

RESEARCH

Open Access



Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond

Zarif Bin Akhtar^{1*}

*Correspondence:
zarifbinakhtarg@gmail.com;
zarifbinakhtar@ieee.org

¹ Department of Engineering,
University of Cambridge, Cambridge,
UK

Abstract

This comprehensive exploration of recent breakthroughs in artificial intelligence (AI) traversed the realms of language models, computer vision, and generative models, unraveling the intricacies of cutting-edge technologies such as GPT-3.5, GPT-4, Pix-2Seq, and multimodal models in terms of generative AI. In this multifaceted journey, the focus extended beyond technological prowess to ethical considerations, emphasizing responsible AI practices guided by Google's AI Principles. The nuanced discussions encapsulated the transformative impact of AI on user experiences across various Google products and toolsets, paving the way for a future where natural language interaction, creative content generation, and multimodal understanding redefine human–computer interactions. The research investigation showcased not only the advancements themselves but also the critical lens through which these innovations are approached, underscoring the importance of ethical and responsible AI in shaping the technological landscape.

Keywords: Artificial intelligence (AI), Computer vision, Deep learning (DL), Generative artificial intelligence (GAI), Large language models (LLM), Machine intelligence, Machine learning (ML)

Introduction

Generative artificial intelligence (AI) refers to algorithms, such as ChatGPT, that have the capacity to create diverse content, including audio, code, images, text, simulations, and videos. This research exploration and investigation discusses the transformative impact of recent breakthroughs in generative AI, highlighting the potential to revolutionize content creation across various domains.

The piece introduces ChatGPT, an AI chatbot developed by OpenAI, focusing on its ability to generate answers to a wide array of questions [1]. The exploration underscores ChatGPT's popularity, with over a million users signing up within five days of its public release in November 2022 [1, 2]. It explores the range of content that ChatGPT can produce, from computer code and essays to poems, showcasing its versatility and effectiveness. The discussions expand toward encompassing the broader landscape of generative

AI, mentioning DALL-E, a tool for AI-generated art, and the overall impact of generative AI on diverse job roles. While some individuals express fear regarding the technology, the investigation emphasizes the positive contributions of machine learning, particularly in areas like computer vision object detection, medical imaging analysis, and model forecasting generation and deployments, with various areas of interest highlighted by a 2022 McKinsey survey [3].

To understand the foundations of generative AI, this research delves into the distinctions between artificial intelligence and machine learning. AI is defined as the emulation of human intelligence by machines, with examples like voice assistants (Siri and Alexa) and customer service chatbots [4–6]. Machine learning, a subset of AI, involves creating AI models that can learn from data patterns without explicit human guidance. The research also outlines the historical development of machine learning models, starting from classical statistical techniques to the recent advancements in self-supervised learning. It introduces the predecessors of ChatGPT, namely GPT-3 and Google's BERT, highlighting the challenges faced by text-based machine learning models in the past [7–9]. Addressing the question of how generative AI models work, the exploration explains the transition from supervised learning, where models are trained by humans based on labeled data, to self-supervised learning. ChatGPT, representing the latest advancements, is lauded for its improved performance, generating high-quality responses in various contexts [9, 10]. The complexities of building a generative AI model are discussed, emphasizing the substantial resources and expertise required. Companies like OpenAI, DeepMind, and Meta, with substantial funding and top-tier talent, are at the forefront of developing such models. The investigation touches upon the cost associated with training these models, indicating that GPT-3 was trained on approximately 45 terabytes of text data, equivalent to a substantial investment. Generative AI's diverse outputs, ranging from text and images to code and business simulations, are explored, highlighting the lifelike and creative nature of the generated content. The discussion underscores the potential applications in various industries, such as IT, software development, marketing, and medical imaging. While recognizing the opportunities presented by generative AI, the research addresses the limitations and risks associated with biased or inaccurate outputs. It stresses the importance of careful data selection during model training, the potential need for human oversight, and the ongoing need for regulatory frameworks as the technology evolves. The research paints a comprehensive picture of generative AI, covering its evolution, applications, challenges, and future prospects. It serves as a valuable resource for understanding the dynamics of this rapidly advancing field and its implications across industries.

Methods and experimental analysis explorations

The methodology employed in examining recent advancements in artificial intelligence (AI) was methodically structured to provide a comprehensive overview. Commencing with a thorough background research review, the exploration delved into specific domains of AI, including natural language processing, computer vision, and generative models. Focused discussions elucidated the intricacies of language models like GPT-3.5, GPT-4, Pix2Seq, and transformative applications across various types of Google products and platforms with associated toolset and models. Considerable attention was

dedicated to user experiences and the integration of AI into real-world scenarios. Ethical considerations and responsible AI practices, guided by Google's AI Principles, were thoroughly addressed, emphasizing transparency and accountability. Forward-looking perspectives highlighted the potential of multimodal models and the ongoing evolution of AI technologies. The methodology, with its structured approach, offered a wide array toward understanding of the diverse facets shaping contemporary artificial intelligence (AI) within a framework, an experimental simulation for generative AI model set for visualization.

Generative artificial intelligence (generative AI or GenAI) refers to artificial intelligence capable of producing text, images, or other media using generative models. These models learn patterns and structures from training data to generate new data with similar characteristics. Notable examples include large language model (LLM) chatbots like ChatGPT, Copilot, Bard, and LLaMA, as well as text-to-image AI art systems such as Stable Diffusion, Midjourney, and DALL-E [11]. The applications of generative AI span various industries, including art, writing, software development, healthcare, finance, gaming, marketing, and fashion.

The early 2020s witnessed a surge in investment in generative AI by major companies like Microsoft, Google, and Baidu, as well as numerous smaller firms. However, concerns have arisen regarding potential misuse, including cybercrime, fake news creation, and the production of deepfakes for deceptive purposes [12]. The history of artificial intelligence dates back to the 1956 Dartmouth College workshop, with subsequent waves of advancement and ethical discussions about creating artificial beings with human-like intelligence. Automated art concepts trace back to ancient Greece, while the idea of AI captivated society from the mid-twentieth century, culminating in Alan Turing's seminal 1950 paper, "Computing Machinery and Intelligence."

The evolution of AI in the 1950s saw slow progress due to high costs and limited computing capabilities. However, the 1956 Dartmouth Summer Research Project on AI became a landmark event, setting the stage for two decades of rapid advancements. Artists and researchers have utilized AI for artistic works since the early 1970s, exemplified by Harold Cohen's AARON program generating paintings [4, 13].

Markov chains, developed by Andrey Markov in the early twentieth century, have been crucial for modeling natural languages. In the late 2000s, deep learning advancements, including variational autoencoders and generative adversarial networks, enabled the practical development of deep neural networks capable of learning generative models for complex data like images [14]. The Transformer network's introduction in 2017 marked a significant leap in generative models, leading to the first generative pretrained transformer (GPT-1) in 2018. Subsequent models, such as GPT-2 and DALL-E, showcased advancements in generative AI art. In 2021, the release of DALL-E, followed by Midjourney and Stable Diffusion, marked practical high-quality AI art generation from natural language prompts [5]. GPT-4's release in March 2023 stirred discussions about whether it could be considered an early version of artificial general intelligence (AGI). While some argue it's a step toward AGI, others contend that generative AI is still far from reaching the benchmark of general human intelligence as of 2023 [15]. Generative AI systems can operate in different modalities, including text, code, images, audio, video, molecules, robotics, and business intelligence. These systems can be unimodal,

accepting one type of input, or multimodal, accepting multiple input types. Various generative AI models, such as GPT-4, OpenAI Codex, and DALL-E, have specific applications in text, code, and image generation. These models have been integrated into products like Microsoft Office, Google Photos, and Adobe Photoshop. Smaller models can run on smartphones and personal computers, while larger models with tens of billions of parameters may require accelerators like GPUs or AI accelerator chips. Very large models with hundreds of billions of parameters, such as GPT-4, typically run on datacenter computers as cloud services. Generative AI represents a transformative force with diverse applications across industries, raising both possibilities and concerns about its ethical use and potential societal impact [8,9].

Generative artificial intelligence (generative AI) has sparked significant concerns and challenges across various domains. These apprehensions have prompted protests, legal actions, and calls for the pause of AI experiments, leading multiple governments to take regulatory actions. Secretary-General António Guterres highlighted the dual potential of generative AI in a July 2023 United Nations Security Council briefing, acknowledging its enormous capacity for both positive and negative impacts on a global scale. He emphasized the potential for AI to contribute trillions to the global economy by 2030 but warned of catastrophic consequences if misused. One major concern revolves around job losses. The rise of generative AI, particularly in image generation, has led to significant unemployment in certain sectors. For instance, in China, 70% of jobs for video game illustrators were reportedly lost due to image generation AI. The 2023 Hollywood labor disputes also saw generative AI contributing to concerns, with industry figures expressing fears that artificial intelligence poses an existential threat to creative professions, impacting jobs in areas such as voice acting and video game illustration [7, 16].

The intersection of AI and employment disparities, especially among underrepresented groups globally, is a critical facet. While AI promises efficiency enhancements and skill acquisition, concerns about job displacement and biased recruiting persist. Strategies to address these issues involve regulation, inclusive design, and education to maximize benefits while minimizing harms [17–22]. In the financial sector, generative AI has led to significant investment surges, resulting in transformative tools like robo-advisors. However, economists like Daron Acemoglu have raised warnings about potential adverse consequences, including data harvesting, customer manipulation, and labor market disparities, underscoring the complex impact of AI on society [6, 23].

The integration of AI with social identities holds both promises and challenges. While AI has the potential to transform traditional research methods, biases ingrained in AI systems perpetuate stereotypes and marginalize certain groups. The need to address these biases for inclusivity is emphasized, calling for ethical considerations in the development and deployment of AI systems [24]. Deepfakes, AI-generated media that replace a person's likeness in existing images or videos, have raised widespread concerns due to their potential misuse in creating deceptive content, such as revenge porn, fake news, and financial fraud. This has led to industry and government responses to detect and limit their use [8, 25–28]. Audio deepfakes, where users manipulate software to generate controversial statements in the vocal style of celebrities, have raised ethical concerns. Instances of AI-generated music using cloned voices of famous musicians have gained

both popularity and criticism, leading to debates about the ethics and impact of such technology on the music industry [2, 29–32]. Generative AI's realistic content creation capabilities have been exploited in cybercrime, including phishing scams and disinformation campaigns. Recent research in 2023 revealed vulnerabilities in generative AI, enabling criminals to manipulate systems for harmful purposes, such as social engineering attacks and phishing attempts [33]. Misuse in journalism has been observed, with reports of CNET using an undisclosed internal AI tool to write stories and a German tabloid publishing a fake AI-generated interview with a public figure. These incidents highlight the ethical considerations and potential consequences of AI in journalistic practices [34].

Regulation is a critical aspect of addressing these concerns. In the European Union, the proposed Artificial Intelligence Act includes requirements to disclose copyrighted material used to train generative AI systems and label any AI-generated output as such. In the USA, voluntary agreements have been signed to watermark AI-generated content. China has introduced Interim Measures for the Management of Generative AI Services, regulating public-facing generative AI with requirements for watermarking, data collection restrictions, and adherence to socialist core values. These regulatory efforts aim to strike a balance between harnessing AI's potential and mitigating its risks.

Generative artificial intelligence (generative AI) is a technology that can create diverse content types, such as text, images, audio, and synthetic data. While the concept of generative AI dates back to the 1960s with chatbots, it gained significant traction in 2014 with the advent of generative adversarial networks (GANs). These machine learning algorithms enabled generative AI to produce authentic-looking images, videos, and audio. Recent advancements, particularly in large language models (LLMs) like transformers, have propelled generative AI into the mainstream, allowing for the rapid creation of high-quality text, graphics, and videos. Transformers, coupled with breakthrough language models, have played a crucial role in the widespread adoption of generative AI. These models, with billions or trillions of parameters, have the capacity to write engaging text, create realistic images, and even generate entertaining sitcoms. Multimodal AI innovations further extend generative AI's capabilities, enabling content creation across various media types, including text, graphics, and video. Despite the progress, early implementations of generative AI faced challenges such as accuracy issues, biases, hallucinations, and odd responses. Nevertheless, the technology's potential to transform enterprise operations is significant, with envisioned applications ranging from code writing and drug design to product development and supply chain transformation. Generative AI operates by receiving a prompt, which can be in various forms like text, images, or videos, and then leveraging AI algorithms to produce new content in response [35, 36]. Recent developments focus on enhancing user experiences, allowing users to describe requests in plain language and customize results based on feedback about style, tone, and other elements [1].

Generative AI models combine different algorithms to process and represent content. Natural language processing techniques transform raw characters into sentences, parts of speech, entities, and actions, represented as vectors. Images undergo a similar transformation into various visual elements. Neural networks, such as GANs and variational autoencoders, play a crucial role in generating realistic content, while recent progress

in transformers enables the generation of text, images, and proteins. The date models also introduce popular generative AI interfaces, including DALL-E, ChatGPT, and Bard. DALL-E, trained on a large dataset of images and text descriptions, is an example of multimodal AI connecting words to visual elements. ChatGPT is an AI-powered chatbot built on OpenAI's GPT-3.5, GPT-4 offering interactive feedback via a chat interface. Bard, developed by Google, is a transformer AI-based chatbot that incorporates the LaMDA family of large language models [37, 38]. Use cases for generative AI are diverse, ranging from customer service chatbots and deepfake creation to improving movie dubbing and generating content like emails, dating profiles, art, and music. The technology's accessibility and adaptability make it a powerful tool for various applications, promising transformative impacts across industries. To better understand, Fig. 1 provides an illustration concerning the perspective matter.

Generative AI holds significant promise across various business functions, offering a range of benefits that can streamline processes, enhance creativity, and automate tasks. One key advantage is its ability to automate the manual process of content creation. By leveraging generative AI, businesses can reduce the time and effort spent on writing content, enabling a more efficient workflow. This can be particularly beneficial in areas such as marketing, where the demand for high-quality and engaging content is constant.

Additionally, generative AI can contribute to improving email response processes. Through its capacity to understand context and generate relevant and coherent responses, businesses can leverage generative AI to handle a portion of email communications. This not only saves time for employees but also ensures consistent and timely responses.

In the technical domain, generative AI can be employed to enhance responses to specific technical queries. By providing accurate and context-aware information, it

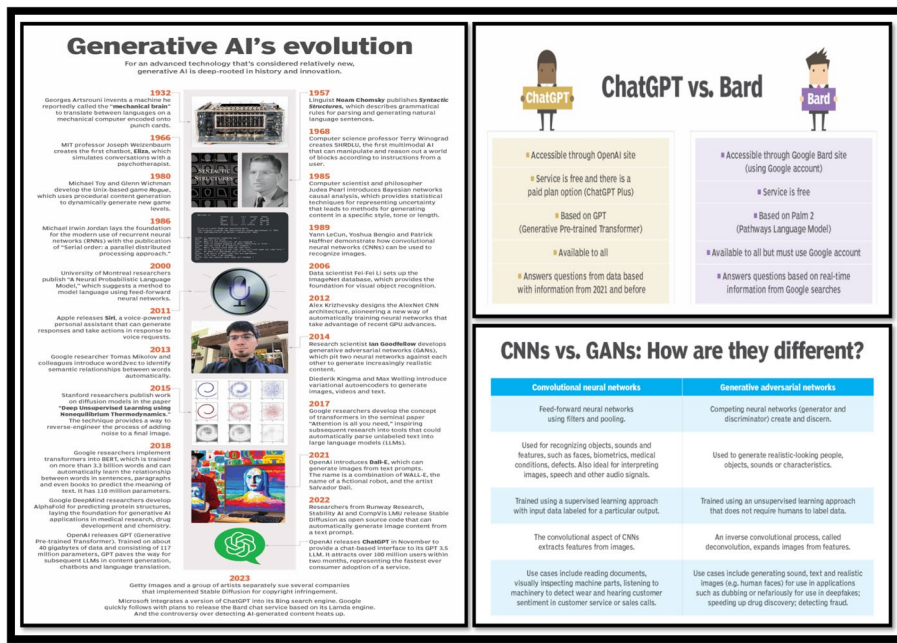


Fig. 1 An overview illustration concerning generative AI

aids in problem-solving and support functions. This can be particularly valuable for technical support teams, enabling them to address inquiries more efficiently. Generative AI's capability to create realistic representations of people is another noteworthy advantage.

This feature finds applications in various industries, including entertainment and advertising, where generating lifelike characters or models are crucial. It opens up possibilities for more immersive and visually appealing content creation. Moreover, generative AI excels in summarizing complex information into coherent narratives. This skill is valuable across different sectors, facilitating the communication of intricate details in a more digestible format. Businesses can use generative AI to distill complex reports, research findings, or data into accessible and comprehensible summaries. Another benefit lies in its ability to simplify the process of creating content in a specific style. Whether it's writing in a formal tone, adopting a casual style, or adhering to brand guidelines, generative AI can be customized to produce content that aligns with specific stylistic preferences. However, despite these advantages, generative AI also poses several challenges and limitations that need careful consideration. The technology's potential for inaccuracies, bias, and the generation of misleading information raises ethical concerns. Furthermore, issues related to trust, source identification, and the potential for disruption to existing business models are important factors to address when implementing generative AI. The introduction of transformers, such as GPT-3, has played a pivotal role in advancing generative AI capabilities. The attention mechanism introduced by transformers allows models to track connections between words across a broader context, enabling more sophisticated content generation. These advancements have propelled generative AI into a new era, where it can create engaging text, realistic images, and even entertainment content on a large scale. Despite its transformative potential, concerns surrounding generative AI are on the rise. The technology's realistic output raises issues related to accuracy, trustworthiness, bias, hallucination, and potential misuse. Detecting AI-generated content becomes challenging, posing risks in scenarios where accuracy is critical, such as in coding or medical advice.

Generative AI has found applications across various modalities, including text, imagery, music, code, and voices. Popular tools like GPT, Dall-E, and Bard exemplify its versatility. These tools cater to diverse use cases, from chat responses and image generation to music composition and code creation. The implementation of generative AI is not solely a technological consideration; it also has implications for ethics and bias. The realistic nature of the content generated by these tools introduces new challenges in terms of accuracy verification, source transparency, and ethical use. Detecting when AI-generated content is incorrect or potentially harmful becomes a complex task. To provide an idea relating to the matter, Fig. 2 represents the business aspect of generative AI. Looking ahead, the future of generative AI is expected to witness further advancements in translation, drug discovery, anomaly detection, and content generation across various domains. The technology's integration into existing tools and workflows is anticipated to redefine how businesses operate, offering more efficient grammar checkers, design tools, and training mechanisms. However, as the technology evolves, ongoing evaluation of its impact on human expertise, ethical considerations, and potential risks will be essential for responsible and effective implementation.

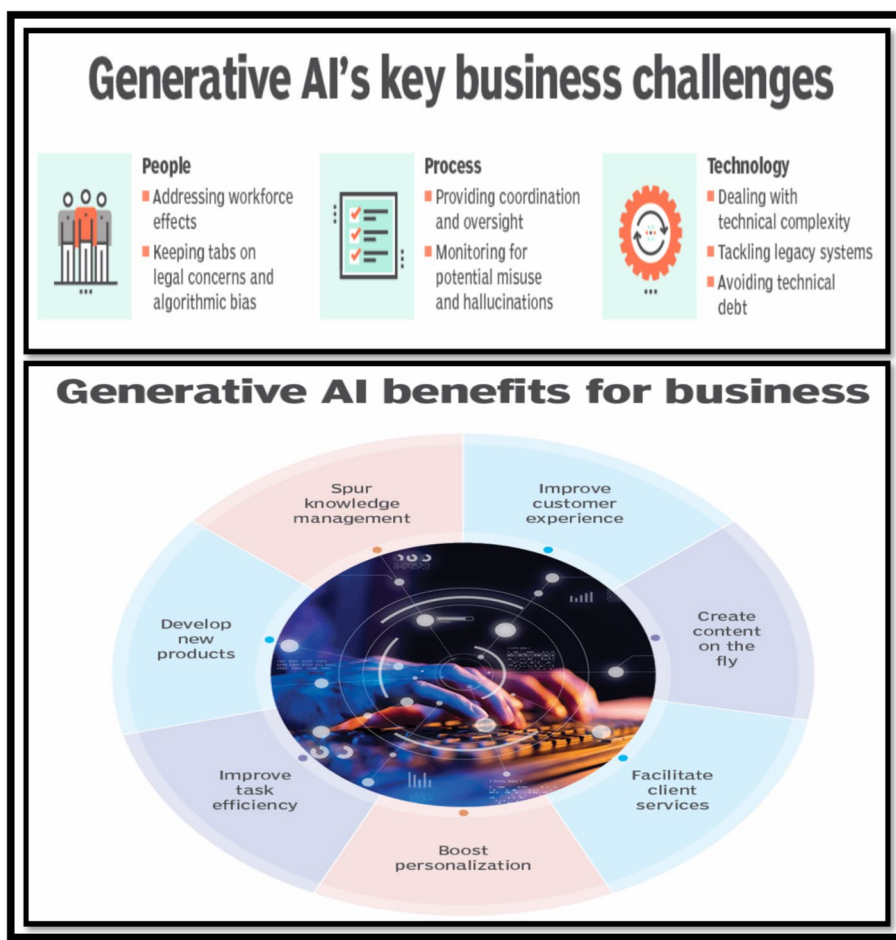


Fig. 2 The generative AI aspect in terms of business

The field of machine learning has witnessed remarkable progress in the last decade, particularly in the development of larger and more powerful language models (LLMs). Advances such as sequence-to-sequence learning and the introduction of the Transformer model, fundamental to recent breakthroughs, have significantly enhanced the capabilities of language models.

Despite being trained on seemingly simple objectives, such as predicting the next token in a sequence, large language models exhibit the ability to generate coherent, contextual, and natural-sounding responses. These models, exemplified by Google’s LaMDA, are versatile and find applications in creative content generation, language translation, coding assistance, and providing informative responses in conversations.

Google’s research on PaLM, a massive 540 billion parameter language model, demonstrated the impact of scale on improving state-of-the-art performance across various natural language, translation, and coding tasks. Language models trained on source code have proven valuable for internal developers, with ML-enhanced code completion reducing coding iteration time for Google software developers. Ongoing efforts involve enhancing these models to further benefit developers. One of the

critical challenges in artificial intelligence is enabling multistep reasoning, allowing models to break down complex problems into smaller tasks.

Google's Chain of Thought prompting encourages models to "show their work" in solving problems, resulting in more structured and accurate responses. This approach has shown particular effectiveness in solving complex mathematical and scientific problems. The Minerva effort, utilizing a general-purpose language model (PaLM) fine-tuned on mathematical and scientific documents, demonstrated substantial improvements over the state of the art in mathematical reasoning and scientific problem-solving. The model's ability to perform multistep reasoning significantly outperformed existing benchmarks, showcasing its proficiency in tackling intricate problems. Learned prompt tuning has emerged as a promising technique, adapting general-purpose language models to specific domains with relatively few examples.

In the medical domain, learned prompt tuning achieved high accuracy on medical questions, surpassing prior ML state of the art. This suggests the potential utility of LLMs in medicine, improving comprehension, knowledge recall, and medical reasoning. Large language models trained on multiple languages have been employed for translation tasks, even for languages without explicit translation training data [39, 40].

This approach, outlined in Google's work on zero-resource machine translation, adds support for 24 new languages to Google Translate, showcasing the versatility of LLMs in handling diverse linguistic tasks. Another intriguing aspect is the phenomenon observed in large language models, where their utility in performing complex tasks increases significantly with scale. As models reach a certain scale, they exhibit sudden improvements in the ability to perform complex tasks effectively. This prompts exploration into what new tasks may become feasible as these models continue to be trained at larger scales.

The advancements in LLMs, exemplified by Google's research, signify a transformative era in natural language processing and understanding. The integration of these models into various applications, from code completion to scientific problem-solving, highlights their potential to revolutionize how humans interact with and leverage computational systems. The ongoing exploration of scale-related phenomena and the adaptability of LLMs to new tasks hint at a future where these models play an increasingly vital role in diverse aspects of our lives. To better understand, Fig. 3 provides a visualization of the LLM models in action in terms of their performance.

Computer vision is an interdisciplinary field that encompasses methods for acquiring, processing, analyzing, and understanding digital images. It involves extracting high-dimensional data from the real world to produce numerical or symbolic information, facilitating decision-making processes. The core objective is to enable computers to gain high-level understanding from digital images or videos, akin to the human visual system.

The theoretical foundation of computer vision involves models constructed with geometry, physics, statistics, and learning theory. Its practical application, as a technological discipline, focuses on constructing computer vision systems. The sub-domains within computer vision are diverse and include scene reconstruction, object detection, event detection, activity recognition, video tracking, 3D pose estimation, and more. Organizations considering adopting computer vision technology may find it challenging, as there is no single-point solution, and only a few companies offer a unified platform for deploying and managing computer vision applications. The history of computer vision

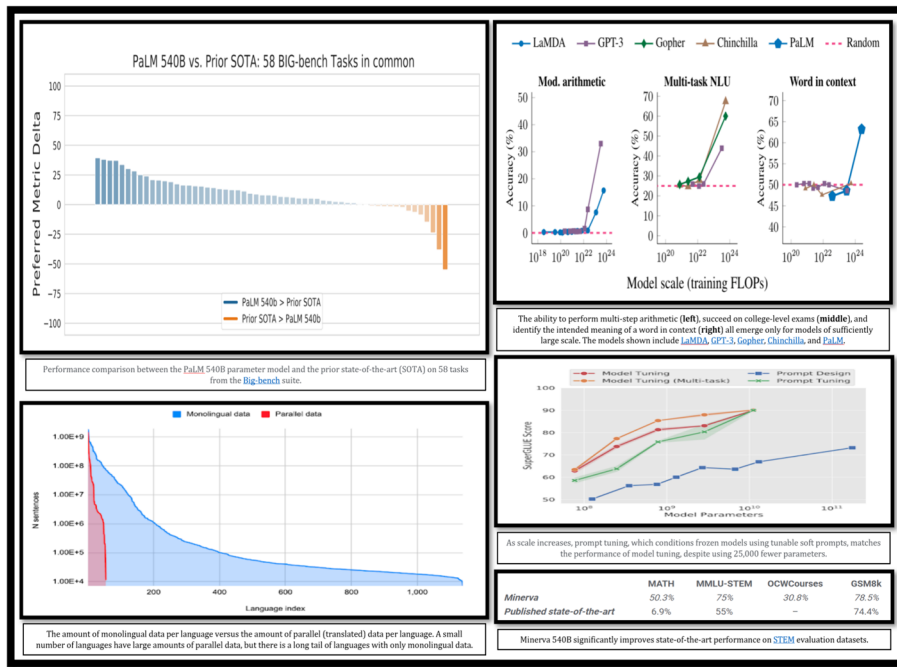


Fig. 3 A visualization of the various LLM models in action

traces back to the late 1960s when it emerged in universities pioneering artificial intelligence. Initially, the goal was to mimic the human visual system, eventually progressing to achieve full scene understanding. The 1970s laid the early foundations for computer vision algorithms, exploring topics such as edge extraction, modeling, optical flow, and motion estimation.

The subsequent decades saw mathematical analysis, quantitative aspects, and the convergence of computer graphics and computer vision. Recent advancements, particularly in deep learning, have revitalized the field of computer vision. Deep learning techniques have significantly improved accuracy in various computer vision tasks, surpassing traditional methods. This progress has breathed new life into feature-based methods, combining machine learning and complex optimization frameworks.

Computer vision is closely related to other scientific disciplines. Solid-state physics plays a crucial role in designing image sensors, explaining the interaction of light with surfaces. Neurobiology has greatly influenced computer vision algorithms, with the neocognitron being an early example inspired by the human visual cortex. Signal processing, robotic navigation, and other fields like object detection in photographs and photogrammetry overlap with computer vision. The distinctions between related fields such as image processing, image analysis, machine vision, and computer graphics provide insights into their specific focuses. While image processing deals with transforming images, computer vision involves 3D analysis from 2D images. Machine vision emphasizes applications in industrial settings, and computer graphics produces images from 3D models. Applications of computer vision span various industries and tasks.

From automatic inspection in manufacturing to assisting in identification tasks and controlling processes, computer vision finds utility in diverse scenarios. In medicine,

it aids in diagnosing conditions by extracting information from medical images, such as detecting tumors or measuring organ dimensions [39–41]. Machine vision supports industrial processes, including quality control in manufacturing. Military applications involve detecting enemy soldiers, missile guidance, and battlefield awareness. Autonomous vehicles leverage computer vision for navigation and obstacle detection.

Tactile feedback, another application area, involves using materials like rubber and silicon to create sensors for detecting micro undulations and calibrating robotic hands. Computer vision is a dynamic and evolving field with broad applications across industries. Its history reflects a journey from mimicking human vision to leveraging advanced algorithms and deep learning. As technology progresses, computer vision is poised to play an increasingly pivotal role in diverse fields, offering solutions to complex problems and enhancing automation in various domains.

Multimodal models: a case study analysis

In the realm of machine learning (ML), the focus has historically been on models dealing with a single modality of data, such as language models, image classifiers, or speech recognition models. However, the future of ML holds even more promise with the emergence of multimodal models capable of handling diverse modalities simultaneously.

Rather than relying on individual models tailored to specific tasks or domains, the next generation of multimodal models aims to activate only the relevant model pathways for a given problem, allowing flexible handling of different modalities both as inputs and outputs. Building effective multimodal models involves addressing two key questions: how much modality-specific processing should be done before merging learned representations, and what is the most effective way to mix these representations.

Recent work on "*Multi-modal bottleneck transformers*" and "*Attention bottlenecks for multimodal fusion*" explores these trade-offs, revealing that merging modalities after a few layers of modality-specific processing and then mixing features through a bottleneck layer is particularly effective. This approach significantly improves accuracy in various video classification tasks by leveraging multiple modalities for decision-making.

Combining modalities not only enhances performance on multimodal tasks but can also improve accuracy on single-modality tasks. Notable examples include the DeViSE framework, which improves image classification accuracy by combining image and word-embedding representations, and Locked-image Tuning (LiT), a method that adds language understanding to pretrained image models, demonstrating substantial improvements in zero-shot image classification.

The utility of multimodal models extends to co-training on related modalities, such as images and videos, leading to enhanced accuracy on video action classification tasks. Additionally, the integration of language with other modalities, like vision, opens avenues for more natural human–computer interactions.

The "PaLI: scaling language-image learning" model, for instance, combines vision transformers with text-based transformers to achieve state-of-the-art results across various language-related tasks and benchmarks.

In applications like FindIt, a unified model for visual grounding, natural language questions about visual images can be answered through a general-purpose visual grounding model. This flexibility enables the model to handle different types of

queries related to grounding and detection. Another interesting application is video question answering, where multi-stream video inputs and text inputs are fused to produce text-based answers by iteratively co-tokenizing the video-language inputs.

In the domain of content creation, "VDTTS: visually-driven text-to-speech" explores a multimodal model for visually-driven text-to-speech tasks, making dialog replacement in videos more efficient. The model, trained on desired text and original video frames, generates synchronized speech output that matches the video, showcasing improvements in video-sync, speech quality, and speech pitch.

Moreover, the integration of multimodal models in Google Assistant, as demonstrated in "Look and talk: natural conversations with google assistant," enhances the naturalness of interactions. On-device multimodal models leverage both video and audio inputs to accurately determine user intent through visual and auditory cues, leading to more natural conversations.

Multimodal models are not limited to human-oriented modalities and have significant implications in real-world applications, particularly in autonomous vehicles and robotics. The fusion of sensor data from Lidar units and vehicle cameras in real-time, as shown in "4D-net for learning multimodal alignment for 3D and image inputs in time," improves accuracy in 3D object recognition. The ability to understand and combine data from different sensors provides better insights into the surrounding environment.

In essence, the advent of multimodal models represents a paradigm shift in machine learning, enabling single models to understand and generate outputs across various modalities fluidly and contextually. This development holds immense potential for diverse applications in Google products, platforms, and promises advancements in fields such as health, science, creativity, robotics, and beyond. Figure 4 provides an overview illustration toward the matter of perspective.



Fig. 4 An overview illustration for Multimodal Models in action

Generative models: an investigative analysis

In 2022, the field of generative models for imagery, video, and audio witnessed remarkable advancements, showcasing extraordinary capabilities. A variety of approaches have been explored, with generative adversarial networks (GANs) being a pioneering model introduced in 2014. GANs involve a generator creating realistic images and a discriminator distinguishing between generated and real images. Over the past decade, these models have evolved significantly, as illustrated in the provided images showing the progress in generative image model capabilities. Diffusion models, introduced in 2015, follow an iterative forward diffusion process to systematically destroy structure in a data distribution. They then learn a reverse diffusion process to restore lost structure, offering controllable generative capabilities.

Autoregressive models, such as PixelRNN, PixelCNN, VQ-VAE, and Image Transformer, have played a crucial role in pixel-level generation using deep neural networks. While these models initially produced relatively low-quality images, recent advances, including CLIP pretraining, larger language model encoders, and more extensive training datasets, have elevated the quality of generated images significantly. Two notable generative models from Google Research, Imagen, and Parti have made substantial contributions. Imagen leverages a large language model pretrained on text-only corpora for effective encoding of text for image synthesis. Imagen introduces new architectural advancements, including efficient U-Net and classifier-free diffusion guidance, enhancing performance.

Parti, utilizing an autoregressive Transformer architecture, generates image pixels based on text input, achieving improved results by scaling the Transformer encoder-decoder to 20 billion parameters. User control in generative models has become a focus, enabling users to fine-tune models like Imagen or Parti for subject-driven image generation. DreamBooth allows users to combine text and input images for more personalized generation, while other methods like "Prompt-to-prompt image editing with cross attention control" and Imagen Editor enable iterative editing of generated images using text prompts. The evolution of generative models extends beyond still images to generative video. Imagen Video and Phenaki tackle the challenge of generating high-resolution, temporally consistent videos with controllability. Imagen Video utilizes cascaded diffusion models, while Phenaki introduces a Transformer-based model for variable-length video generation. Combining these models offers the potential to benefit from high-resolution frames and long-form videos. Generative models for audio have also seen significant progress. The AudioLM approach leverages language modeling for audio generation without relying on annotated data. By separating the audio generation process into coarse semantic tokens and fine-grained audio tokens, AudioLM produces syntactically and semantically plausible speech and coherent piano music continuations. The model maintains speaker identity and prosody for unseen speakers in speech generation.

The landscape of generative models in 2022–2024 showcases substantial advancements across imagery, video, and audio domains, paving the way for more realistic and controllable generation processes. These developments open new possibilities for creative applications and user-driven content generation. How generative AI enables next level research and future prospect in terms of usage and performance Fig. 5 provides an insight to better understand the significance of the matter.



Fig. 5 An overview illustration for generative models in action

Results and findings

In the research experimental simulation investigation model "*Pix2Seq: a language modeling framework for object detection*," a novel approach to object detection is introduced, departing from traditional task-specific methods. The Pix2Seq framework treats object detection as a language modeling task, where the model is trained to "read out" the locations and attributes of objects in an image based on pixel inputs.

This method proves to be competitive on the COCO dataset, a large-scale object detection benchmark, showcasing comparable performance to existing specialized and optimized detection algorithms. Notably, Pix2Seq's performance can be further enhanced through pretraining on a more extensive object detection dataset. Addressing the challenge of understanding the 3D structure of real-world objects from 2D images, several approaches are explored.

In "Large motion frame interpolation," the creation of short slow-motion videos is demonstrated by interpolating between images taken seconds apart. "View synthesis with transformers" combines light field neural rendering (LFNR) and generalizable patch-based neural rendering (GPNR) to achieve high-quality view synthesis of novel scenes from just a few images.

LFNR accurately reproduces view-dependent effects, and GPNR extends this capability to generalize across different scenes, enabling the synthesis of views for new scenes. Taking a step further in "LOLNerf: learn from one look," the focus shifts to learning a high-quality representation from a single 2D image.

By training on various examples of specific object categories, such as multiple images of different cats, the model gains enough knowledge about the expected 3D structure to create a 3D model from a single image of a novel category. This approach, exemplified in LOLCats, demonstrates the potential for learning robust 3D representations from minimal visual input.

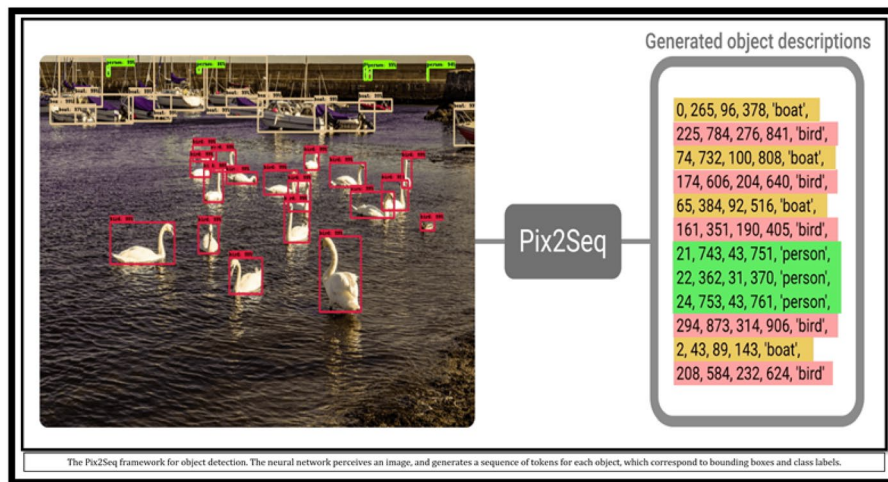


Fig. 6 The research findings for generative AI models in experimentation 1

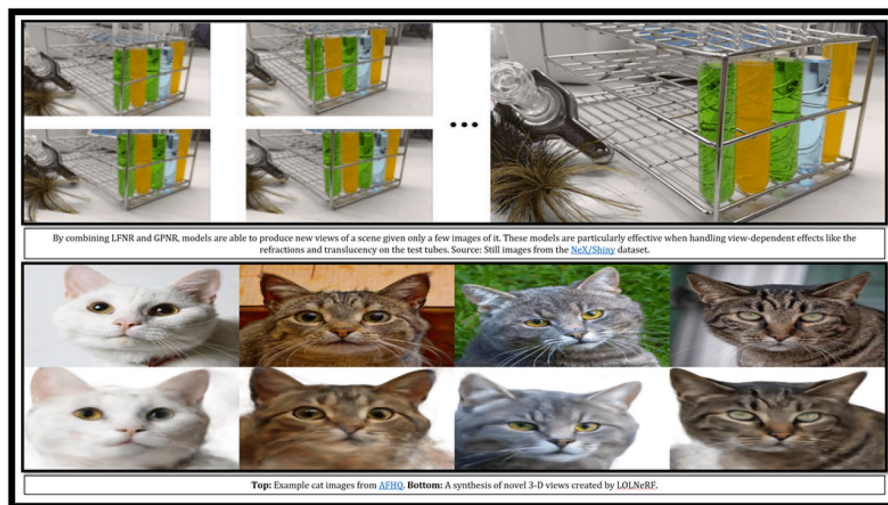


Fig. 7 The research findings for generative AI models in experimentation 2

The overarching objective of this body of work is to advance techniques that empower computers to better comprehend the 3D world, reflecting a longstanding aspiration within the field of computer vision. These advancements hold promise for a wide range of applications, from object detection to scene understanding and 3D model synthesis, contributing to the broader goal of enhancing artificial intelligence’s understanding of the visual environment. To provide a better understanding, the visualizations in Fig. 6, 7, 8, 9, 10 illustrate the matter of perspective approached

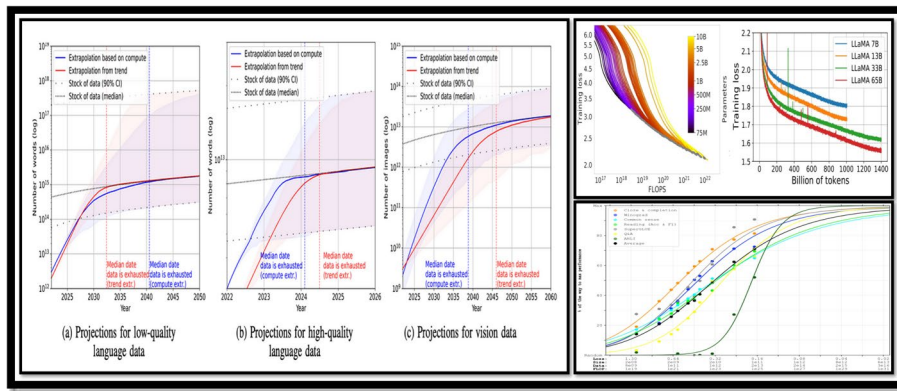


Fig. 8 The research findings for generative AI models in experimentation 3

PROVIDER	MODEL	DATASET								
		Banking77	Civil Comments	LEDGAR	Walmart Amazon	Company match	SQuAD2.0	SciQ	CONLL 2003	Average
	gpt-4	0.843	0.88	0.78	0.98	0.965	0.801	0.969	0.856	0.884
	gpt-3.5-turbo	0.776	0.785	0.734	0.962	0.85	0.69	0.944	0.762	0.813
	claude-v1	0.828	0.854	0.715	0.922	0.7	0.759	0.955	0.746	0.81
	text-davinci-003	0.806	0.89	0.673	0.777	0.72	0.586	0.913	0.774	0.767
	text-bison@001	0.775	0.87	0.558	0.949	0.946	0.781	0.947	0.623	0.806
	flan-t5-xxl	0.792	0.88	0.713	0.951	0.805	0.818	0.839	0.419	0.777
	Human annotators	0.77	-	-	0.966	-	-	-	0.849	0.862

Label quality (% agreement with ground truth labels) across a variety of NLP tasks. Highest numbers in each column are highlighted in green

Fig. 9 The research findings for generative AI models in experimentation 4

Model	LAMBADA (OpenAI)	HellaSwag	PIQA	ARC-Easy	ARC-Challenge	BoolQ	COPA	Winograd	Winogrande	TriviaQA	Jeopardy	MMIU
LLaMA-7B	0.703	0.761	0.799	0.813	0.284	0.760	0.835	0.878	0.685	0.343	0.308	0.296
LLaMA-13B	0.738	0.751	0.792	0.852	0.45	0.787	0.779	0.887	0.875	0.441	0.234	0.302
LLaMA-33B	0.533	0.48	0.666	0.435	0.259	0.606	0.672	0.646	0.53	0.649	0.000	0.23
Pythia-7B	0.687	0.638	0.761	0.581	0.325	0.634	0.769	0.785	0.607	0.36	0.022	0.265
Pythia-13B	0.704	0.672	0.768	0.605	0.381	0.675	0.781	0.847	0.627	0.233	0.026	0.253
GPT-J-6B	0.683	0.665	0.762	0.583	0.355	0.648	0.789	0.833	0.641	0.234	0.026	0.23
GPT-NeoX-20B	0.719	0.712	0.780	0.644	0.392	0.681	0.781	0.881	0.686	0.347	0.146	0.289
Cerebras-7B	0.638	0.582	0.744	0.564	0.31	0.625	0.734	0.779	0.603	0.141	0.012	0.259
Cerebras-18B	0.635	0.588	0.740	0.571	0.321	0.61	0.716	0.780	0.602	0.146	0.013	0.258
OPT-7B	0.677	0.676	0.773	0.579	0.329	0.685	0.719	0.840	0.656	0.227	0.020	0.23
OPT-13B	0.692	0.701	0.774	0.586	0.345	0.657	0.805	0.851	0.670	0.282	0.106	0.257

Model	GPT-3	BLOOM	LLaMA	LLaMA-2	T5	PaLM
Developer						
Model Size (Parameters)	175B	176B	70, 130, 330, 65B	70, 130, 330, 70B	118	540B
Training Data (Tokens)	300B	300B	1.4T	2T	34B	769B
Training Compute (FLOPs)	3.2E+23	3.7E+23	3.9E+23	1.9E+24	2.2E+21	2.6E+24
Processor	Manufacturer: Nvidia	Nvidia	Nvidia	Nvidia	Google	Google
Type	GPU	GPU	GPU	GPU	TPU	TPU
Model	V100	A100	A100	A100	TPU v3	TPU v4
Processor Price	3,532,000	1,082,960	1,770,354	3,311,616	245,760	6,604,992
Top Culture Industry Applications	0.429	0.857	0.383	0.423	0.345	0.079
Cost Center Efficiency (PUE)	1.1	1.2	1.1	1.1	1.12	1.08
Energy Consumption (MW)	1,287	520	779	1,400	86	3,438
Carbon Emissions (tCO2e)	552	30	300	593	47	271

$AWA = \text{Hours to train} \times \text{Number of Processors} \times \text{Average Power per Processor} \times PUE \times 1000$
 $ACOG = AWA \times \text{kg CO2e per AWA} \times 1000$

Fig. 10 The research findings for generative AI models in experimentation 5

rendered for the generative AI model framework experimentation in terms of performance action and data analytics.

Discussions and future directions

In 2022–2024, significant strides were made in the field of generative models, enabling computers to interact with natural language more effectively and understand users' creative processes. This has ushered in new possibilities for computers to assist users in creating images, videos, and audio in ways that surpass the capabilities of traditional tools. The focus on advancing text-to-image and text-to-video capabilities, exemplified by initiatives like Dream Booth, reflects a growing interest in empowering users to control the generative process. The coming years are anticipated to witness further advancements in the quality and speed of media generation, accompanied by novel user experiences that enhance creative expression. While these creative tools hold tremendous potential, it is crucial to acknowledge and address associated concerns [42].

The ability of generative models to produce content raises ethical considerations, such as the potential to generate harmful or misleading information, fake imagery, or realistic audio content that blurs the line between reality and fabrication. Responsible deployment of these models is paramount, and the challenges of mitigating risks and ensuring ethical use are areas of active consideration and development. Responsible AI practices play a central role in addressing these challenges. Google, as a leader in machine learning (ML) and AI, emphasizes the importance of responsible AI pursuits. The company's AI Principles prioritize beneficial use, user well-being, safety, and the avoidance of harms.

These principles guide the entire AI process from research priorities to product development and deployment. Google applies the scientific method rigorously in AI research and development, incorporating peer review, readiness reviews, and responsible access and externalization. Collaboration with multidisciplinary experts, including social scientists and ethicists, is integral to Google's responsible AI approach. The company emphasizes continuous learning and improvement based on feedback from developers, users, governments, and affected communities [43–50].

Regular reviews of AI research and development, transparent reporting of findings, and staying abreast of emerging concerns and risks are key components of Google's commitment to responsible AI. The other AI companies also take a leadership role in shaping responsible governance, accountability, and regulation in the AI space. This involves contributing to the development of policies that encourage innovation while effectively managing the risks associated with AI technologies. Moreover, AI companies are dedicated to fostering public understanding of AI by providing clear information on what AI is, its potential benefits, and how users and society can leverage its capabilities responsibly.

In a forthcoming future research exploration, leaders from Google's Responsible AI team will delve deeper into the work undertaken in 2022–2024 for AI research, offering detailed insights and outlining their vision for the field in the coming years. This continued commitment to responsible AI underscores Google's and other AI tech giants' dedication to ethical and impactful contributions in the realm of artificial intelligence.

Conclusions

The concluding thoughts concerning the transformative advances achieved in various domains, with a particular focus on enhancing Google products and platforms to benefit billions of users. These advancements span across products such as Search, Assistant, Ads, Cloud, Gmail, Maps, YouTube, Workspace, Android, Pixel, Nest, and Translate.

The application of cutting-edge technologies, including Transformer models and sequence-to-sequence learning in language models, has enabled natural conversations with computers, providing surprisingly good responses. In computer vision, innovations allow users to create and interact in three dimensions, marking a shift from traditional 2D interactions. Generative models have opened up new possibilities for users to create images, videos, and audio in ways previously unattainable with conventional tools.

A significant transformation arises from the increasing capabilities of multimodal models, aiming to create a unified model proficient in understanding various modalities and generating diverse modes in context. Google has introduced a unified language model capable of performing vision, language, question answering, and object detection tasks across more than 100 languages. The goal is to enable users to engage multiple senses when interacting with computers, enhancing the naturalness of the interaction. On-device multimodal models have already demonstrated improved interactions with Google Assistant, and ongoing advancements promise more exciting developments in this space.

The emphasis on responsible AI practices is reiterated, acknowledging the need to ensure the safety of state-of-the-art technologies before their broad release. Tech giants like Google is committed to adhering to its AI Principles, prioritizing user safety and societal well-being. The narrative underscores the ongoing mission of organizing the world's information and making it universally accessible and useful, indicating that, even after two decades, this mission remains as bold as ever.

The excitement lies in how these AI advances are applied to enhance user experiences, enabling more people to better understand the world and accomplish tasks efficiently. The closing remarks emphasize a vision of computers playing a pivotal role in achieving these transformative goals.

The discussions and future directions throughout this research have provided a comprehensive overview of the groundbreaking advancements and research breakthroughs in the field of artificial intelligence, particularly within the domains of natural language processing, computer vision, and generative models. The exploration of language models, exemplified by the Transformer architecture, has showcased how AI systems, such as Bard, GPT-3.5, and GPT-4 can engage in natural and context-aware conversations, demonstrating remarkable proficiency in understanding and generating human-like responses. The strides in computer vision have unveiled innovative approaches to object detection and 3D understanding of real-world scenes. Notably, Pix2Seq presented a novel perspective on object detection by framing it as a language modeling task, achieving competitive results on the COCO dataset.

The challenges of grasping 3D structures from 2D images were addressed through techniques like Large Motion Frame Interpolation and View Synthesis with Transformers,

providing insights into generating realistic views of scenes from limited visual input. Generative models, a focal point of the conversation, showcased remarkable progress in imagery, video, and audio generation. From the early days of generative adversarial networks (GANs) to recent advancements like Imagen and Parti, the capacity to create high-resolution, detailed, and contextually meaningful content has significantly evolved. The integration of language models, such as CLIP, demonstrated the potential to control and guide generative processes effectively, bridging the gap between textual prompts and visually rich outputs.

The narrative extended to the exciting realm of generative video and audio, where Imagen Video and Phenaki presented solutions for high-resolution video generation and audio synthesis, respectively. These developments reflect the ongoing efforts to extend generative capabilities into multimodal experiences, enabling a more holistic and immersive interaction with AI systems.

The overarching theme of responsibility in AI development was emphasized, acknowledging the potential risks associated with misinformation, toxicity, and the generation of harmful content. Google's commitment to responsible AI practices, as outlined in its principles, underscores the importance of ethical considerations in deploying these advanced technologies.

Looking forward, the exploration anticipates continued progress in media generation, user control over generative processes, and addressing challenges related to responsible AI deployment. The intersection of language understanding, computer vision, and generative models holds the promise of transformative user experiences, aligning with Google's, OpenAI, Meta and other tech giants of AI, their mission to organize the world's information and make it universally accessible and useful. The future trajectory, guided by both innovation and responsibility, highlights the ongoing evolution of AI technologies and their profound impact on human-computer interactions.

Acknowledgements

The idea representation with the research focuses along with the context concerning the investigative exploration and manuscript writing was done by the author himself. All the datasets, data models, data materials, data information, computing toolsets used, and retrieved for the conduction concerning this research are mentioned within the manuscript and acknowledged with its associated references where appropriate.

Author contributions

Described in details within the Acknowledgements section.

Funding

No funding was provided for the conduction of this research.

Availability of data and materials

The various original data models and datasets of which are not all publicly available, because they contain various private information. The available platform provided datasets and data models that support the findings and information of the research investigations are referenced where appropriate.

Code availability

Mentioned in details within the Acknowledgements section.

Declarations

Ethics approval and consent to participate

The author has read and approved the manuscript and has agreed to its publication.

Competing interests

There is no conflict of interest or any type of competing interests for this research.

Received: 2 January 2024 Accepted: 29 May 2024
Published: 12 June 2024

References

1. ChatGPT – Release Notes. Archived from the original on January 12, 2024. Retrieved January 16, 2024.
2. A history of generative AI: from GAN to GPT-4. March 21, 2023.
3. Lock S (2022) What is AI chatbot phenomenon ChatGPT and could it replace humans? *The Guardian*. Archived from the original on January 16, 2023
4. Chui M, Kamalnath V, McCarthy B (2018) *An executive's guide to AI*. McKinsey & Company, New York
5. Griffith E, Metz C (2023) Anthropic said to be closing in on \$300 million in new AI funding. *The New York Times*, New York
6. Simon FM, Altay S, Mercier H (2023) Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinf Rev* <https://doi.org/10.37016/mr-2020-127>
7. Metz C (2023) Open AI plans to up the ante in tech's AI race. *The New York Times*, New York
8. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, Nori H, Palangi H, Ribeiro MT, Zhang Y (2023). Sparks of artificial general intelligence: early experiments with GPT-4. [arXiv:2303.12712](https://arxiv.org/abs/2303.12712) [cs.CL].
9. Weise K, Metz C, Grant N, Isaac M (2023) Inside the AI arms race that changed silicon valley forever. *The New York Times*, New York
10. Metz C, Mickle T (2024) OPENAI COMPLETES DEAL THAT VALUES THE COMPANY AT \$80 BILLION. *The New York Times*, New York
11. Chui M, Hall B, Mayhew H, Singla A, Sukharevsky A, McKinsey AI (2022) The state of AI in 2022- and a half decade in review. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review/>
12. Chui M, Roberts R, Yee L (2022) *McKinsey technology trends outlook 2022*. McKinsey & Company, New York
13. Chui M, Manyika J, Miremadi M (2018) What AI can and can't do (yet) for your business. *McKinsey Quarterly* 1(97–108):1
14. Newsom G, Weber SN (2023) Executive order N-12–23 (PDF). Executive Department, State of California, California
15. Lanxon N, Bass D, Davalos J (2023) A cheat sheet to AI buzzwords and their meanings. *Bloomberg News*, New York
16. Karpathy A, Abbeel P, Brockman G, Chen P, Cheung V, Duan Y, Goodfellow I, Kingma D, Ho J, Rein H, Tim S, John S, Ilya S, Wojciech Z (2016) Generative models. *OpenAI*
17. Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A (2022) LaMDA: language models for dialog applications. [arXiv:2201.08239](https://arxiv.org/abs/2201.08239) [cs.CL].
18. Roose K (2022) A coming-out party for generative AI. *Silicon Valley's New Craze*. *The New York Times*, New York
19. Don't fear an AI-induced jobs apocalypse just yet. *The Economist*, New York
20. Eapen T, Finkenstadt DJ, Folk J, Venkataswamy L (2023) How generative AI can augment human creativity. *Harvard Bus Rev* 101(4):16
21. The race of the AI labs heats up. *The Economist*. January 30, 2023. Retrieved March 14, 2023.
22. Yang J, Gokturk B (2023) Google cloud brings generative AI to developers, businesses, and governments.
23. Hendrix J (2023) Transcript: senate judiciary subcommittee hearing on oversight of AI. *Techpolicy Press*, Austin
24. SITTFlash (2017) The history of artificial intelligence. *Science in the News*, Washington, Dc
25. Bergen N, Huang A (2023) A brief history of generative AI (PDF). *Dichotomies: generative AI: Navigating Towards a Better Future* 2(4).
26. Cao Y, Li S, Liu Y, Yan Z, Dai Y, Yu PS, Sun L (2023) A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT. [arXiv:2303.04226](https://arxiv.org/abs/2303.04226) [cs.AI].
27. finetune-transformer-lm. *GitHub*. Retrieved May 19, 2023.
28. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners. *OpenAI Blog* 1(8):9
29. Schlagwein D, Willcocks L (2023) Chatgpt et al: the ethics of using (generative) artificial intelligence in research and science. *J Inf Technol* 38(2): 232–238
30. Explainer: what is generative AI, the technology behind OpenAI's ChatGPT? *Reuters*. March 17, 2023. Retrieved March 17, 2023.
31. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, Brynjolfsson E (2021) On the opportunities and risks of foundation models. [arXiv:2108.07258](https://arxiv.org/abs/2108.07258) [cs.LG].
32. Chen M, Tworek J, Jun H, Yuan Q, Pinto HPDO, Kaplan J, Edwards H, Burda Y, Joseph N, Brockman G Ray A (2021) Evaluating large language models trained on code. [arXiv:2107.03374](https://arxiv.org/abs/2107.03374) [cs.LG].
33. Epstein Z, Hertzmann A, Akten M, Farid H, Fjeld J, Frank MR, Groh M, Herman L, Leach N, Mahariq R, Pentland AS, Rusakovsky O, Schroeder H, Smith A (2023) Art and the science of generative AI. *Science* 380(6650):1110–1111
34. Nellis S, Lee J (2022). U.S. officials order Nvidia to halt sales of top AI chips to China. *Reuters*.
35. OpenAI API. platform.openai.com. Archived from the original on March 3, 2023. Retrieved March 3, 2023.
36. OpenAI (2022). ChatGPT: Optimizing Language Models for Dialogue. Archived from the original on November 30, 2022. Retrieved December 5, 2022.
37. What's the next word in large language models? *Nat Mach Intell* 5(4): 331–332
38. What is ChatGPT and why does it matter? Here's what you need to know. *ZDNET*. May 30, 2023. Archived from the original on February 15, 2023. Retrieved June 22, 2023.
39. Akhtar ZB (2024) The design approach of an artificial intelligent (AI) medical system based on electronic health records (EHR) and priority segmentations. *J Eng* 2024:1–10. <https://doi.org/10.1049/tje2.12381>

40. Akhtar ZB, Gupta AD (2024) Integrative approaches for advancing organoid engineering: from mechanobiology to personalized therapeutics. *J Appl Artif Intell* 5(1):1–27
41. Akhtar ZB, Gupta AD (2024) Advancements within molecular engineering for regenerative medicine and biomedical applications an investigation analysis towards a computing retrospective. *J Electron Electromed Eng Med Inform* 6(1):54–72
42. Akhtar Z (2024) Securing operating systems (OS): a comprehensive approach to security with best practices and techniques. *Int J Adv Netw Monit Controls* 9(1):100–111. <https://doi.org/10.2478/ijanmc-2024-0010>
43. Pinaya WHL, Graham MS, Kerfoot E, Tudosiu PD, Dafflon J, Fernandez V, Sanchez P, Wolleb J, da Costa PF, Patel A (2023) Generative AI for medical imaging: extending the MONAI framework. [arXiv:2307.15208](https://arxiv.org/abs/2307.15208) [eess.IV].
44. Pasick A (2023) Artificial intelligence glossary: neural networks and other terms explained. *The New York Times*, New York
45. Douglas W (2023). The inside story of how ChatGPT was built from the people who made it. *MIT Technology Review*. Archived from the original on March 3, 2023. Retrieved March 6, 2023.
46. Vincent J (2022). ChatGPT proves AI is finally mainstream – and things are only going to get weirder. *The Verge*. Archived from the original on January 11, 2023. Retrieved December 8, 2022.
47. Roth E (2023). Microsoft spent hundreds of millions of dollars on a ChatGPT supercomputer. *The Verge*. Archived from the original on March 30, 2023. Retrieved March 30, 2023.
48. Press Center - TrendForce Says with Cloud Companies Initiating AI Arms Race, GPU Demand from ChatGPT Could Reach 30,000 Chips as It Readies for Commercialization | TrendForce - Market research, price trend of DRAM, NAND Flash, LEDs, TFT-LCD and green energy, PV. TrendForce. Archived from the original on November 2, 2023. Retrieved November 2, 2023.
49. Badawy M, Ramadan N (2023) Hefny HA (2023) Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *J Electr Syst Inf Technol* 10:40. <https://doi.org/10.1186/s43067-023-00108-y>
50. Abdalla PA, Mohammed BA (2023) Saeed AM (2023) The impact of image augmentation techniques of MRI patients in deep transfer learning networks for brain tumor detection. *J Electr Syst Inf Technol* 10:51. <https://doi.org/10.1186/s43067-023-00119-9>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.