallel multi-head attention mechanism (*Vaswani et al., 2017*). The transformer is used to convert one sequence into another. However, it distinguishes itself from conventional Seq2Seq models by not relying on Recurrent Networks such as GRU or LSTM. Because RNNs have a smaller reference window, they cannot access words generated earlier in the sequence as stories get longer. Even though GRU and LSTM have a greater capacity to achieve longer-term memory and, consequently, a longer window to reference from, they still face

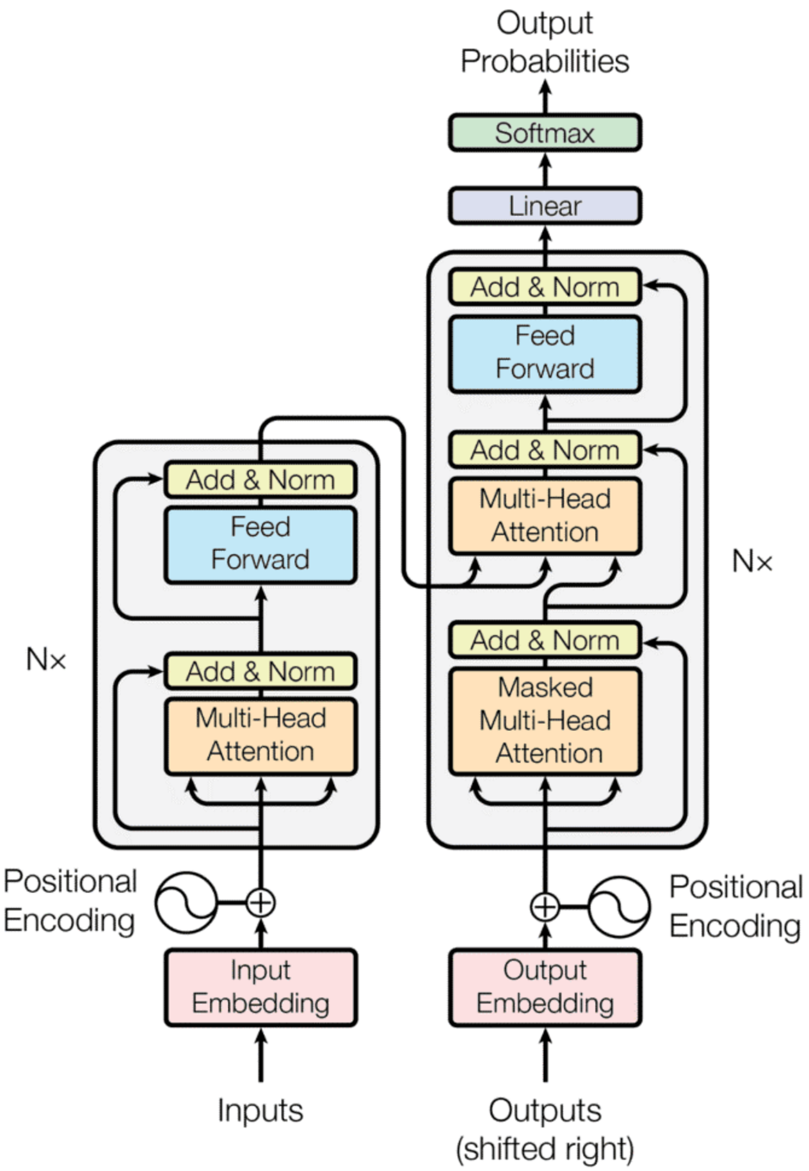


Figure 9 **Transformer architecture.** Full-size DOI: 10.7717/peerj-cs.3276/fig-9

the issue of accessing words generated before. The strength of the attention mechanism of the transformer lies in its lack of short-term memory problems. The attention mechanism can reference an unlimited information window with sufficient computational resources. This allows it to utilize the entire context of the story when generating text.

The transformer structure consists of an encoder and Decoder, as presented in Fig. 9. Encoder and Decoder are modules that may be layered on each other several times, as shown in the figure by Nx. The encoder's role, situated on the left side of the Transformer architecture, is to convert an input sequence into a series of continuous representations. These representations are subsequently sent to the Decoder. The Decoder, located on the right side of the architecture, takes in the output of the encoder and the preceding time

Table 1 **Comparison of GANs, VAEs, diffusion models, and transformers in terms of architecture, performance, applications, and limitations.**

Model	Architecture	Performance	Applications	Limitations
GANs	Generator + Discriminator	High-quality but unstable	Image generation, style transfer	Mode collapse, unstable training, high computational costs, ethical issues (*e.g.*, deepfakes)
VAEs	Encoder + Decoder (Latent Space)	Stable training, but blurry outputs	Anomaly detection, image compression	Less sharp images, not fully generative, struggles with fine details, suboptimal for complex datasets
Diffusion models	Noising + Denoising process	High-quality and stable outputs	Image generation, video synthesis	Slow generation, high computational cost, limited understanding of conditional diffusion models
Transformers	Encoder–Decoder (NLP), Vision transformers	Excellent long-range understanding	NLP, computer vision, bioinformatics	Resource-heavy, struggles with long-range dependencies, overfitting on small datasets, lacks recursive computation for hierarchical structures

step's decoder output to produce an output sequence. Initially, the input is sent through a word embedding layer. A word embedding layer may be conceptualized as a retrieval table that retrieves a learned vector representation for each word. Due to the absence of recurrence in the transformer encoder, positional information must be incorporated into the input embeddings. Thus, Positional encoding is utilized by employing sine and cosine functions. Now, the input embeddings with their positions are fed to the encoder, which has two sub-modules: multi-headed attention and a fully connected network. In addition, residual connections surround each of the two sublayers, which are then followed by a normalization layer. The Multi-headed attention is a component within the transformer network that calculates the attention weights for the input and generates an output vector containing encoded information on how each word should prioritize its attention toward all other words in the sequence. Consequently, the feedforward layer transforms the attention outputs, potentially enhancing their representation. The Decoder is autoregressive, commencing with a start token and accepting a sequence of prior outputs as inputs, in addition to the encoder outputs that encompass the attention information derived from the input. The decoding process terminates when a token is produced as an output. Transformers pave the way for powerful models like BERT, GPT series, T5 (*Raffel et al., 2020*) and PaLM, revolutionizing NLP. Though initially limited in high-fidelity image and video generation, newer models integrate Transformer elements to enhance performance. However, they demand high computational power, particularly for long sequences, though later optimizations help improve efficiency (*Fan et al., 2020*).

Comparing GANs, VAEs, diffusion models, and transformers

Different AI models perform better in specific tasks based on their strengths and limitations. GANs create realistic images and are great for style transfer, but they often face training instability and mode collapse, limiting diversity in their outputs (*Bengesi et al., 2024*). VAEs offer stable training and are useful for anomaly detection and data compression, though their images can be blurry. Diffusion Models generate highly detailed content, especially in text-to-image tasks, but they are slower and computationally demanding. Transformers, known for natural language processing, are now advancing in computer vision, but they require large datasets and high computing power. Table 1 provides a good overview of the key characteristics and limitations of GANs, VAEs, Diffusion Models, and Transformers.

GAI TOOLS

Recently, numerous companies and startups have developed software frameworks and platforms that leverage foundational GAI models to generate text, speech, images, video, music, code, and scientific content. Since the rise of ChatGPT, the development of GAI tools has surged, expanding beyond research into everyday use and commercial applications. Businesses and individuals are increasingly adopting these tools to boost productivity and streamline workflows. Table 2 offers a comprehensive overview of a wide range of GAI tools, as identified by the author at the time of writing. This includes well-known language models like ChatGPT (*OpenAI, 2023c*), ChatGPT-4o (*OpenAI, 2024a*), Claude (*Anthropic, 2023a*), Gemini (*Bard, 2023*; *Google, 2023*), Grok 3 (*X AI, 2023*), DeepSeek (*DeepSeek, 2023*), and Google NotebookLM (*Google, 2023*), along with research- and productivity-focused platforms such as Perplexity (*Perplexity.ai, 2022*), Microsoft Copilot (*Microsoft Copilot, 2023*), Copilot (*Microsoft Copilot, 2023*), Cohere (*Cohere, 2022*), and Bardeen.ai (*Bardeen AI, 2025*). Tools designed for code generation, such as GitHub Copilot (*GitHub Copilot, 2021*) and AlphaCode (*DeepMind, 2022*), are also included.

In addition to text and code, the table highlights AI tools that work with images and video, including DALL.E Mini (*Bayma & Cuenca, 2021*), DALL-E (*OpenAI, 2021c*), DALL-E 2 (*OpenAI, 2022*), DALL-E 3 (*OpenAI, 2023a*), Midjourney (*Midjourney*), Crayion (*Craiyon, 2022*), Stable Diffusion (*Stability AI, 2022*), Type Studio (*DreamStudio, 2022*), Designs.ai (*Designs AI, 2019*), Runway GEN (*Runway Research, 2024*), Pika (*Pika, 2023*), Descript (*Descript, 2017*), Synthesia (*Synthesia, 2023*), and Imagen Video (*Google, 2022*). For audio and music generation, tools like Soundraw (*SOUNDRAW, 2021*), Lyria (*Google DeepMind, 2024b*), and Murf.ai (*Murf AI, 2020*) are featured. Each tool is summarized in terms of its release date, underlying model, affiliated company, estimated usage cost, and key limitations. These tools differ in how they take input and generate output, with applications that span from chat-based assistants and code-writing companions to fully multimodal systems capable of handling text, image, video, and audio. Together, they reflect the diversity and rapid evolution of GAI technologies.

DATASETS

The datasets used for training GAI are typically vast and diverse, encompassing a wide range of text, image, and video sources to capture linguistic patterns, visual details, and knowledge across various domains. The datasets used to train unimodal LLM often include text from books, websites, research articles, social media, and other publicly available content, ensuring a rich and varied input. For multimodal LLMs, additional datasets consisting of labeled images, captions, video content, and other visual data are incorporated to train the model, enabling it to generate and understand both text, images, and video. This combination helps the model interpret and produce information across

Table 2 **Comparison of GAI tools.**

Tools	Comp.	In.	Out.	Models	Cost	Datasets	Limitations
ChatGPT (2021)	OpenAI	T,S,C	T,C	GPT-3.5	Free	Web text	Limited reasoning, no multimodal
ChatGPT-4o (2022)	OpenAI	T,I,S,C, D	T,I, F, C	GPT-4, DALL-E 3	subscr.	Large web *corpus*	Weak long-context coherence
Gemini (2024), Bard (2023)	Google DeepMind	T,I,A, V,C, D	T,I	Gemini 1.5 + Imagen 3	Free	Multimodal & multilingual	Not as creative as GPT-4o, limited multimodal
Grok 3 (2025)	xAI	T,I,D	T,I, C	Proprietary transformer, DeepSearch + Think Mode	subscr.	Real-time, legal, diverse data X	Costly, X data bias risk.
DeepSeek (2025)	DeepSeek	T,I,D	T,C	MoE w/DeepThink	Free	Hybrid datasets	Security/reliability issues, poor integration
Google NotebookLM (2025)	Google	T,D	T	Proprietary	Free	Google Drive & external datasets	Limited to docs, lacks reasoning
Claude (2023)	Anthropic	T	T	Claude	Free tier	Public text sources	Unimodal, strict safety limits use
Cohere (2022)	Cohere	T	T	Command Model, Embed	Free tier	Web, books & articles	Smaller, less powerful than GPT
Bardeen.ai (2021)	Bardeen	T	T	GPT-3, GPT-4 Turbo	Free tier	APIs, web & user data	Limited deployment, no full autonomy
Perplexity (2022)	Perplexity AI	T,I,C,D	T,I, C	GPT-4 Turbo, Claude 3, DALL-E 3, SDXL	Free tier	Internet data	Refining multimodal, research-ready
Microsoft Copilot (2025)	Microsoft	T,D	T	GPT-4 + ML	subscr.	Curated M365/ D365/Power data	MS-dependent, limited cross-platform
Copilot (2021)	GitHub + Microsoft	C	C	Microsoft Bing + GPT-3	Free tier	Public GitHub repos	Less flexible, limited explanations
AlphaCode (2022)	DeepMind	T,C	C	Transformer	subscr.	GitHub, coding platforms	Low efficiency, has scalability issues.
GitHub Copilot (2021)	GitHub (Microsoft)	T	C	OpenAI Codex	subscr.	Public code repos	Struggles with multi-step code
DALL-E (2021)	OpenAI	T	I	GPT-3 Architecture	Free	Image-text pairs dataset	Limited resolution and accuracy.
DALL-E 2 (2022)	OpenAI	T	I	CLIP + Diffusion	subscr.	Millions of image-text pairs	Less detailed/realistic than DALL·E 3
DALL-E 3 (2023)	OpenAI	T	I	ChatGPT + Diffusion	subscr.	Web image-text data	Weak with complex visuals
Midjourney (2022)	Midjourney, Inc.	T	I	Proprietary	subscr.	curated datasets	Inconsistent with complex prompts
Crayion (2021)	Boris Dayma	T	I	VQGAN + CLIP	Free	Vast internet images	Lower quality, complex prompt issues
Stable Diffusion V1 (2021), Stable Diffusion V3 (2024)	Stability AI, CompVis & LAION	T	I	LDM + U-Net + CLIP	Free tier	LAION-5B	Ethical risks, prompt issues, resource-heavy

(Continued)

Table 2 (continued)

Tools	Comp.	In.	Out.	Models	Cost	Datasets	Limitations
Descript (2017)	Descript	T,V,A	T,V, A	Lyrebird AI + others	Free tier	Text, audio, video files	Accuracy issues, speech overlap challenges
Type Studio (2021)	Streamlabs	T,V,A	T,V, A	Speech-to-text, NLP models	Free tier	Audio & transcript data	Less capable than descript
Designs AI (2020)	Mediaclip	T,V,I	T,V, I	Proprietary	Free tier	Templates & images data	Less flexible *vs.* design AIs
Soundraw (2021)	Soundraw	Control	M	Proprietary	Free tier	Musical datasets	Less versatile *vs.* Jukebox
Lyria (2023)	Google DeepMind + YouTube	T, M	M	Proprietary	Free	YouTube + other sources	Limited access, Google-only
Duet AI (2023)	Google	T,M	T, M, C	Gemini	subscr.	Text, code & Google data	Less reach *vs.* ChatGPT
Murf AI (2020)	Murf	T,A	A	Proprietary	Free tier	Large speech datasets	Limited Emotional Range
Pika (2024)	Pika Labs	T,I	V	Pika 1.5	Free for 10 vids	Uncited	Video length & motion limits, poor accessibility
Imagen Video (2023)	Google DeepMind	T	V	Proprietary	subscr.	curated video data	Resource-heavy, misuse concerns
Runway GEN (2024)	Runway	T,I	V	Proprietary	Free tier	diverse video data	Resource-heavy, limited free length
Synthesia (2017)	Synthesia AI Research	T,I,V	V	EXPRESS-1 Model	subscr.	Public video/ audio/text data	Style/motion only, no multilingual

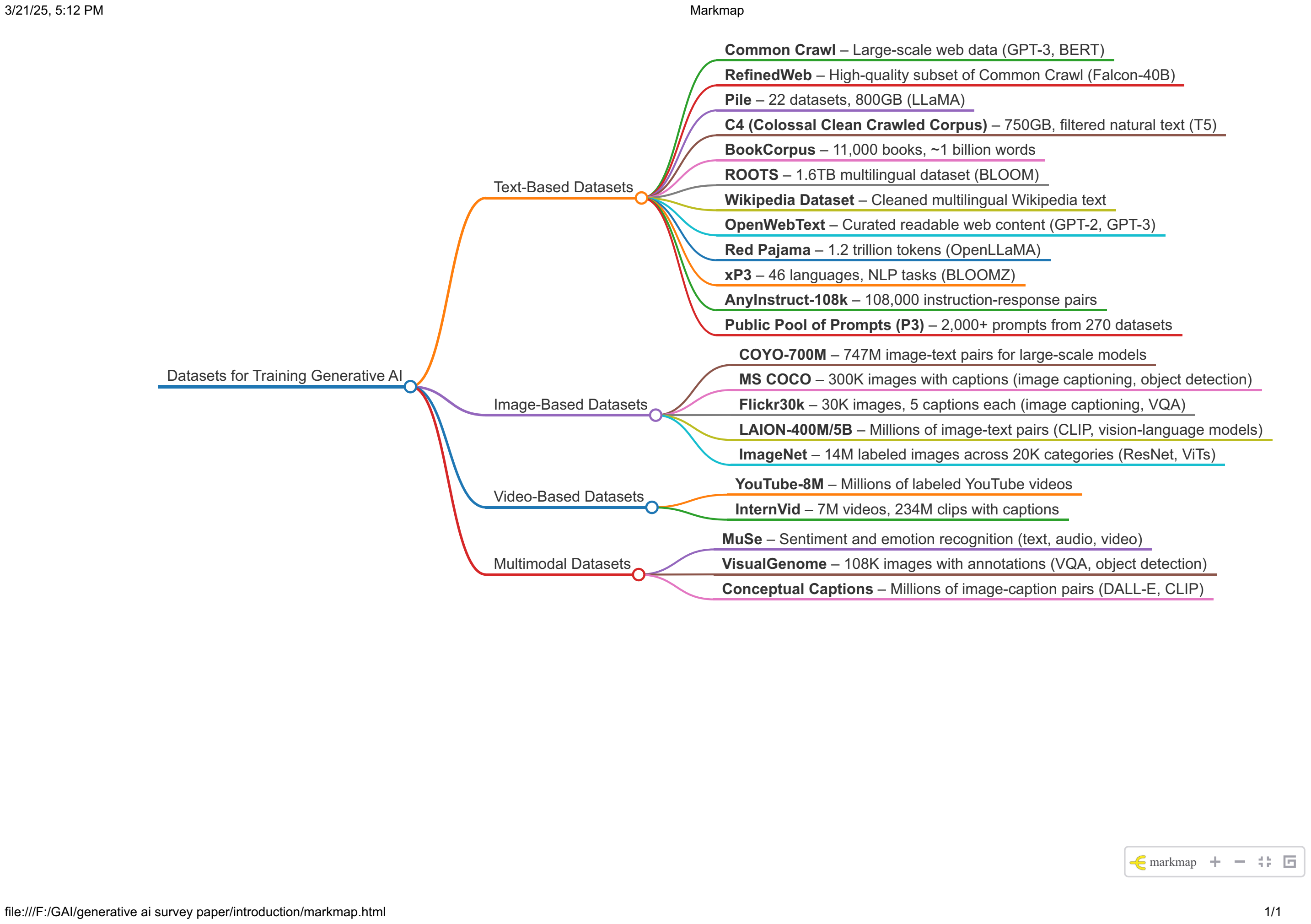
Note:
Text (T), Image (I), Audio (A), Video (V), Code (C), Document (D), File (F), Speech (S), Music (M), Mixture-of-Experts (MoE), Company (Comp.), Input (In.), Output (Out.) Microsoft (MS, Stable Diffusion XL (SDXL), Latent Diffusion Model (LDM).

multiple modalities. The details of some of these datasets are introduced as follows. A graph summarizing these datasets is provided in Fig. 10.

Text-based datasets

Common Crawl dataset provides significant volumes of web data gathered from multiple pages, with frequent monthly updates. It has played a crucial role in training other notable language models, including GPT-3 and BERT. However, it comes with challenges like noise, bias, and ethical concerns around privacy and copyright. Careful filtering and responsible use are essential, but despite these hurdles, it has greatly advanced AI research (*Wenzek et al., 2020*; *Luccioni & Viviano, 2021*; *Baack, 2024*).

RefinedWeb is a high-quality dataset built from Common Crawl, used to train models like Falcon-40B. It combines scale, diversity, and cost-efficiency, making it ideal for pre-training LLM (*Penedo et al., 2023*). While it faces challenges like noise, bias, and ethical concerns, it has pushed GAI forward, powering everything from chatbots to recommendation systems and domain-specific applications.



3/21/25, 5:12 PM

Markmap

Datasets for Training Generative AI
Text-Based Datasets
Common Crawl – Large-scale web data (GPT-3, BERT)
RefinedWeb – High-quality subset of Common Crawl (Falcon-40B)
Pile – 22 datasets, 800GB (LLaMA)
C4 (Colossal Clean Crawled Corpus) – 750GB, filtered natural text (T5)
BookCorpus – 11,000 books, ~1 billion words
ROOTS – 1.6TB multilingual dataset (BLOOM)
Wikipedia Dataset – Cleaned multilingual Wikipedia text
OpenWebText – Curated readable web content (GPT-2, GPT-3)
Red Pajama – 1.2 trillion tokens (OpenLLaMA)
xP3 – 46 languages, NLP tasks (BLOOMZ)
AnyInstruct-108k – 108,000 instruction-response pairs
Public Pool of Prompts (P3) – 2,000+ prompts from 270 datasets
Image-Based Datasets
COYO-700M – 747M image-text pairs for large-scale models
MS COCO – 300K images with captions (image captioning, object detection)
Flickr30k – 30K images, 5 captions each (image captioning, VQA)
LAION-400M/5B – Millions of image-text pairs (CLIP, vision-language models)
ImageNet – 14M labeled images across 20K categories (ResNet, ViTs)
Video-Based Datasets
YouTube-8M – Millions of labeled YouTube videos
InternVid – 7M videos, 234M clips with captions
Multimodal Datasets
MuSe – Sentiment and emotion recognition (text, audio, video)
VisualGenome – 108K images with annotations (VQA, object detection)
Conceptual Captions – Millions of image-caption pairs (DALL-E, CLIP)
markmap

file:///F:/GAI/generative ai survey paper/introduction/markmap.html

1/1

Figure 10 **Datasets for training GAI models.**

Full-size DOI: 10.7717/peerj-cs.3276/fig-10

Pile is a collection of 22 datasets, including a total of 800 GB of data. It is designed to enhance the performance of models such as LLaMA by enhancing their understanding of various settings (*Gao et al., 2020*). Its high-quality, human-generated content helps improve AI reliability, but it also comes with challenges like bias, copyright concerns, and data redundancy. Since it's mostly in English, its multilingual potential is limited, and ethical concerns arise from sensitive forum content and the environmental cost of training (*Biderman, Bicheno & Eleutherai, 2022*).

Colossal Clean Crawled *Corpus* (C4) is a 750 GB dataset from Common Crawl, designed to filter meaningful text for training AI models like Google's T5. It supports tasks like translation, summarization, and Q&A by providing diverse web content (*Dodge et al., 2021*). While it enhances AI generalization, challenges like bias, quality inconsistencies, and ethical concerns from web scraping remain.

BookCorpus dataset is a dataset of about 11,000 unpublished books with nearly a billion words, widely used in AI to train models like BERT and GPT. Its rich linguistic diversity helps with language modeling and creative text generation, but it has drawbacks like genre imbalance, inconsistent quality, and lack of metadata. Ethical concerns include the use of books without author consent, and its English-only content limits multilingual applications (*Bandy & Vincent, 2021*; *Liu et al., 2024a*).

Responsible Open-science Open-collaboration Text Sources (ROOTS) is a 16 TB multilingual dataset, created under the BigScience initiative to train models like BLOOM. Sourced from Common Crawl, GitHub, and more, it supports AI tasks like language modeling, translation, and bias research (*Ostendorff et al., 2024*). ROOTS promote ethical AI through open collaboration, but challenges like data quality, bias, and legal constraints remain (*Piktus et al., 2023*).

Wikipedia dataset is a multilingual collection of cleaned Wikipedia text, used to train AI models like GPT, BERT, and T5 (*Ostendorff et al., 2024*). Its structured, factual content makes it great for language models, knowledge graphs, and Q&A systems. However, it has biases, quality inconsistencies, and gaps in less-represented topics and languages (*Shen, Qi & Baldwin, 2017*).

OpenWebText: **OpenWebText** is an open-source version of OpenAI's WebText, featuring 38 GB of curated web content used to train models like GPT-2 and GPT-3 (*Perełkiewicz & Poświata, 2025*). It enhances AI by providing diverse, readable text but has challenges like quality inconsistencies, biases, and ethical concerns.

The Red Pajama dataset is a 1.2 trillion-token open-source collection from Common Crawl, GitHub, books, and more, designed to replicate LLaMA's training data. It powers models like OpenLLaMA, promoting transparency and open AI research. However, it faces challenges like bias and legal concerns (*Penedo et al., 2023*).

Cross-lingual Public Pool of Prompts (xP3) is a comprehensive repository containing datasets and prompts across 46 languages and encompassing 16 tasks related to NLP. It is an essential resource for training advanced multilingual language models such as BLOOMZ (*Muennighoff et al., 2022*). It supports instruction tuning, bias mitigation, and cross-lingual learning. While offering rich linguistic diversity, it faces challenges like data inconsistencies and bias (*Üstün et al., 2024*).

AnyInstruct-108k The **AnyInstruct-108k dataset** contains 108,000 instruction-response pairs used to train AnyGPT, a multimodal LLM capable of processing text, images, speech, and music. By leveraging discrete representations, AnyGPT integrates these modalities without modifying its core architecture. The dataset enables zero-shot and few-shot learning, but challenges like bias, limited real-world complexity, and ethical concerns around data privacy remain (*Zhan et al., 2024*).

Public Pool of Prompts (P3) is a collection of over 2,000 English prompts from more than 270 datasets, designed for NLP tasks like zero-shot learning. It has trained models like

Flan-T5, enhancing multitask learning across reasoning and instructional tasks. However, P3 faces challenges like bias, quality inconsistencies, and a static nature, limiting long-term adaptability (*Bach et al., 2022*; *Sanh et al., 2021*).

Image-based datasets

The **COYO-700M dataset** contains 747 million image-text pairs with meta-attributes, designed for training multimodal AI models. Sourced from HTML documents, it supports text-to-image generation and has powered models like Vision Transformer (ViT). While promoting accessibility and collaboration, it faces challenges like data noise, bias, and ethical concerns, highlighting the need for careful filtering and curation (*Lu et al., 2023*).

Microsoft Common Objects in Context (MS COCO) contains over 300,000 images with captions. It has been used to train GAI models like Bootstrapped Language-Image Pretraining 2 (BLIP-2) (*Li et al., 2023*), enabling advancements in vision-language tasks such as image captioning and visual question answering. However, it faces challenges like dataset limitations, biases, and environmental concerns (*Lin et al., 2014*).

Flickr30k contains 30,000 images, each with five captions, making it a key resource for image captioning, visual question answering, and multimodal AI. It has helped train models like BLIP-2, advancing cross-modal AI applications (*Liu et al., 2024b*). However, it faces limitations such as its small scale, potential biases, and ethical concerns around privacy and demographic representation (*Plummer et al., 2015*).

Large AI Open Network-400M/5B (LAION-400M/5B) is a massive collection of image-text pairs (400M and 5.85B, respectively) used to train GAI models like CLIP, BLIP, and DALL-E. It powers tasks like text-to-image generation and visual question answering. However, it faces challenges such as noisy data, biases, insufficient curation, privacy concerns, and high computational costs (*Schuhmann et al., 2022*).

ImageNet: contains over 14 million labeled images across 20,000 categories, making it a cornerstone for computer vision and GAI. It has powered models like ResNet, AlexNet, VGG, and Vision Transformers (ViTs) for tasks like image classification and object detection. However, it faces challenges such as biases, labeling errors, static imagery, privacy concerns, and environmental impact (*Krizhevsky, Sutskever & Hinton, 2012*).

Video-based datasets

YouTube-8M is a massive video collection with 8 million labeled videos across 4,800 classes, supporting tasks like video classification and content recommendation. It has helped train models like Deep Bag-of-Frames (DBoF), LSTM, and MoE. However, its reliance on pre-computed features limits detailed analysis, and ethical concerns around privacy and copyright remain challenges (*Abu-El-Haija et al., 2016*).

InternVid includes over seven million videos and 234 million clips with detailed captions, making it a key resource for video-text understanding. It supports tasks like video captioning, retrieval, and generative video synthesis, training models like **ViCLIP** for

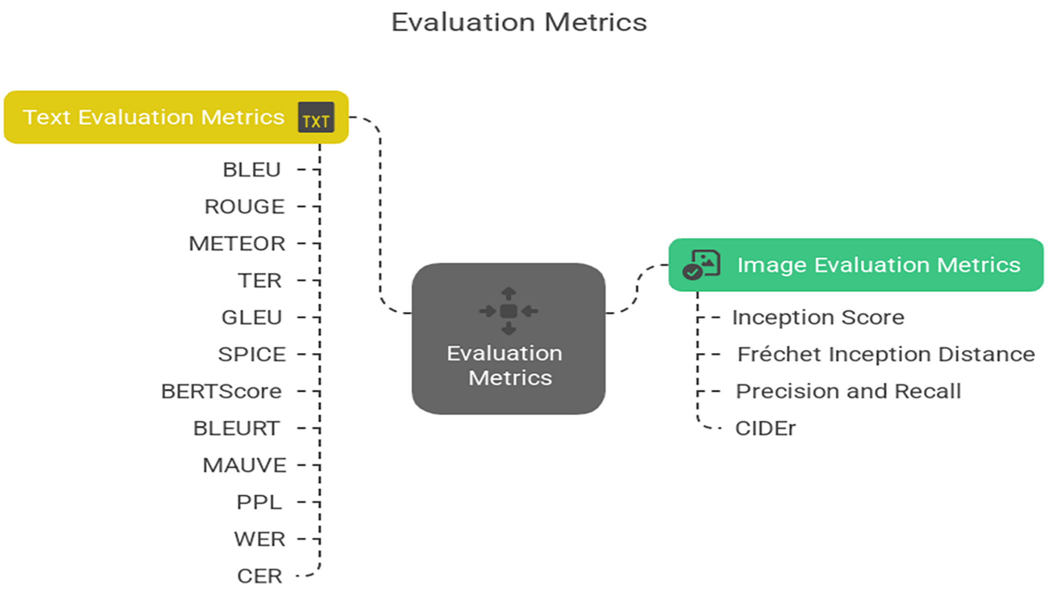


Figure 11 **Evaluation metrics for generative models' performance.**
Full-size DOI: 10.7717/peerj-cs.3276/fig-11

zero-shot action recognition. However, challenges include biases, noisy data, high computational demands, privacy concerns, and costly storage and annotation (*Wang et al., 2023*).

Multimodal datasets

Multimodal Sentiment Understanding and Emotion Recognition (MuSe) combines text, audio, and visual data with 20,000 labeled video segments for sentiment and emotion analysis. It supports tasks like emotion recognition and human-computer interaction but faces challenges like demographic biases, data inconsistencies, and privacy concerns. Models like GRU-RNN and GRUs have leveraged MuSe for tasks such as mimicked emotions and cross-cultural humor detection in events like MuSe 2023 (*Brooks et al., 2023*).

VisualGenome features over 108,000 images with rich textual annotations, enabling tasks like visual question answering, image captioning, and scene graph generation. While it's a key resource for multimodal learning, it has challenges like annotation errors, scalability issues, and biases (*Kim et al., 2024*).

Conceptual Captions: s a massive dataset of image-caption pairs, widely used to train text-to-image models like CLIP and DALL-E. It provides linguistic diversity and real-world relevance but comes with challenges like caption inaccuracies, lack of context, and biases. Ethical concerns, including copyright and privacy issues, also pose challenges (*Ivezić & Babac, 2023*).

EVALUATION METRICS

Evaluating LLMs that generate both text and images requires a diverse set of metrics to assess the quality, diversity, and realism of their outputs across different modalities. This article highlights various evaluation metrics introduced in GAI to measure the performance of LLMs. Figure 11 provides a concise summary of these evaluation metrics, categorizing them for both text and image generation tasks. These metrics are essential for assessing the quality, diversity, and realism of outputs produced by generative models. However, these existing metrics are insufficient for accurately capturing the full scope of LLM capabilities. There remains a need for new and more effective evaluation metrics that can provide a comprehensive assessment of LLMs' quality across both text and image generation.

Image evaluation metrics

The inception score (IS)

IS is a metric used to evaluate the quality of images produced by GANs (*Barratt & Sharma, 2018*). The score quantifies two aspects of the generated samples: the entropy of each individual sample with respect to the class labels, and the entropy of the class distribution across a large number of samples to assess their diversity. In order for a model to be considered well-trained, the entropy of a class for a single sample should be minimized, while the entropy of the class distribution across all generated samples should be maximized (*Raut & Singh, 2024*).

$$IS = \exp\left(E_{x\sim pg} D_{KL}(p(y|x) \parallel p(y))\right),$$

where D_{KL} is the KL divergence between the conditional and marginal distributions, p(y|x) indicates the conditional probability distribution, p(y) is the marginal probability distribution, and $E_{x\sim pg}$ is the sum and average of all results.

The Fréchet inception distance (FID)

FID is a statistic used to measure the level of realism and diversity in images produced by GANs (*Yu, Zhang & Deng 2021*). FID is employed for the analysis of images and is not typically utilized for text, sounds, or other modalities. It measures the influence of modifications in neural network models on realism and compares the advantages of various GAN models in generating images. It effectively evaluates both the visual quality and diversity using a single metric. A lower score indicates a higher resemblance between generated images and genuine images.

Precision and recall

Precision and recall are two metrics used to assess the diversity and quality of the generated samples, specifically addressing the problem of mode dropping. The system generates a two-dimensional score that evaluates the quality of the created images based on precision, which measures the accuracy, and recall, which reflects the extent of coverage by the generative model (*Assefa et al., 2018*; *Kynkäänniemi et al., 2019*).

$$Percision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN},$$

where True Positives (TP) indicate the number of correct positive predictions, False Positives (FP) signify the number of incorrect positive predictions, and False Negatives (FN) represent the number of actual positives that were incorrectly predicted as negatives.

Consensus-based Image Description Evaluation (CIDEr)

The CIDEr metric measures the similarity between a sentence generated by a machine and a collection of sentences authored by humans. This statistic exhibits a high degree of agreement with human judgment on consensus. Through the utilization of sentence similarity, CIDEr automatically encompasses multiple linguistic and evaluative elements, such as grammaticality, saliency, importance, and accuracy (both precision and recall) (*Vedantam, Zitnick & Parikh, 2014*).

$$CIDEr = \frac{1}{|C|}\sum_{i=1}^{|C|}\frac{\sum_{n=1}^{N} w_n.count}{\sum_{n=1}^{N} w_n},$$

where $|C|$ represents the number of reference captions, w_n are weights assigned to diverse n-grams, N is the total number of documents and count indicates the count of n-grams in the created caption that match those in the reference caption.

Text evaluation metrics

Bilingual Evaluation Understudy (BLEU)

BLEU metric is utilized to evaluate the accuracy of machine-generated translations by comparing them to one or more reference translations (*Papineni et al., 2002*). BLEU measures the likeness between translations produced by machines and reference translations by examining the presence of overlapping n-grams, which are consecutive sequences of n words, in both. The BLEU score is a numerical value that falls within the range of 0 to 1. A higher score signifies a stronger similarity between the generated translation and the reference. A score of 1 indicates a flawless alignment, while a score of 0 indicates no overlap between the two translations.

$$BLEU = BP.exp\left(\frac{1}{N}\sum_{n=1}^{N} w_n.\log(p_n)\right),$$

where BP is the Brevity Penalty (to penalize short outputs), and w_n are weights allocated to each precision score.

Recall-Oriented Understudy for Gisting Evaluation

Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a widely employed measure for assessing the excellence of automatically produced summaries from text summarization systems (*Lin, 2004*). ROUGE evaluates the likeness of the created summary and one or more reference summaries by computing precision and recall scores. This is done by comparing n-gram units (such as individual words or word sequences) in the generated and reference summaries. The statistic prioritizes the recall score, which measures the degree to which the essential information from the reference summaries is included in the created summary.

$$ROUGE = \frac{number\ of\ overlapping\ words}{total\ words\ in\ system\ summary}.$$

Metric for Evaluation of Translation with Explicit Ordering

The Metric for Evaluation of Translation with Explicit Ordering (METEOR) is an evaluation metric that is more comprehensive and reliable. It evaluates both precisions, by comparing matching n-grams, and recall, by considering overlapping n-grams (*Lavie & Denkowski, 2009*). Additionally, METEOR considers variations in word order between the language model outputs and the expected results. In addition, METEOR makes use of external language resources, such as WordNet, to include synonyms in the evaluation process. The ultimate score is calculated by taking the harmonic mean of precision and recall and factoring in penalties for inconsistencies in word order.

$$METEOR = F_{mean} \times (1 - P),$$

where P is the penalty based on how well the matched words are sequenced, and F-mean combines precision and recall using a weighted harmonic mean.

Translation Edit Rate

The Translation Edit Rate (TER) is a widely used metric for assessing the precision of machine translations by comparing them to a reference translation. The algorithm calculates the lowest number of editing operations required to align the hypothesis with the reference translation. These operations include shifts, replacements, deletions, and insertions. The TER is computed by dividing the total number of revisions by the average number of words in the reference translation (*Lee et al., 2023*).

$$TER = \frac{E}{N} \times 100,$$

where E signifies the number of edits required to correct the machine-generated translation and N is the total number of words in the reference translation.

Google BLEU

Google BLEU (GLEU) is a specialized metric used to assess the quality of shorter texts. It is often used in activities such as machine translation and language understanding. The functioning of this system is comparable to BLEU, but it is specifically designed to perform better on shorter sentences. It is commonly utilized within Google's internal operations. An advantage of this system is its enhanced ability to detect faults in short messages. However, similar to BLEU, it possesses limitations, particularly in its capacity to comprehensively capture (*Brown, 2024*).

$$GLEU = \min\ (Percision,\ Recall).$$

Semantic Propositional Image Caption Evaluation

Semantic Propositional Image Caption Evaluation (SPICE) is a metric used to assess the quality of image captions based on their semantic content. The assessment process involves comparing the generated pictures against reference pictures using scene graphs that signify the attributes, objects, and relationships within the image (*Anderson et al., 2016*).

$$\mathrm{SPICE} = \frac{2 \times Recall \times Percision}{Recall + Percision}.$$

Bidirectional Encoder Representations from Transformers Score

Bidirectional Encoder Representations from Transformers Score (BERTScore) is a metric that improves the evaluation of text similarity. Built upon the BERT model, it offers a deeper comprehension of textual content, enabling the generation of more insightful and meaningful similarity scores (*Zhang et al., 2019*).

BERT-based Language Understanding Evaluation Reference Tool

BERT-based Language Understanding Evaluation Reference Tool (BLEURT) is a metric that evaluates the quality of generated text by utilizing the advanced comprehension abilities of the BERT model. The objective is to enhance text evaluation by offering assessments that are more sophisticated and sensitive to context (*Lu & Han, 2023*).

Measuring the Gap between Neural Text and Human Text

Measuring the Gap between Neural Text and Human Text (MAUVE) is a metric designed to assess the similarity between neural text produced by LLMs and human writing (*Pillutla et al., 2021*).

Perplexity

Perplexity (PPL) is an assessment criterion for language modeling, is utilized to illustrate the model's capacity to handle extended material (*Hu et al., 2024*).

$$P(w) = \left(\prod_{i=1}^{N} p(w_i)\right)^{-\frac{1}{N}},$$

where W represents the sequence of the words, N indicates the total number of words in the sequence, $p(w_i)$ refers to the predicted probability of the ith word in the sequence.

Word Error Rate

Word Error Rate (WER) assesses the error rate between the reference transcription and the automatic speech recognition-generated transcription (*Matasyoh, Zeineldin & Mathis-Ullrich, 2024*).

$$WER = \frac{NO.\ of\ incorrect\ words}{Total\ no.\ of\ words\ in\ the\ reference\ text} \times 100\%.$$

Character Error Rate

Character Error Rate (CER) assesses the efficacy of text recognition systems, such as Optical Character Recognition (OCR) engines, by calculating the Levenshtein gap between the recognized text and the reference text, then dividing this value by the total number of characters in the reference text to yield an accuracy metric (*Thomas, Gaizauskas & Lu, 2024*).

$$CER = \frac{NO.\ of\ incorrect\ characters}{Total\ no.\ of\ characters\ in\ the\ reference\ text} \times 100\%.$$

GAI APPLICATIONS

GAI models find extensive applications across diverse industries (*Gozalo-Brizuela & Garrido-Merchán, 2023*). These applications encompass:

Healthcare: In the healthcare sector, GAI models may be used in clinical decision support, clinical administration support, synthetic data generation, virtual assistants, query response, medical education, a medical chatbot for customer support, and creating high-resolution images from low-resolution ones (*Javaid, Haleem & Singh, 2023*; *Zhang & Kamel Boulos, 2023*; *Pahune & Rewatkar, 2023*; *Varghese & Chapiro, 2024*). In addition, GAI, including VAE, GAN, flow-based, and diffusion models, offers valuable tools for analyzing large-scale brain imaging data. These methods help researchers understand brain structure and function, especially in reconstructing brain network connectivity (*Gong et al., 2023*; *Wang et al., 2025*). A notable example of this potential in practice is the work of Insilico Medicine. The company has developed Pharma.AI, a GAI platform that played a central role in advancing a cystic fibrosis treatment to human trials in less than 18 months—a remarkable improvement compared to the typical 5–10 year development timeline. Insilico has also developed drug delivery platforms that utilize GAI, Chemistry42 (*Insilico Medicine, 2023*), and PandaOmics (*Pharma.AI*). These platforms are available to other companies for the purpose of enhancing their drug discovery systems. In parallel, the company is working to fully digitize its research and development process, with the goal of reducing errors, accelerating timelines, and cutting costs (*Zhang, Mastouri & Zhang, 2024*).

Education: GAI has the potential to transform the educational system completely by improving several aspects of teaching and learning. It can be used to create content such as quizzes, study material, and worksheets. Moreover, it analyzes the essay content, structure, and grammar to automate the grading process. GAI can also help language learning by producing conversational conversations, offering immediate feedback on pronunciation and grammar, and facilitating interactive language exercises. Besides, it can develop real virtual reality applications with educational objectives that can create simulated worlds about several academic disciplines, such as science, history, and geography (*Wang, 2023*; *Hutson & Cotroneo, 2023*; *Hutson & Lang, 2023*; *Murugesan & Cherukuri, 2023*; *Coleman, 2023*; *Okaiyeto, Bai & Xiao, 2023*; *Bahroun et al., 2023*; *Relmasira, Lai & Donaldson, 2023*;

Jauhiainen & Guerra, 2023). Building on these capabilities, a growing number of GAI tools are now being applied in real-world educational settings, helping to reshape the way students learn and teachers teach. For instance, ChatGPT (OpenAI) is increasingly used for personalized tutoring, writing assistance, and generating formative feedback, offering learners tailored support across disciplines. Similarly, tools such as Google's Bard and Anthropic's Claude provide comparable functionalities, with added emphasis on conversational depth and ethical guidance. In programming education, GitHub Copilot is being adopted to help students write and debug code using natural language prompts, thereby lowering barriers to learning complex technical skills. At the K-12 level, platforms like Socratic (Google) deliver step-by-step explanations that support students in solving homework problems. Meanwhile, Khanmigo, developed by Khan Academy and powered by GPT-4, serves as an interactive tutor, engaging students through Socratic questioning and contextual feedback. Collectively, these tools demonstrate how GAI is supporting more personalized, responsive, and dynamic learning experiences across educational levels and subject areas (*Sandhu et al., 2024*).

Gaming: GAI has potential in game development, namely in creating various game elements like sceneries, levels, maps, environments, textures, character designs, animations, and visual effects. Using GAI, Non-Player Characters (NPC) in video games may be made to behave intelligently and realistically. Furthermore, dialogues, narratives, storylines, sound effects, and music for video games that can reply to player actions may all be created using GAI. Gamers' tastes and behavior may be analyzed using GAI to customize the gaming experience. AI systems can learn from player behavior and modify the level of difficulty, pace, and content to fit the abilities and preferences of each player, making for a more personalized and immersive experience. Besides, processes like game testing and balance may be aided by GAI through simulating and testing games to find any problems, glitches, and imbalances (*Gozalo-Brizuela & Garrido-Merchán, 2023*). Several tools and platforms are already demonstrating how these capabilities are being applied in practice. Ludo AI (*Ludo.ai, 2021*), for instance, helps developers generate game concepts, mechanics, and visual moodboards from simple text prompts. Tools like MarioGPT (*Sudhakaran et al., 2023*) and Puck (*Puck, 2025*) can create level ideas and 3D assets, while Inworld AI enables the development of communicative, AI-driven NPCs. Other platforms, such as Google's Chimaera and modl.ai's virtual testers (*modl.ai, 2023*), showcase how GAI can be used to automate game balancing and playtesting. These innovations not only accelerate development workflows but also open the door to new game genres and creative approaches. Tools like NVIDIA ACE (*NVIDIA Developer, 2023*) for Games further extend the possibilities by enabling voice-driven interactions with in-game characters, allowing for more natural and engaging gameplay (*Werning, 2024*).

Advertising: Chatbots and virtual assistants driven by GAI may be used in advertising to engage with consumers, respond to their questions, and provide tailored suggestions. These chatbots help consumers locate what they are looking for, make purchases, and provide a more tailored shopping experience. Moreover, GAI may help design by producing layouts, images, logos, invitations, digital postcards, graphics, and business cards. Content of all kinds, including text, photos, videos, music, reports, code, HR documents, legal documents, and documentation, may be produced using GAI. Besides, GAI algorithms may develop tailored and targeted adverts by identifying trends and preferences and analyzing large consumer data volumes. This boosts the efficacy of advertising efforts by enabling marketers to target consumers with relevant material (*Mayahi & Vidrih, 2022*; *Hong et al., 2023*; *Israfilzade, 2023*; *Kowalczyk, Röder & Thiesse, 2023*; *Tafesse & Wien, 2024*). A notable example of GAI in advertising is Heinz's use of OpenAI's DALL·E to generate creative visuals and design variations for its ketchup bottles. This campaign highlights how generative tools can enhance brand storytelling and personalization, reflecting a broader shift toward AI-driven, consumer-focused marketing strategies (*Magro-Vela, Sánchez-López & Navarro-Sierra, 2024*).

Fashion and retail: GAI plays a substantial and influential role within the fashion and retail sectors. It can support fashion designers in generating novel and distinctive design concepts. Moreover, it may assist in creating textures, color schemes, and patterns for textiles, giving designers a plethora of alternatives for materials and fabrics. Besides, retailers may design digital showrooms and storefronts where consumers can browse and engage with their designs. Retailers may also save time and ensure consistent and attractive product listings by using GAI to produce marketing material and product descriptions. Furthermore, customers may preview how clothing items will appear on them before purchasing by using GAI to create virtual try-on experiences (*Luce, 2018*; *Raffiee & Sollami, 2020*; *Singh et al., 2020*; *Yan et al., 2023*). To support these innovations, a range of GAI tools is being adopted across design, production, and retail workflows. Platforms like Edited (*EDITED, 2009*), Heuritech (*Heuritech, 2013*), and WGSN (*WGSN*) predict fashion trends by analyzing sales data, social media, and runway activity, enabling brands to align their collections with consumer preferences. Virtual try-on technologies, powered by AI and AR, further enhance the online shopping experience while minimizing return rates. On the operational side, AI supports supply chain optimization by forecasting demand and managing inventory, while AI-driven recommendation engines and automated product labeling improve both user experience and efficiency. Together, these tools are helping the fashion industry become more agile, customer-centric, and sustainable (*Shi, Wong & Zou, 2025*).

Software: The software development business is being revolutionized by GAI, which is effectively decreasing time requirements and enhancing efficiency. GAI enables developers to streamline their workflow by generating code, suggesting code snippets, and automating repetitive processes such as UI creation, completing lines of code, testing, and documentation. This allows developers to allocate their time and energy towards more intricate and demanding jobs. Additionally, generative models can be utilized to create synthetic data for software testing, identify potential vulnerabilities, and even predict bugs through the analysis of code patterns and structure (*Ebert & Louridas, 2023*). A growing number of tools are making these capabilities accessible in practice. Platforms such as

Amazon CodeWhisperer (*Amazon, 2023*), GitHub Copilot, and OpenAI's Codex (*OpenAI, 2021b*) enable developers to generate functions, fix bugs, suggest completions, and translate natural language prompts into executable code across multiple programming languages. These tools not only streamline development workflows but also reduce syntax errors, improve code quality, and accelerate the prototyping process. Moreover, GAI enhances documentation and testing processes, supporting the creation of cleaner, more maintainable codebases (*Huynh & Lin, 2025*).

GAI CHALLENGES, LIMITATIONS, AND ETHICAL IMPLICATIONS

GAI models are advanced forms of AI that can generate novel and realistic content by learning from existing data, including images, text, music, and videos. These models hold great promise for various applications across entertainment, education, marketing, design, and healthcare domains. However, deploying generative models in real-world scenarios presents several challenges that require careful consideration. The following are a few of the main challenges that GAI faces (*Fui-Hoon Nah et al., 2023*; *Koohi-Moghadam & Bae, 2023*):

Data quality and quantity: To create realistic and varied content, generative models need high-quality data to learn from. Unfortunately, gathering and preparing data may be expensive, time-consuming, and biased. Data privacy and security concerns may also restrict data accessibility and availability for generative models.

Model performance and evaluation: Using GAI models for training and inference may pose significant computational and resource burdens. Cutting-edge models often need high-performance hardware and substantial computing resources. The accessibility and scalability of GAI are constrained, especially for individuals and organizations with small computational capabilities. Moreover, assessing the originality and quality of created data is still a challenge. Objective assessment measures may not adequately capture the intricacies and complexity of GAI models. Evaluation is subjective and time-consuming since it often involves human judgment to determine if the product meets coherence, realism, and relevance criteria.

Furthermore, GAI also gives rise to a range of ethical concerns that need to be considered (*Fenwick & Jurcys, 2023*; *Koohi-Moghadam & Bae, 2023*; *Parikh, 2023*; *Zohny, McMillan & King, 2023*).

Intellectual property: GAI can create content similar to or copy existing works, raising intellectual property rights concerns. Therefore, it is essential to determine ownership when AI produces content that could violate patents or copyrights.

Bias: GAI models can inherit biases from their training data, leading to unfair or discriminatory results, such as biased language, stereotypes, or skewed recommendations. To mitigate this, techniques like careful dataset curation, adversarial training, and fairness-aware algorithms can help create more balanced and inclusive AI-generated content.

Misinformation or manipulation: GAI can create highly realistic text, images, and videos, increasing the risk of misinformation and public manipulation. This raises concerns about the authenticity of AI-generated content and the need for safeguards to detect and prevent misuse. Implementing solutions like AI detection algorithms, watermarking, and traceability can help ensure transparency and reduce the spread of false information.

Data privacy violations: GAI relies on large amounts of data, including personal information, raising privacy concerns. Setting clear standards and obtaining user consent is essential for responsible data use. Techniques like anonymization, secure training methods, and protective noise help safeguard sensitive information. Moreover, following privacy laws ensures ethical AI development.

Human creativity and employment: There are worries that as GAI develops, human creativity may be replaced or diminished in several contexts. This may have an impact on jobs and people's lives who depend on creative activity. An essential ethical problem is striking a balance between the advantages of GAI and the preservation of human creativity and job security.

Transparency: Organizations should aim for transparency and create documentation on how the model works, its limitations and risks, and the data used to train it. "Black-box" models are challenging to interpret by nature; many of these LLMs have billions of parameters that make them uninterpretable.

Accountability: If content created by GAI is inaccurate or harmful, who is responsible? Who has the legal liability? Users and creators of GAI models should be held responsible for the models' outputs through explicit processes in place. This might be through creating new legislation and regulations.

Environmental impact: The environmental effect of the computing resources needed to train and implement GAI models at scale may be substantial. For instance, massive energy consumption occurs during the large-scale training of AI models, which increases carbon emissions and exacerbates the issue of climate change. This topic urges the AI industry to adopt sustainable practices and create more energy-efficient algorithms to lessen the environmental impact of GAI. To tackle these ethical concerns effectively, it is imperative for AI engineers, legislators, ethicists, and the general public to collaborate and adopt a multidisciplinary approach. This approach aims to safeguard social values and minimize potential harm by establishing ethical frameworks, regulations, and legislation that promote responsible and accountable usage of GAI.

CONCLUSION

A remarkable transformation has occurred in the domain of GAI, reshaping the direction of AI research and applications. This review article extensively analyzes GAI, including its historical development, various models, essential tools, datasets, evaluation metrics, challenges, and significant ethical implications. The review article begins by outlining the chronological progression of GAI development, highlighting significant milestones from its early stages to current improvements of the models and tools. The review examines foundational generative models like GANs and VAEs, their progression toward more advanced architectures such as Transformers, and offers a comparative analysis of their strengths and weaknesses. Additionally, it explores the capabilities of these models in generating images, text, and music, along with their impact on industries like healthcare, education, and entertainment. Lastly, the article discusses critical challenges, limitations, and ethical concerns, providing insights into key issues that researchers must address to ensure responsible AI development.

The future of GAI is filled with exciting possibilities, but also significant challenges that must be addressed to ensure responsible progress. One of the most transformative advancements will be the rise of multimodal AI, seamlessly integrating text, images, audio, and video for more natural and intuitive interactions. AI will also become more hyper-personalized, tailoring content and recommendations to individual preferences, making digital experiences more engaging and meaningful. As AI becomes more efficient, it will require fewer computational resources, making powerful tools accessible to businesses, developers, and individuals. However, ethical concerns must be prioritized. Bias and fairness remain critical issues, as AI models often reflect biases from their training data, particularly in sensitive areas like healthcare and education. Stronger techniques are needed to reduce bias, especially in diverse and multilingual contexts. Privacy and security risks also pose challenges, with AI tools potentially exposing sensitive data, making privacy-preserving architectures and stricter regulations essential. AI's ability to generate hallucinations raises concerns, particularly in fields like medicine where accuracy is crucial. Improving fact-checking systems and domain-specific accuracy metrics is necessary to ensure reliability. Additionally, adversarial attacks can bypass AI safety measures, requiring stronger security defenses. Regulatory frameworks must evolve to keep pace with rapidly advancing AI models, ensuring adaptive policies that balance innovation with accountability. Beyond technical risks, overreliance on AI may diminish human creativity and decision-making. Collaboration between AI developers, ethicists, and industry experts is vital to designing systems that enhance, rather than replace, human input. Sustainability is another pressing challenge, as training large AI models consumes vast amounts of energy. Developing energy-efficient architectures and measuring AI's carbon footprint will be key to minimizing environmental impact. By addressing these challenges GAI can evolve responsibly, unlocking its full potential while ensuring ethical and societal well-being.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was funded by the Korea Agency for Technology and Standards in 2023, project numbers 1415181629 (20022340, Development of International Standard Technologies based on AI Model Lightweighting Technologies) and 1415180835 (20020734, Development of International Standard Technologies based on AI Learning and Inference Technologies). The APC was funded by the Ministry of Oceans and Fisheries (RS-2022-KS221571) The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Korea Agency for Technology and Standards: 1415181629 and 1415180835.
Ministry of Oceans and Fisheries: RS-2022-KS221571.

Competing Interests

Author Mohamed S. Abdallah was employed by the company DeltaX Co., Ltd., South Korea. The remaining authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Author Contributions

- Rasha Shoitan conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Mona M. Moussa conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Nahed Tawfik conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Young-Im Cho performed the computation work, prepared figures and/or tables, and approved the final draft.
- Mohamed S. Abdallah conceived and designed the experiments, performed the experiments, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data was not generated in this Literature Review.

REFERENCES

Abdin M, Aneja J, Awadalla H, Awadallah A, Awan AA, Bach N, Bahree A, Bakhtiari A, Bao J, Behl H, Benhaim A, Bilenko M, Bjorck J, Bubeck S, Cai M, Cai Q, Chaudhary V, Chen D, Chen D, Chen W, Chen Y-C, Chen Y-L, Cheng H, Chopra P, Dai X, Dixon M, Eldan R, Fragoso V, Gao J, Gao M, Gao M, Garg A, Del Giorno A, Goswami A, Gunasekar S, Haider E, Hao J, Hewett RJ, Hu W, Huynh J, Iter D, Jacobs SA, Javaheripi M, Jin X, Karampatziakis N, Kauffmann P, Khademi M, Kim D, Kim YJ, Kurilenko L, Lee JR, Lee YT, Li Y, Li Y, Liang

C, Liden L, Lin X, Lin Z, Liu C, Liu L, Liu M, Liu W, Liu X, Luo C, Madan P, Mahmoudzadeh A, Majercak D, Mazzola M, Mendes CCT, Mitra A, Modi H, Nguyen A, Norick B, Patra B, Perez-Becker D, Portet T, Pryzant R, Qin H, Radmilac M, Ren L, de Rosa G, Rosset C, Roy S, Ruwase O, Saarikivi O, Saied A, Salim A, Santacroce M, Shah S, Shang N, Sharma H, Shen Y, Shukla S, Song X, Tanaka M, Tupini A, Vaddamanu P, Wang C, Wang G, Wang L, Wang S, Wang X, Wang Y, Ward R, Wen W, Witte P, Wu H, Wu X, Wyatt M, Xiao B, Xu C, Xu J, Xu W, Xue J, Yadav S, Yang F, Yang J, Yang Y, Yang Z, Yu D, Yuan L, Zhang C, Zhang C, Zhang J, Zhang LL, Zhang Y, Zhang Y, Zhang Y, Zhou X. 2024. Phi-3 technical report: a highly capable language model locally on your phone. ArXiv DOI 10.48550/arxiv.2404.14219.

Abu-El-Haija S, Kothari N, Lee J, Natsev P, Toderici G, Varadarajan B, Vijayanarasimhan S. 2016. YouTube-8M: a large-scale video classification benchmark. ArXiv DOI 10.48550/arXiv.1609.08675.

Ackley DH, Hinton GE, Sejnowski J. 1985. A learning algorithm for Boltzmann machines. *Cognitive Science* **9**:147–169 DOI 10.1016/S0364-0213(85)80012-4.

Alexa. 2025. Amazon Alexa voice AI. *Available at https://developer.amazon.com/en-US/alexa* (accessed 19 November 2023).

Amazon. 2023. Amazon CodeWhisperer. *Available at https://aws.amazon.com/ar/codewhisperer/* (accessed 17 July 2025).

Anderson P, Fernando B, Johnson M, Gould S. 2016. SPICE: Semantic propositional image caption evaluation. In: Leibe B, Matas J, Sebe N, Welling M, eds. *Computer Vision–ECCV 2016. ECCV 2016. Lecture Notes in Computer Science.* Vol. 9909. Cham: Springer, 382–398 DOI 10.1007/978-3-319-46454-1_24.

Anil R, Dai AM, Firat O, Johnson M, Lepikhin D, Passos A, Shakeri S, Taropa E, Bailey P, Chen Z, Chu E, Clark JH, El SL, Huang Y, Meier-Hellstern K, Mishra G, Moreira E, Omernick M, Robinson K, Ruder S, Tay Y, Xiao K, Xu Y, Zhang Y, Abrego GH, Ahn J, Austin J, Barham P, Botha J, Bradbury J, Brahma S, Brooks K, Catasta M, Cheng Y, Cherry C, Choquette-Choo CA, Chowdhery A, Crepy C, Dave S, Dehghani M, Dev S, Devlin J, Díaz M, Du N, Dyer E, Feinberg V, Feng F, Fienber V, Freitag M, Garcia X, Gehrmann S, Gonzalez L, Gur-Ari G, Hand S, Hashemi H, Hou L, Howland J, Hu A, Hui J, Hurwitz J, Isard M, Ittycheriah A, Jagielski M, Jia W, Kenealy K, Krikun M, Kudugunta S, Lan C, Lee K, Lee B, Li E, Li M, Li W, Li Y, Li J, Lim H, Lin H, Liu Z, Liu F, Maggioni M, Mahendru A, Maynez J, Misra V, Moussalem M, Nado Z, Nham J, Ni E, Nystrom A, Parrish A, Pellat M, Polacek M, Polozov A, Pope R, Qiao S, Reif E, Richter B, Riley P, Ros AC, Roy A, Saeta B, Samuel R, Shelby R, Slone A, Smilkov D, So DR, Sohn D, Tokumine S, Valter D, Vasudevan V, Vodrahalli K, Wang X, Wang P, Wang Z, Wang T, Wieting J, Wu Y, Xu K, Xu Y, Xue L, Yin P, Yu J, Zhang Q, Zheng S, Zheng C, Zhou W, Zhou D, Petrov S, Wu Y. 2023. PaLM 2 technical report. ArXiv DOI 10.48550/arxiv.2305.10403.

Anthropic. 2023a. Claude 2. *Available at https://www.anthropic.com/index/claude-2* (accessed 5 December 2023).

Anthropic. 2023b. Claude. *Available at https://claude.ai/onboarding* (accessed 25 November 2023).

Apple. 2023. Siri. *Available at https://www.apple.com/siri/* (accessed 19 November 2023).

Arora S, Risteski A, Zhang Y. 2017. Theoretical limitations of Encoder-Decoder GAN architectures. ArXiv DOI 10.48550/arXiv.1711.02651.

Artbreeder. 2025. Artbreeder. *Available at https://www.artbreeder.com/* (accessed 8 January 2024).

Assefa SA, Dervovic D, Mahfouz M, Tillman RE, Reddy P, Veloso M. 2018. Assessing generative models via precision and recall. In: *Proceedings of the First ACM International Conference on AI in Finance*. New York, NY, USA: ACM, 5228–5237 DOI 10.1145/3383455.3422554.

Baack S. 2024. A critical analysis of the largest source for generative AI training data: Common crawl. In: *024 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2024*, 2199–2208 DOI 10.1145/3630106.3659033.

Bach SH, Sanh V, Yong ZX, Webson A, Raffel C, Nayak NV, Sharma A, Kim T, Bari MS, Fevry T, Alyafeai Z, Dey M, Santilli A, Sun Z, Ben-David S, Xu C, Chhablani G, Wang H, Fries JA, Al-Shaibani MS, Sharma S, Thakker U, Almubarak K, Tang X, Radev D, Jiang MTJ, Rush AM. 2022. PromptSource: an integrated development environment and repository for natural language prompts. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 93–104 DOI 10.18653/v1/2022.acl-demo.9.

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* **15(17)**:12983 DOI 10.3390/SU151712983.

Baidu Research. 2023. Introducing ERNIE 3.5: Baidu's knowledge-enhanced foundation model takes a giant leap forward. *Available at https://research.baidu.com/Blog/index-view?id=185*.

Bălan C. 2023. Chatbots and voice assistants: digital transformers of the company-customer interface—a systematic review of the business research literature. *Journal of Theoretical and Applied Electronic Commerce Research* **18(2)**:995–1019 DOI 10.3390/jtaer18020051.

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* **15(8)**:260 DOI 10.3390/fi15080260.

Bandy J, Vincent N. 2021. Addressing documentation debt in machine learning research: a retrospective datasheet for bookcorpus. ArXiv DOI 10.48550/arXiv.2105.05241.

Bard. 2023. Google Bard. *Available at https://bard.google.com/chat/5e216179b647b0dd* (accessed 26 November 2023).

Bardeen AI. 2025. GTM copilot for workflow automation. *Available at https://www.bardeen.ai/* (accessed 27 March 2025).

Barratt S, Sharma R. 2018. A note on the inception score. ArXiv DOI 10.48550/arxiv.1801.01973.

Bayma B, Cuenca P. 2021. DALL·E mini – generate images from any text prompt|dalle-mini. Weights & Biases. *Available at https://wandb.ai/dalle-mini/dalle-mini/reports/DALL-E-mini-Generate-images-from-any-text-prompt–VmlldzoyMDE4NDAy* (accessed 24 November 2023).

Bengesi S, El-Sayed H, Sarker MK, Houkpati Y, Irungu J, Oladunni T. 2024. Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEE Access* **12(3)**:69812–69837 DOI 10.1109/ACCESS.2024.3397775.

Biderman S, Bicheno K, Eleutherai LG. 2022. Datasheet for the pile. ArXiv DOI 10.48550/arXiv.2201.07311.

Brooks JA, Tiruvadi V, Baird A, Tzirakis P, Li H, Gagne C, Oh M, Cowen A. 2023. Emotion Expression estimates to measure and improve multimodal social-affective interactions. In: *ACM International Conference Proceeding Series*, 353–358 DOI 10.1145/3610661.3616129.

Brown NB. 2024. Enhancing trust in LLMs: Algorithms for comparing and interpreting LLMs. ArXiv DOI 10.48550/arXiv.2406.01943.

Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems* **33**:1877–1901 DOI 10.5555/3495724.3495883.

Card M. 2023. Model card and evaluations for claude models. *Anthropic* 1–14.

Castelli M, Manzoni L. 2022. Generative models in artificial intelligence and their applications. *Applied Sciences* **12(9)**:10–12 DOI 10.3390/app12094127.

Chang Z, Koulieris GA, Shum HPH. 2023. On the design fundamentals of diffusion models: a survey. ArXiv DOI 10.48550/arXiv.2306.04542.

Chen M, Mei S, Fan J, Wang M. 2024. Opportunities and challenges of diffusion models for generative AI. *National Science Review* **11(12)**:186 DOI 10.1093/NSR/NWAE348.

Chen M, Tworek J, Jun H, Yuan Q, Pinto HPdeO, Kaplan J, Edwards H, Burda Y, Joseph N, Brockman G, Ray A, Puri R, Krueger G, Petrov M, Khlaaf H, Sastry G, Mishkin P, Chan B, Gray S, Ryder N, Pavlov M, Power A, Kaiser L, Bavarian M, Winter C, Tillet P, Such FP, Cummings D, Plappert M, Chantzis F, Barnes E, Herbert-Voss A, Guss WH, Nichol A, Paino A, Tezak N, Tang J, Babuschkin I, Balaji S, Jain S, Saunders W, Hesse C, Carr AN, Leike J, Achiam J, Misra V, Morikawa E, Radford A, Knight M, Brundage M, Murati M, Mayer K, Welinder P, McGrew B, Amodei D, McCandlish S, Sutskever I, Zaremba W. 2021. Evaluating large language models trained on code. ArXiv DOI 10.48550/arxiv.2107.03374.

Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, Barham P, Chung HW, Sutton C, Gehrmann S, Schuh P, Shi K, Tsvyashchenko S, Maynez J, Rao A, Barnes P, Tay Y, Shazeer N, Prabhakaran V, Reif E, Du N, Hutchinson B, Pope R, Bradbury J, Austin J, Isard M, Gur-Ari G, Yin P, Duke T, Levskaya A, Ghemawat S, Dev S, Michalewski H, Garcia X, Misra V, Robinson K, Fedus L, Zhou D, Ippolito D, Luan D, Lim H, Zoph B, Spiridonov A, Sepassi R, Dohan D, Agrawal S, Omernick M, Dai AM, Pillai TS, Pellat M, Lewkowycz A, Moreira E, Child R, Polozov O, Lee K, Zhou Z, Wang X, Saeta B, Diaz M, Firat O, Catasta M, Wei J, Meier-Hellstern K, Eck D, Dean J, Petrov S, Fiedel N. 2022. PaLM: scaling language modeling with pathways. ArXiv DOI 10.48550/arxiv.2204.02311.

Christiano PF, Leike J, Brown TB, Martic M, Legg S, Amodei D. 2017. Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems* **30**:4299–4307.

Cohere. 2022. Cohere Labs. *Available at https://cohere.com/research* (accessed 20 October 2024).

Coleman K. 2023. Generative AI and education ecologies. *Pacific Journal of Technology Enhanced Learning* **5(1)**:19–20 DOI 10.24135/PJTEL.V5I1.175.

Craiyon. 2022. Your FREE AI image generator tool: create AI art!. *Available at https://www.craiyon.com/* (accessed 19 November 2023).

Dasgupta D, Venugopal D, Gupta KD. 2023. A review of generative AI from historical perspectives. *TechRxiv* DOI 10.36227/techrxiv.22097942.v1.

Daunhawer I, Sutter TM, Chin-Cheong K, Palumbo E, Vogt JE. 2021. On the limitations of multimodal VAEs. ArXiv DOI 10.48550/arXiv.2110.04121.

DeepArt.io. 2015. DeepArt.io: your FREE AI image generator tool. *Available at https://deepart.io/* (accessed 17 July 2025).

DeepMind. 2022. AlphaCode. *Available at https://alphacode.deepmind.com/* (accessed 27 March 2025).

DeepSeek. 2023. DeepSeek. *Available at https://www.deepseek.com/* (accessed 27 March 2025).

Descript. 2017. Edit videos & podcasts like a doc|AI video editor. *Available at https://www.descript.com/* (accessed 27 March 2025).

Designs AI. 2019. Free online logo, image, AI chat videos & voice generator. *Available at https://designs.ai/* (accessed 27 March 2025).

Devlin J, Chang MW, Lee K, Toutanova K. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2019). Vol. 1, 4171–4186.

DiPietro R, Hager GD. 2019. Deep learning: RNNs and LSTM. In: *Handbook of Medical Image Computing and Computer Assisted Intervention,* 503–519 DOI 10.1016/B978-0-12-816176-0.00026-0.

Dodge J, Sap M, Marasovic A, Agnew W, Ilharco G, Groeneveld D, Mitchell M, Gardner M. 2021. Documenting large webtext corpora: a case study on the colossal clean crawled corpus. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing.* Stroudsburg, PA: Association for Computational Linguistics, 1286–1305.

DreamStudio. 2022. DreamStudio. *Available at https://beta.dreamstudio.ai/generate* (accessed 5 December 2023).

Ebert C, Louridas P. 2023. Generative AI for software practitioners. *IEEE Software* **40(4)**:30–38 DOI 10.1109/MS.2023.3265877.

EDITED. 2009. Empowering intelligent retail. *Available at https://edited.com/* (accessed 17 July 2025).

Falcon LLM & Technology Innovation Institute (TII). 2023. Falcon-H1. *Available at https://falconllm.tii.ae/* (accessed 25 March 2024).

Fan A, Lavril T, Grave E, Joulin A, Sukhbaatar S. 2020. Addressing some limitations of transformers with feedback memory. ArXiv DOI 10.48550/arxiv.2002.09402.

Fenwick M, Jurcys P. 2023. Originality and the future of copyright in an age of generative AI. *Computer Law and Security Review* **51**:105892 DOI 10.1016/j.clsr.2023.105892.

Feuerriegel S, Hartmann J, Janiesch C, Zschech P. 2023. Generative AI. *Business and Information Systems Engineering* **66(1)**:111–126 DOI 10.1007/s12599-023-00834-7.

Fuest M, Ma P, Gui M, Schusterbauer J, Hu VT, Ommer B. 2024. Diffusion models and representation learning: a survey. ArXiv DOI 10.48550/arXiv.2407.00783.

Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. 2023. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research* **25(3)**:277–304 DOI 10.1080/15228053.2023.2233814.

Gad AF, Jarmouni FE. 1995. Introduction to artificial neural networks (ANN). In: *Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics*. Piscataway: IEEE, 33–37 DOI 10.1016/b978-0-323-90933-4.00007-3.

Gao L, Biderman S, Black S, Golding L, Hoppe T, Foster C, Phang J, He H, Thite A, Nabeshima N, Presser S, Leahy C. 2020. The pile: An 800GB dataset of diverse text for language modeling. ArXiv DOI 10.48550/arXiv.2101.00027.

Gao J, Shen T, Wang Z, Chen W, Yin K, Li D, Litany O, Gojcic Z, Fidler S. 2022. GET3D: a generative model of high quality 3D textured shapes learned from images. *Advances in Neural Information Processing Systems* **35**:1–39.

Gatys LA, Ecker AS, Bethge M. 2015. A neural algorithm of artistic style. *Journal of Vision* **16(12)**:326 DOI 10.1167/16.12.326.

GitHub Copilot. 2021. Your AI pair programmer. *Available at https://github.com/features/copilot* (accessed 19 November 2023).

Gong C, Jing C, Chen X, Pun CM, Huang G, Saha A, Nieuwoudt M, Li HX, Hu Y, Wang S. 2023. Generative AI for brain image computing and brain network computing: a review. *Frontiers in Neuroscience* **17**:663 DOI 10.3389/fnins.2023.1203104.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. 2014. Generative adversarial networks. In: *NIPS'14: Proceedings of the 27th International*

Conference on Neural Information Processing Systems. New York: ACM, 2672–2680 DOI 10.1145/3422622.

Google. 2023. Gemini. *Available at https://gemini.google.com/app* (accessed 27 March 2025).

Google Assistant. 2016. Your own personal Google. *Available at https://assistant.google.com/* (accessed 19 November 2023).

Google AI. 2023. Introducing PaLM 2. *Available at https://blog.google/technology/ai/google-palm-2-ai-large-language-model/* (accessed 20 October 2024).

Google. 2022. Imagen Video. *Available at https://imagen.research.google/video/* (accessed 27 March 2025).

Google AI for Developers. 2023. Gemma: a family of lightweight, state-of-the art open models from Google. *Available at https://ai.google.dev/gemma* (accessed 17 March 2024).

Google. 2023. NotebookLM. *Available at https://notebooklm.google.com/* (accessed 27 March 2025).

Google DeepMind. 2024a. Imagen 3. *Available at https://deepmind.google/technologies/imagen-3/*.

Google DeepMind. 2024b. Lyria. *Available at https://deepmind.google/models/lyria/* (accessed 17 July 2025).

Google DeepMind. 2023. Transforming the future of music creation. *Available at https://deepmind.google/discover/blog/transforming-the-future-of-music-creation/* (accessed 19 November 2023).

Gozalo-Brizuela R, Garrido-Merchán EC. 2023. A survey of generative AI applications. ArXiv DOI 10.48550/arxiv.2306.02781.

Heuritech. 2013. Fashion trend forecasting & prediction with AI. *Available at https://heuritech.com/* (accessed 17 July 2025).

Ho J, Chan W, Saharia C, Whang J, Gao R, Gritsenko A, Kingma DP, Poole B, Norouzi M, Fleet DJ, Salimans T. 2022. Imagen video: high definition video generation with diffusion models. ArXiv DOI 10.48550/arXiv.2210.02303.

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* **9(8)**:1735–1780 DOI 10.1162/neco.1997.9.8.1735.

Hong W, Ding M, Zheng W, Liu X, Tang J. 2022. CogVideo: large-scale pretraining for text-to-video generation via transformers. ArXiv DOI 10.48550/arxiv.2205.15868.

Hong MK, Hakimi S, Chen Y-Y, Toyoda H, Wu C, Klenk M. 2023. Generative AI for product design: Getting the right design and the design right. ArXiv DOI 10.48550/arxiv.2306.01217.

Hu Y, Huang Q, Tao M, Zhang C, Feng Y. 2024. Can perplexity reflect large language model's ability in long text understanding? ArXiv DOI 10.48550/arXiv.2405.06105.

Hutson J, Cotroneo P. 2023. Generative AI tools in art education: exploring prompt engineering and iterative processes for enhanced creativity. *Metaverse* **4(1)**:14 DOI 10.54517/m.v4i1.2164.

Hutson J, Lang M. 2023. Content creation or interpolation: AI generative digital art in the classroom. *Metaverse* **4(1)**:1–15 DOI 10.54517/m.v4i1.2158.

Huynh N, Lin B. 2025. Large language models for code generation: a comprehensive survey of challenges, techniques, evaluation, and applications. ArXiv DOI 10.48550/arXiv.2503.01245.

IBM Newsroom. 2023. IBM unveils the Watsonx platform to power next-generation foundation models for business. *Available at https://newsroom.ibm.com/2023-05-09-IBM-Unveils-the-Watsonx-Platform-to-Power-Next-Generation-Foundation-Models-for-Business* (accessed 1 April 2024).

IBM Research. 2023. Building AI for business: IBM's Granite foundation models. *Available at https://www.ibm.com/think/news/granite-foundation-models* (accessed 26 March 2025).

Iglesias G, Talavera E, Díaz-Álvarez A. 2023. A survey on GANs for computer vision: recent research, analysis and taxonomy. *Computer Science Review* **48(12)**:100553 DOI 10.1016/j.cosrev.2023.100553.

Insilico Medicine. 2023. Chemistry42. *Available at https://insilico.com/chemistry42_fr* (accessed 17 July 2025).

Isola P, Zhu JY, Zhou T, Efros AA. 2017. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. Piscataway: IEEE, 5967–5976 DOI 10.1109/CVPR.2017.632.

Israfilzade K. 2023. Beyond automation: the impact of anthropomorphic generative AI on conversational marketing. In: *International European Conference On Interdisciplinary Scientific Research-VIII*, Rome, Italy, 13–15 DOI 10.5281/zenodo.8253308.

Ivezić D, Babac MB. 2023. Trends and challenges of text-to-image generation: sustainability perspective. *Croatian Regional Development Journal* **4(1)**:56–77 DOI 10.2478/crdj-2023-0004.

Jauhiainen JS, Guerra AG. 2023. Generative AI and ChatGPT in school children's education: evidence from a school lesson. *Sustainability* **15(18)**:14025 DOI 10.3390/SU151814025.

Javaid M, Haleem A, Singh RP. 2023. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* **3(1)**:100105 DOI 10.1016/J.TBENCH.2023.100105.

Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, de las CD, Bressand F, Lengyel G, Lample G, Saulnier L, Lavaud LR, Lachaux M-A, Stock P, Le ST, Lavril T, Wang T, Lacroix T, Sayed WE. 2023. Mistral 7B. ArXiv DOI 10.48550/arxiv.2310.06825.

Karras T, Laine S, Aila T. 2021. A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43(12)**:4217–4228 DOI 10.1109/TPAMI.2020.2970919.

Kim K, Yoon K, Jeon J, In Y, Moon J, Kim D, Park C. 2024. LLM4SGG: large language models for weakly supervised scene graph generation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 28306–28316 DOI 10.1109/CVPR52733.2024.02674.

Kingma DP, Welling M. 2014. Auto-encoding variational bayes. In: *In Proceedings of the International Conference on Learning Representations (ICLR)*. Banff, AB, Canada.

Kingma DP, Welling M. 2019. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning* **12**:307–392 DOI 10.48550/arXiv.1906.02691.

Koohi-Moghadam M, Bae KT. 2023. Generative AI in medical imaging: applications, challenges, and ethics. *Journal of Medical Systems* **47(1)**:1–4 DOI 10.1007/s10916-023-01987-4.

Kowalczyk P, Röder M, Thiesse F. 2023. Nudging creativity in digital marketing with generative artificial intelligence: opportunities and limitations. In: *European Conference on Information Systems*, Kristiansand, Norway, 1–9.

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. In: *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097–1105.

Kynkäänniemi T, Karras T, Laine S, Lehtinen J, Aila T. 2019. Improved precision and recall metric for assessing generative models. In: *Advances in neural information processing systems.* Neural Information Processing Systems Foundation, 32.

Lavie A, Denkowski MJ. 2009. The METEOR metric for automatic evaluation of machine translation. *Machine Translation* **23**:105–115 DOI 10.1007/S10590-009-9059-4.

Lee S, Lee J, Moon H, Park C, Seo J, Eo S, Koo S, Lim H. 2023. A survey on evaluation metrics for machine translation. *Mathematics* **11**:1006 DOI 10.3390/MATH11041006.

Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih WT, Rocktäschel T, Riedel S, Kiela D. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems* **33**:9459–9474.

Li J, Li D, Savarese S, Hoi S. 2023. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. *Proceedings of Machine Learning Research* **202**:20351–20383.

Lin C-Y. 2004. ROUGE: a package for automatic evaluation of summaries. In: *Workshop on Text Summarization Branches Out, Association for Computational Linguistics (ACL)*, 74–81.

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. 2014. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. *Computer Vision–ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*. Vol. 8693. Cham: Springer, 740–755 DOI 10.1007/978-3-319-10602-1_48.

Link D. 2006. Traces of the mouth: Andrei Andreyevich Markov's mathematization of writing. *History of Science* **44(3)**:321–348 DOI 10.1177/007327530604400302.

Liu Y, Cao J, Liu C, Ding K, Jin L. 2024a. Datasets for large language models: a comprehensive survey. ArXiv DOI 10.48550/arXiv.2402.18041.

Liu H, Song Y, Wang X, Xiangru Z, Li Z, Song W, Li T. 2024b. Flickr30K-CFQ: a compact and fragmented query dataset for text-image retrieval. ArXiv DOI 10.1007/978-981-97-5555-4_30.

Lu X, Han C. 2023. Automatic assessment of spoken-language interpreting based on machine-translation evaluation metrics: a multi-scenario exploratory study. *Interpreting* **25(1)**:109–143 DOI 10.1075/INTP.00076.LU.

Lu C-Z, Jin X, Hou Q, Liew JH, Cheng M-M, Feng J. 2023. Delving deeper into data scaling in masked image modeling. ArXiv DOI 10.48550/arXiv.2305.15248.

Luccioni A, Viviano J. 2021. What's in the box? An analysis of undesirable content in the common crawl corpus. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 182–189 DOI 10.18653/V1/2021.ACL-SHORT.24.

Luce L. 2018. Generative models as fashion designers. In: *Artificial Intelligence for Fashion: How AI is Revolutionizing the Fashion Industry*, 125–139 DOI 10.1007/978-1-4842-3931-5.

Ludo.ai. 2021. Create hit games with the power of AI. *Available at https://ludo.ai/* (accessed 17 July 2025).

Magro-Vela S, Sánchez-López P, Navarro-Sierra N. 2024. The revolution will be artificial: an analysis of AI-generated audio-visual creation. *Tripodos* **55**:75–98 DOI 10.51698/TRIPODOS.2024.55.05.

Matasyoh NM, Zeineldin RA, Mathis-Ullrich F. 2024. Optimising speech recognition using LLMs: an application in the surgical domain. *Current Directions in Biomedical Engineering* **10**:45–48 DOI 10.1515/CDBME-2024-0112.

Mayahi S, Vidrih M. 2022. The impact of generative AI on the future of visual content marketing. ArXiv DOI 10.48550/arxiv.2211.12660.

Meta AI. 2023. Introducing LLaMA: a foundational, 65-billion-parameter language model. *Available at https://ai.meta.com/blog/large-language-model-llama-meta-ai/* (accessed 27 March 2025).

Microsoft. 2014. Cortana. *Available at https://www.microsoft.com/en-us/cortana* (accessed 19 November 2023).

Microsoft Azure. 2024. Phi open models: small language models. *Available at https://azure.microsoft.com/en-us/products/phi* (accessed 17 July 2025).

Microsoft Copilot. 2023. Your everyday AI companion. *Available at https://www.microsoft.com/en-us/microsoft-copilot/organizations* (accessed 8 January 2024).

Microsoft Research. 2023. Phi-2: the surprising power of small language models. *Available at https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/* (accessed 20 October 2024).

Midjourney. *Available at https://www.midjourney.com/home?callbackUrl=%2Fexplore* (accessed 19 November 2023).

Mistral AI. 2023. La-plateforme. *Available at https://mistral.ai/news/la-plateforme/*.

Mistral AI. 2024. Le chat. *Available at https://mistral.ai/news/le-chat-mistral/*.

modl.ai. 2023. AI engine for game development. *Available at https://modl.ai/* (accessed 17 July 2025).

Mordvintsev A, Olah C, Tyka M. 2015. DeepDream: a code example for visualizing neural networks. *Available at https://ai.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html* (accessed 26 March 2025).

Muennighoff N, Wang T, Sutawika L, Roberts A, Biderman S, Le Scao T, Bari MS, Shen S, Yong ZX, Schoelkopf H, Tang X, Radev D, Aji AF, Almubarak K, Albanie S, Alyafeai Z, Webson A, Raff E, Raffel C. 2022. Crosslingual generalization through multitask finetuning. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* **1**:15991–16111 DOI 10.18653/v1/2023.acl-long.891.

Mukhamediev RI, Popova Y, Kuchin Y, Zaitseva E, Kalimoldayev A, Symagulov A, Levashenko V, Abdoldina F, Gopejenko V, Yakunin K, Muhamedijeva E, Yelis M. 2022. Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges. *Mathematics* **10(15)**:1–25 DOI 10.3390/math10152552.

Murf AI. 2020. Free AI voice generator: versatile text to speech software. *Available at https://murf.ai/* (accessed 27 March 2025).

Murugesan S, Cherukuri AK. 2023. The rise of generative artificial intelligence and its impact on education: the promises and perils. *Computer* **56(5)**:116–121 DOI 10.1109/MC.2023.3253292.

Narang S, Chowdhery A. 2022. Pathways Language Model (PaLM): scaling to 540 billion parameters for breakthrough performance. *Available at https://research.google/blog/pathways-language-model-palm-scaling-to-540-billion-parameters-for-breakthrough-performance/* (accessed 27 March 2025).

NVIDIA. 2025. Build a custom LLM with chat with RTX. *Available at https://www.nvidia.com/en-me/ai-on-rtx/chat-with-rtx-generative-ai/* (accessed 17 March 2024).

NVIDIA. 2021. GauGAN2. *Available at http://ww7.gaugan.org/gaugan2/?usid=25&utid=4643646231* (accessed 8 January 2024).

NVIDIA Developer. 2023. ACE for games. *Available at https://developer.nvidia.com/ace-for-games* (accessed 17 July 2025).

Okaiyeto SA, Bai J, Xiao H. 2023. Generative AI in education: to embrace it or not? *International Journal of Agricultural and Biological Engineering* **16**:285–286.

OpenAI. 2021a. CLIP: connecting text and images. *Available at https://openai.com/index/clip/* (accessed 27 March 2025).

OpenAI. 2021b. Codex. *Available at https://openai.com/codex/* (accessed 17 July 2025).

OpenAI. 2022. DALL·E 2. *Available at https://openai.com/dall-e-2* (accessed 25 November 2023).

OpenAI. 2023a. DALL·E 3. *Available at https://openai.com/dall-e-3* (accessed 25 November 2023).

OpenAI. 2023b. GPT-4. *Available at https://openai.com/research/gpt-4* (accessed 19 November 2023).

OpenAI. 2023c. ChatGPT. *Available at https://chat.openai.com/* (accessed 19 November 2023).

OpenAI. 2024a. Hello GPT-4o. *Available at https://openai.com/index/hello-gpt-4o/* (accessed 27 March 2025).

OpenAI. 2019. MuseNet. *Available at https://openai.com/research/musenet* (accessed 20 November 2023).

OpenAI. 2024b. Sora. *Available at https://openai.com/sora*.

OpenAI. 2021c. DALL·E: creating images from text. *Available at https://openai.com/research/dall-e* (accessed 19 November 2023).

OpenAI. 2023d. GPT-4 technical report. ArXiv DOI 10.48550/arxiv.2303.08774.

Originality.ai. 2025. Baidu Ernie Bot statistics. *Available at https://originality.ai/blog/baidu-ernie-bot-statistics*.

Ostendorff M, Suarez PO, Lage LF, Rehm G. 2024. LLM-datasets: an open framework for pretraining datasets of large language models. *Available at https://openreview.net/forum?id=5RdIMlGLXL*.

Pahune S, Rewatkar N. 2023. Healthcare: a growing role for large language models and generative AI. *International Journal for Research in Applied Science and Engineering Technology* **11(8)**:2288–2301 DOI 10.22214/ijraset.2023.55573.

Pharma.AI. PandaOmics. *Available at https://pharma.ai/pandaomics* (accessed 17 July 2025).

Papineni K, Roukos S, Ward T, Zhu W-J. 2002. BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia*, 311–318.

Parikh NA. 2023. Empowering business transformation: the positive impact and ethical considerations of generative AI in software product management: a systematic literature review. *Available at https://www.igi-global.com/gateway/chapter/332749*.

Penedo G, Malartic Q, Hesslow D, Cojocaru R, Alobeidli H, Cappelli A, Pannier B, Almazrouei E, Launay J. 2023. The RefinedWeb Dataset for Falcon LLM: outperforming curated corpora with web data only. ArXiv DOI 10.48550/arXiv.2306.01116.

Perełkiewicz M, Poświata R. 2025. A review of the challenges with massive web-mined corpora used in large language models pre-training. In: Rutkowski L, Scherer R, Korytkowski M, Pedrycz W, Tadeusiewicz R, Zurada JM, eds. *Artificial Intelligence and Soft Computing. ICAISC 2024. Lecture Notes in Computer Science*. Cham: Springer, 153–163 DOI 10.1007/978-3-031-81596-6_14.

Perplexity.ai. 2022. Perplexity.ai. *Available at https://www.perplexity.ai/*.

Pika. 2023. Pika. *Available at https://pika.art/* (accessed 8 January 2024).

Piktus A, Akiki C, Villegas P, Laurencon H, Dupont G, Luccioni AS, Jernite Y, Rogers A. 2023. The ROOTS search tool: data transparency for LLMs. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* **3**:304–314 DOI 10.18653/V1/2023.ACL-DEMO.29.

Pillutla K, Swayamdipta S, Zellers R, Thickstun J, Welleck S, Choi Y, Harchaoui Z. 2021. MAUVE: measuring the gap between neural text and human text using divergence frontiers. *Advances in Neural Information Processing Systems* **6**:4816–4828.

Plummer BA, Wang L, Cervantes CM, Caicedo JC, Hockenmaier J, Lazebnik S. 2015. Flickr30k entities: collecting region-to-phrase correspondences for richer image-to-sentence models. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. New York: Springer, 2641–2649 DOI 10.1007/s11263-016-0965-7.

Puck. 2025. Puck. *Available at https://gamesbypuck.itch.io/puck* (accessed 17 July 2025).

Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. 2021. Learning transferable visual models from natural language supervision. ArXiv DOI 10.48550/arXiv.2103.00020.

Radford A, Metz L, Chintala S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. In: *Conference Track Proceedings of the 4th International Conference on Learning Representations (ICLR, 2016)*.

Radford L, Narasimhan K, Salimans T, Sutskever I. 2018. Improving language understanding by generative pre-training. OpenAI Blog. *Available at https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf*.

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. 2019. Language models are unsupervised multitask learners. *OpenAI Blog. Available at https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf*.

Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* **21**:1–67.

Raffiee AH, Sollami M. 2020. GarmentGAN: photo-realistic adversarial fashion transfer. In: *Proceedings - International Conference on Pattern Recognition*, 3923–3930 DOI 10.1109/ICPR48806.2021.9412908.

Raut G, Singh A. 2024. Generative AI in vision: a survey on models, metrics and applications. ArXiv DOI 10.48550/arXiv.2402.16369.

Relmasira SC, Lai YC, Donaldson JP. 2023. Fostering AI literacy in elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) education in the age of generative AI. *Sustainability* **15(18)**:13595 DOI 10.3390/SU151813595.

runway. 2023a. Gen-1: the next step forward for generative AI. *Available at https://research.runwayml.com/gen1*.

runway. 2023b. Gen-2: generate novel videos with text, images or video clips. *Available at https://runwayml.com/research/gen-2*.

Runway Research. 2024. Introducing Gen-3 Alpha: a new frontier for video generation. *Available at https://runwayml.com/research/introducing-gen-3-alpha* (accessed 27 March 2025).

Saghiri AM, Vahidipour SM, Jabbarpour MR, Sookhak M, Forestiero A. 2022. A survey of artificial intelligence challenges: analyzing the definitions, relationships, and evolutions. *Applied Sciences (Switzerland)* **12(8)**:4054 DOI 10.3390/app12084054.

Saharia C, Chan W, Saxena S, Li L, Whang J, Denton E, Ghasemipour SKS, Ayan BK, Mahdavi SS, Gontijo-Lopes R, Salimans T, Ho J, Fleet DJ, Norouzi M. 2022. Photorealistic text-to-image diffusion models with deep language understanding. ArXiv DOI 10.48550/arXiv.2205.11487.

Sáiz-Manzanares MC, Marticorena-Sánchez R, Ochoa-Orihuel J. 2020. Effectiveness of using voice assistants in learning: a study at the time of covid-19. *International Journal of Environmental Research and Public Health* **17(15)**:1–20 DOI 10.3390/ijerph17155618.

Salem FM. 2017. Gated RNN: the long short-term memory (LSTM) RNN. In: *Recurrent Neural Networks*. Cham: Springer, 71–82.

Salem FM. 2022. Gated RNN: the gated recurrent unit (GRU) RNN. In: *Recurrent Neural Networks*, 85–100 DOI 10.1007/978-3-030-89929-5_5.

Sandhu R, Channi HK, Ghai D, Cheema GS, Kaur M. 2024. An introduction to generative AI tools for education 2030. In: *Integrating Generative AI in Education to Achieve Sustainable Development Goals*. IGI Global Scientific Publishing, 1–28 DOI 10.4018/979-8-3693-2440-0.CH001.

Sanh V, Webson A, Raffel C, Bach SH, Sutawika L, Alyafeai Z, Chaffin A, Stiegler A, Le ST, Raja A, Dey M, Bari MS, Xu C, Thakker U, Sharma SS, Szczechla E, Kim T, Chhablani G, Nayak N, Datta D, Chang J, Jiang MT-J, Wang H, Manica M, Shen S, Yong ZX, Pandey H, Bawden R, Wang T, Neeraj T, Rozen J, Sharma A, Santilli A, Fevry T, Fries JA, Teehan R, Bers T, Biderman S, Gao L, Wolf T, Rush AM. 2021. Multitask prompted training enables zero-shot task generalization. ArXiv DOI 10.48550/arxiv.2110.08207.

Saxena D, Cao J. 2020. Generative adversarial networks (GANs): challenges, solutions, and future directions. ArXiv DOI 10.48550/arxiv.2005.00065.

Scao TL, Fan A, Akiki C, Pavlick E, Ilić S, Hesslow D, Castagné R, Luccioni AS, Yvon F, Gallé M, Tow J, Rush AM, Biderman S, Webson A, Ammanamanchi PS, Wang T, Sagot B, Muennighoff N, del Moral AV, Ruwase O, Bawden R, Bekman S, McMillan-Major A, Beltagy I, Nguyen H, Saulnier L, Tan S, Suarez PO, Sanh V, Laurençon H, Jernite Y, Launay J, Mitchell M, Raffel C, Gokaslan A, Simhi A, Soroa A, Aji AF, Alfassy A, Rogers A, Nitzav AK, Xu C, Mou C, Emezue C, Klamm C, Leong C, van Strien D, Adelani DI, Radev D, Ponferrada EG, Levkovizh E, Kim E, Natan EB, De Toni F, Dupont G, Kruszewski G, Pistilli G, Elsahar H, Benyamina H, Tran H, Yu I, Abdulmumin I, Johnson I, Gonzalez-Dios I, de la Rosa J, Chim J, Dodge J, Zhu J, Chang J, Frohberg J, Tobing J, Bhattacharjee J, Almubarak K, Chen K, Lo K, Von Werra L, Weber L, Phan L, Allal LB, Tanguy L, Dey M, Muñoz MR, Masoud M, Grandury M, Šaško M, Huang M, Coavoux M, Singh M, Jiang MT-J, Vu MC, Jauhar MA, Ghaleb M, Subramani N, Kassner N, Khamis N, Nguyen O, Espejel O, de Gibert O, Villegas P, Henderson P, Colombo P, Amuok P, Lhoest Q, Harliman R, Bommasani R, López RL, Ribeiro R, Osei S, Pyysalo S, Nagel S, Bose S, Muhammad SH, Sharma S, Longpre S, Nikpoor S, Silberberg S, Pai S, Zink S, Torrent TT, Schick T, Thrush T, Danchev V, Nikoulina V, Laippala V, Lepercq V, Prabhu V, Alyafeai Z, Talat Z, Raja A, Heinzerling B, Si C, Taşar DE, Salesky E, Mielke SJ, Lee WY, Sharma A, Santilli A, Chaffin A, Stiegler A, Datta D, Szczechla E, Chhablani G, Wang H, Pandey H, Strobelt H, Fries JA, Rozen J, Gao L, Sutawika L, Bari MS, Al-shaibani MS, Manica M, Nayak N, Teehan R, Albanie S, Shen S, Ben-David S, Bach SH, Kim T, Bers T, Fevry T, Neeraj T, Thakker U, Raunak V, Tang X, Yong Z-X, Sun Z, Brody S, Uri Y, Tojarieh H, Roberts A, Chung HW, Tae J, Phang J, Press O, Li C, Narayanan D, Bourfoune H, Casper J, Rasley J, Ryabinin M, Mishra M, Zhang M, Shoeybi M, Peyrounette M, Patry N, Tazi N, Sanseviero O, von Platen P, Cornette P, Lavallée PF, Lacroix R, Rajbhandari S, Gandhi S, Smith S, Requena S, Patil S, Dettmers T, Baruwa A, Singh A, et al. 2022. BLOOM: a 176B-parameter open-access multilingual language model. ArXiv DOI 10.48550/arXiv.2211.05100.

Schuhmann C, Beaumont R, Vencu R, Gordon C, Wightman R, Cherti M, Coombes T, Katta A, Mullis C, Wortsman M, Schramowski P, Kundurthy S, Crowson K, Schmidt L, Kaczmarczyk R, Jitsev J. 2022. LAION-5B: an open large-scale dataset for training next generation image-text models. ArXiv DOI 10.48550/arXiv.2210.08402.

Sengar SS, Bin HA, Kumar S, Carroll F. 2024. Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications* **84(21)**:23661–23700 DOI 10.1007/s11042-024-20016-1.

Shen A, Qi J, Baldwin T. 2017. A hybrid model for quality assessment of Wikipedia articles. In: *Proceedings of Australasian Language Technology Association Workshop*, 43–52.

Shi W, Wong W, Zou X. 2025. Generative AI in fashion: overview. *ACM Transactions on Intelligent Systems and Technology* **8**:2575 DOI 10.1145/3718098.

Shokrollahi Y, Yarmohammadtoosky S, Nikahd MM, Dong P, Li X, Gu L. 2023. A comprehensive review of generative AI in healthcare. ArXiv DOI 10.48550/arXiv.2310.00795.

Singer U, Polyak A, Hayes T, Yin X, An J, Zhang S, Hu Q, Yang H, Ashual O, Gafni O, Parikh D, Gupta S, Taigman Y. 2022. Make-a-video: text-to-video generation without text-video data. ArXiv DOI 10.48550/arxiv.2209.14792.

Singh M, Bajpai U, Vijayarajan V, Prasath S. 2020. Generation of fashionable clothes using generative adversarial networks: a preliminary feasibility study. *International Journal of Clothing Science and Technology* **32(2)**:177–187 DOI 10.1108/IJCST-12-2018-0148.

Sohl-Dickstein J, Weiss EA, Maheswaranathan N, Ganguli S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: *Proceedings of the 32nd International Conference on Machine Learning*. Lille, France.

SOUNDRAW. 2021. AI music generator. *Available at https://soundraw.io/?ref=bohoangc&gad_source=1&gclid=CjwKCAjw7pO_BhAlEiwA4pMQvOAdSXTEEYmFeFzJnFo2mStfzVrQP5eashBmFnClz5rlvJSqlaKJcxoCplMQAvD_BwE* (accessed 27 March 2025).

Stability AI. 2022. Stable diffusion. *Available at https://stability.ai/stable-diffusion* (accessed 27 November 2023).

Stability AI. 2023. Stable video diffusion now available on stability AI developer platform API. *Available at https://stability.ai/news/introducing-stable-video-diffusion-api* (accessed 27 March 2025).

Sudhakaran S, González-Duque M, Freiberger M, Glanois C, Najarro E, Risi S. 2023. MarioGPT: open-ended text2level generation through large language models. ArXiv DOI 10.48550/arXiv.2302.05981.

Sun Y, Wang S, Li Y, Feng S, Chen X, Zhang H, Tian X, Zhu D, Tian H, Wu H. 2019. ERNIE: enhanced representation through knowledge integration. ArXiv DOI 10.48550/arxiv.1904.09223.

Sun Y, Wang S, Li Y, Feng S, Tian H, Wu H, Wang H. 2020. ERNIE 2.0: a continual pre-training framework for language understanding. In: *AAAI, 2020 - 34th AAAI Conference on Artificial Intelligence*. Washington, D.C.: AAAI, 8968–8975 DOI 10.1609/aaai.v34i05.6428.

Sutskever I, Vinyals O, Le QV. 2014. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems* **4**:3104–3112.

Sutskever I, Vinyals O, Le QV, Cho K, van MB, Schwenk H, Bengio Y. 2014. Learning phrase representation using RNN encoder-decoder. *Proceedings of the Empherical Methods in Natural Language Processing* **4**:1724–1734.

Synthesia. 2023. #1 AI video generator. *Available at https://www.synthesia.io/?via=bestaitool/* (accessed 5 December 2023).

Tafesse W, Wien A. 2024. ChatGPT's applications in marketing: a topic modeling approach. *Marketing Intelligence and Planning* **42**:666–683 DOI 10.1108/MIP-10-2023-0526/FULL/PDF.

Tahmid M, Alam MS, Rao N, Ashrafi KMA. 2016. Image-to-image translation with conditional adversarial networks. In: *Proceedings of 2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE 2023)*, 468–472 DOI 10.1109/WIECON-ECE60392.2023.10456447.

Taulli T. 2023. *Generative AI: how ChatGPT and other AI tools will revolutionize business*. New York: Apress.

Thomas A, Gaizauskas R, Lu H. 2024. Leveraging LLMs for post-OCR correction of historical newspapers. In: *Proceedings of the 1st Workshop on Language Technology for Historical and Ancient Languages (LT4HALA 2024)*, 116–121.

Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng H-T, Jin A, Bos T, Baker L, Du Y, Li Y, Lee H, Zheng HS, Ghafouri A, Menegali M, Huang Y, Krikun M, Lepikhin D, Qin J, Chen D, Xu Y, Chen Z, Roberts A, Bosma M, Zhao V, Zhou Y, Chang C-C, Krivokon I, Rusch W, Pickett M, Srinivasan P, Man L, Meier-Hellstern K, Morris MR, Doshi T, Santos RD, Duke T, Soraker J, Zevenbergen B, Prabhakaran V, Diaz M, Hutchinson B, Olson K, Molina A, Hoffman-John E, Lee J, Aroyo L, Rajakumar R, Butryna A, Lamm M, Kuzmina V, Fenton J, Cohen A, Bernstein R, Kurzweil R, Aguera-Arcas B, Cui C, Croak M, Chi E, Le Q. 2022. LaMDA: language models for dialog applications. ArXiv DOI 10.48550/arxiv.2201.08239.

Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G. 2023. LLaMA: open and efficient foundation language models. ArXiv DOI 10.48550/arXiv.2302.13971.

Trufinescu A. 2024. Discover the new multi-lingual, high-quality Phi-3.5 SLMs. *Available at https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/discover-the-new-multi-lingual-high-quality-phi-3-5-slms/ba-p/4225280* (accessed 20 October 2024).

Üstün A, Aryabumi V, Yong Z-X, Ko W-Y, D'Souza D, Onilude G, Bhandari N, Singh S, Ooi H-L, Kayid A, Vargus F, Blunsom P, Longpre S, Muennighoff N, Fadaee M, Kreutzer J, Hooker S. 2024. Aya model: an instruction finetuned open-access multilingual language model. ArXiv DOI 10.48550/arxiv.2402.07827.

van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K. 2016. WaveNet: a generative model for raw audio. ArXiv DOI 10.48550/arxiv.1609.03499.

Varghese J, Chapiro J. 2024. ChatGPT: the transformative influence of generative AI on science and healthcare. *Journal of Hepatology* **80(6)**:977–980 DOI 10.1016/J.JHEP.2023.07.028.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. The transformer. In: *Advances in Neural Information Processing Systems*, 5998–6008 DOI 10.1007/978-3-319-29409-4_3.

Vedantam R, Zitnick CL, Parikh D. 2014. CIDEr: consensus-based image description evaluation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **5302**:4566–4575 DOI 10.1109/CVPR.2015.7299087.

Wang T. 2023. Navigating Generative AI (ChatGPT) in higher education: opportunities and challenges. In: *Smart Learning for A Sustainable Society: Proceedings of the 7th International Conference on Smart Learning Environments*. Singapore: Springer Nature Singapore, 215–225.

Wang Y, He Y, Li Y, Li K, Yu J, Ma X, Li X, Chen G, Chen X, Wang Y, Luo P, Liu Z, Wang Y, Wang L, Qiao Y. 2023. InternVid: a large-scale video-text dataset for multimodal understanding and generation. In: *12th International Conference on Learning Representations, ICLR 2024*.

Wang S, Sun Y, Xiang Y, Wu Z, Ding S, Gong W, Feng S, Shang J, Zhao Y, Pang C, Liu J, Chen X, Lu Y, Liu W, Wang X, Bai Y, Chen Q, Zhao L, Li S, Sun P, Yu D, Ma Y, Tian H, Wu H, Wu T, Zeng W, Li G, Gao W, Wang H. 2021. ERNIE 3.0 Titan: exploring larger-scale knowledge enhanced pre-training for language understanding and generation. ArXiv DOI 10.48550/arxiv.2112.12731.

Wang Y, Wang M, Manzoor MA, Liu F, Georgiev G, Das RJ, Nakov P. 2024. Factuality of large language models: a survey. ArXiv DOI 10.48550/arXiv.2402.02420.

Wang S, Zhou T, Shen Y, Li Y, Huang G, Hu Y. 2025. Generative AI enables EEG super-resolution via spatio-temporal adaptive diffusion learning. *IEEE Transactions on Consumer Electronics* **71(1)**:1034–1045 DOI 10.1109/TCE.2025.3528438.

Wenzek G, Lachaux M-A, Conneau A, Chaudhary V, Guzmán F, Joulin A, Grave É. 2020. CCNet: extracting high quality monolingual datasets from web crawl data. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 4003–4012.

Werning S. 2024. *Generative AI and the Technological imaginary of game design*. London: Palgrave Macmillan, 67–90.

WGSN. Trend forecasting & analytics 2025–2032. *Available at https://www.wgsn.com/en* (accessed 17 July 2025).

Wu C, Liang J, Ji L, Yang F, Fang Y, Jiang D, Duan N. 2022. NÜWA: visual synthesis pre-training for Neural visUal World creAtion. In: *European Conference on Computer Vision*, 720–736 DOI 10.1007/978-3-031-19787-1_41.

X AI. 2023. Announcing Grok. *Available at https://x.ai/* (accessed 19 November 2023).

Yacoby Y, Pan W, Doshi-Velez F. 2020. Failure modes of variational autoencoders and their effects on downstream tasks. ArXiv DOI 10.48550/arXiv.2007.07124.

Yan H, Zhang H, Liu L, Zhou D, Xu X, Zhang Z, Yan S. 2023. Toward intelligent design: an AI-based fashion designer using generative adversarial networks aided by sketch and rendering generators. *IEEE Transactions on Multimedia* **25**:2323–2338 DOI 10.1109/TMM.2022.3146010.

Yu Y, Zhang W, Deng Y. 2021. Frechet inception distance (FID) for evaluating GANs. *Available at https://www.researchgate.net/publication/354269184_Frechet_Inception_Distance_FID_for_Evaluating_GANs*.

Zant TVD, Kouw M, Schomaker L. 2013. Generative artificial intelligence. In: Müller V, ed. *Philosophy and Theory of Artificial Intelligence: Studies in Applied Philosophy, Epistemology and Rational Ethics*. Berlin, Heidelberg: Springer, 107–120.

Zhan J, Dai J, Ye J, Zhou Y, Zhang D, Liu Z, Zhang X, Yuan R, Zhang G, Li L, Yan H, Fu J, Gui T, Sun T, Jiang Y, Qiu X. 2024. AnyGPT: unified multimodal LLM with discrete sequence modeling. ArXiv DOI 10.48550/arxiv.2402.12226.

Zhang P, Kamel Boulos MN. 2023. Generative AI in medicine and healthcare: promises, opportunities and challenges. *Future Internet* **15(9)**:286 DOI 10.3390/FI15090286.

Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y. 2019. BERTScore: evaluating text generation with BERT. In: *8th International Conference on Learning Representations (ICLR 2020)*.

Zhang Q, Lu J, Jin Y. 2021. Artificial intelligence in recommender systems. *Complex and Intelligent Systems* **7(1)**:439–457 DOI 10.1007/s40747-020-00212-w.

Zhang Y, Mastouri M, Zhang Y. 2024. Accelerating drug discovery, development, and clinical trials by artificial intelligence. *Medicine* **5**:1050–1070 DOI 10.1016/j.medj.2024.07.026.

Zhang S, Roller S, Goyal N, Artetxe M, Chen M, Chen S, Dewan C, Diab M, Li X, Lin XV, Mihaylov T, Ott M, Shleifer S, Shuster K, Simig D, Koura PS, Sridhar A, Wang T, Zettlemoyer L. 2022. OPT: open pre-trained transformer language models. ArXiv DOI 10.48550/arxiv.2205.01068.

Zhong J, Huyan J, Zhang W, Cheng H, Zhang J, Tong Z, Jiang X, Huang B. 2023. A deeper generative adversarial network for grooved cement concrete pavement crack detection. *Engineering Applications of Artificial Intelligence* **119(13)**:105808 DOI 10.1016/J.ENGAPPAI.2022.105808.

Zohny H, McMillan J, King M. 2023. Ethics of generative AI. *Journal of Medical Ethics* **49(2)**:79–80 DOI 10.1136/JME-2023-108909.