Journal of Computer Science

Review

A Survey of Generative AI Applications

Roberto Gozalo-Brizuela and Eduardo Eduardo Garrido Merchan

Department of Quantitative Methods, Universidad Pontificia Comillas, Madrid, Spain

Article history
Received: 22-09-2023
Revised: 09-01-2024
Accepted: 05-03-2024

Corresponding Author:
Eduardo Garrido Merchan
Department of Quantitative Methods, Universidad Pontificia Comillas, Madrid, Spain
Email: ecgarrido@comillas.edu

Abstract: Generative AI has experienced remarkable growth in recent years, leading to a wide array of applications across diverse domains. In this study, we present a comprehensive survey of more than 350 generative AI applications, providing a structured taxonomy and concise descriptions of various unimodal and even multimodal generative AIs. The survey is organized into sections, covering a wide range of unimodal generative AI applications such as text (ChatGPT or Bard), images (DALL-E 3 or Midjourney), video (Colossyan AI and Synthesia AI), gaming (CSM and ConvAI) and brain (Meta's speech from brain) information. Our survey provides an updated outlook of the generative AGI landscape, giving several examples of each technology, something not achieved in a paper before to our knowledge. Other surveys have focused on reviewing the impact of generative on different industries while not diving deep into which technologies or models to use. A thorough web and industry analysis was conducted to get examples for each of the sections. Ethical considerations were followed for the composition of this study.

Keywords: Artificial Intelligence, AI, Generative AI, Machine Learning and Deep Learning

Introduction

Generative AI technologies are supposed a turning point in technological development. Artificial intelligence has traditionally been used for classification tasks such as fraud detection, customer segmentation, and risk assessment very successfully. The breakthrough with Generative AI is the ability to generate new output differing from training data and enabling creativity (Feuerriegel *et al*., 2024). Generative AI refers to artificial intelligence that can generate novel content, rather than simply analyzing or acting on existing data like expert systems (Murphy, 2022). Generative AI models, equipped with vast data sets and intricate designs, have the extraordinary capability to create new and diverse content. They can process and learn from information gathered from a multitude of sources, such as Wikipedia (Schick *et al*., 2022), Github (Chen *et al*., 2021), and others. By tapping into this wealth of data, these models can generate an extensive range of multimedia formats, including video, audio, and images. Our survey taps into the extensiveness of this technology by researching the different Generative AI models that have been created and categorizing them.

Groundbreaking Generative AI models include ChatGPT and DALL-E (Ramesh *et al*., 2022), which have catalyzed a new era in the synthesis and manipulation of digital content. The use and combination of these technologies have allowed for use cases throughout many different fields and applications which will be explored throughout this piece. Businesses' productivity has already been boosted by Generative AI, with research finding that 15% of tasks could be achieved significantly faster through the use of AI (Eloundou *et al*., 2023).

During recent years, the continuous growth in computing power has used deep neural networks, transformers, and other innovative models like generative adversarial networks (Creswell *et al*., 2018) and variational autoencoders (Murphy, 2022). All these models can effectively capture the complexity of data, making them adept at modeling high-dimensional probability distributions of language or images from specific or general domains. By complementing generative models with additional techniques that map the latent high-dimensional semantic space of language or images to multimedia representations of text, audio, or video, it becomes possible to transform any input format, such as text, into a variety of output formats like video. This versatility allows for a seamless conversion between multimedia formats, making generative models invaluable in numerous applications. One of the most significant aspects of generative AI is its potential for

endless applications. These models can be trained to generate genuinely different multimedia formats, like video, audio, or text, from various input formats. For instance, generative AI can be used to create realistic images from textual descriptions, produce video content from audio, or even generate music compositions based on specific styles or emotions. Furthermore, generative AI has the potential to revolutionize industries such as advertising, entertainment, and education by automating content creation and providing personalized experiences. With the ability to learn from diverse data sources and generate a wide array of multimedia outputs, these models can help businesses and individuals alike save time and resources while tapping into new creative possibilities. In conclusion, generative AI models, bolstered by their access to extensive data and complex designs, offer unparalleled potential in content creation and transformation. Their ability to learn from various sources, generate diverse multimedia formats, and convert inputs from one format to another opens up a vast array of applications in multimedia generation and conversion, making them indispensable tools in today's technology-driven world.

In more recent work, there have been surveys of LLMs and Generative AI, talking about different applications of the technology (Zhao *et al*., 2023; Cao *et al*., 2023; Zhang *et al*., 2023a-c; Aydın and Karaarslan, 2023). Existing surveys have focused on reviewing specific types of generative AI technologies such as text-to-image (Zhang *et al*., 2023a-c). Other surveys such as (Cao *et al*., 2023) have focused on reviewing the impact of generative A on different industries while not diving deep into which technologies or models to use. In contrast to prior surveys, this comprehensive review aims to offer a unique perspective by highlighting not only the most prominent generative models and their underlying technologies but also emphasizing all the different uses of this technology. In addition, we give an up-to-date competitive outlook in this growing industry and the models behind this growth.

This resource encompasses 13 categories, which include text, images, video, 3D, code and software, speech, AI understanding, business, gaming, music, biotechnology, brain, multimodal and computational efficiency. Within each section, a thorough taxonomy of the current technologies is presented, detailing both the models and tools available. By offering a systematic exploration of these diverse AI applications, the survey serves as an essential reference for researchers, academics, and professionals, enabling them to comprehend better the evolving landscape of generative AI and its far-reaching implications.

As an example, a 3D game designer may have various generative AI needs for a project of his. He may find a solution for his 3D AI needs under both 3D and gaming, getting more specific results and different answers. He may also find solutions for more business needs of his under both business and text. With this survey, we believe that users will get a very good outlook on how generative AI is shaping up and where they may find their needed technology.

We introduce, in this article, the proposition of an extensive survey centered on the most sought-after generative AI applications, which are notably reshaping industries such as the videogames (Liu *et al*., 2021), design (Kulkarni *et al*., 2023), and business operations sectors. The challenge users experience in identifying the developed programs within each distinct application field substantiates the demand for a comprehensive reference tool that categorizes models.

To ensure the inclusivity and comprehensiveness of the study, extensive online research was conducted to gather information on generative Artificial Intelligence (AI) applications. The investigation involved a meticulous exploration of various sources, including companies' official websites, where detailed documentation and technical specifications were scrutinized. Additionally, an examination of companies' social media platforms was undertaken to extract valuable insights into real-world applications, user feedback, and updates related to generative AI technologies. This dual-pronged approach aimed to capture both official, structured information and dynamic, community-driven perspectives surrounding these applications.

Subsequently, the gathered information underwent a rigorous categorization process. The categorization framework was developed based on distinct features, functionalities, and underlying algorithms identified during the review of online sources. Each category was refined through an iterative process, staying attuned to emerging trends in the generative AI landscape. By synthesizing information from diverse online outlets, including both official channels and community discussions, the categorization framework not only enriched the dataset but also provided a holistic view of the practical implementations and evolving trends in the field. This synthesis contributes to the robustness of the study, enhancing the overall validity and relevance of the findings presented in this academic exploration.

Basic Taxonomy of Models

This article explores the burgeoning applications of generative AI, focusing on its transformative potential across diverse sectors, such as art, business, biotechnology, and design. We achieve this by dividing generative AI into 13 parts by both the output that is produced, the context in which they are being used, and the business use that this technology has. The reader may observe that many models could be placed in the text category as the output is text. Or that many copy-writing models could also be placed under the text category. The

classification that is shown below serves for a prospective user of generative AI technologies to quickly find the technology they shall be using based on the use case. In this first part, we introduce the categories in which we taxonomized current Generative AI technology. Here we present a summary of the different categories: Regarding text, generative AI technologies in the text category aim to create and manipulate natural language text. These technologies include language models that can generate human-like text, such as OpenAI's GPT models. While the most famous of these models are chatbots such as OpenAI's ChatGPT or Google's BARD, other types of models are included in this category. These include text-writing assistants, scientific language models, or chatbots. The main criterion for this category was that the models produce text as an output. Concerning images, Generative AI technologies in the images category focus on the creation and manipulation of visual images. The main criterion in this category was that the final output was an image. This can include image-creating models that can create images out of textual descriptions and image-editing models. For simplicity, the category was divided into artistic image creation, realistic image creation, and image editing. Some models that did two or more of these tasks are arbitrarily included in one of the categories. Other models of this category include text-to-layouts and text-to-molecular representations which could not be included in the aforementioned categories. Dealing with video, generative AI technologies in the video category aims to create and manipulate video content. The main criterion for this category was that the final output was a video. This mainly includes video creation models that can generate new video content from textual descriptions. Other models include post-production, text-to-scene generation, text-to-motion capture, image-to-video as well as video dubbing. Also working with 3D, Generative AI technologies in the 3D category focuses on the creation and manipulation of three-dimensional objects and environments. The main criterion was using the output being a fully formed 3D model. Also included, there is a 4-D model and a 3-D model specially designed for metaverse purposes. Inputs include text, a single image, images, and 2D models. Focusing on code and software, Generative AI technologies in the code and software category aim to automate the process of writing code and creating software. The main criterion was that the final output be a code. This includes a variety of categories: Text-to-code, text-to-websites, text-to-software and text-to-apps. Other less frequent models include designs-to-code, text-to-software, text-to-RPA, and a code translator. The text-to-software category was designed to fit Adept, a company that wants users to communicate with computers via just a text input. This is why it is included here. Regarding speech, generative AI technologies in the speech category focus on the creation and manipulation of spoken language. All of these technologies are able to transform an input into a speech output. This is divided into text-to-speech, speech-to-speech, and speech editing. Regarding AI understanding, generative AI technologies in the AI understanding category are those models that convert an input into a text output. This particular category was drawn because of the need for a category that would summarize models that can turn many inputs into speech. Inputs cover: Speech, images audio and video, images, video, metaphors, semi-structured data, structured data, movies, and generative regions. Concerning business, generative AI technologies in the business category focus on the application of AI to improve business processes and decision-making. Many of the said models in the aforementioned categories such as ChatGPT in the text category or Midjourney in the image category could very well be used by businesses. Despite this, the means for this category is for people in general businesses to find models for their operations. These are divided into Marketing, new business models, and business operations. Dealing with gaming, generative AI technologies in the gaming category aims to make game creation much easier for developers. They use text, 3D, and image models for their purposes. They are divided into video game creation and characters. About music, generative AI technologies in the music category focus on the creation and manipulation of musical content. This category includes music generation, musical editing, and a dance-to-music model. Regarding biotechnology, generative AI technologies in the biotechnology category aim to apply generative AI to biological research and medical applications. This can include models that can predict the structure of proteins or DNA sequences, as well as drug discovery tools that can identify new drug candidates. Some of these models could have been included in the Business category, but this category was drawn up because of the abundance of generative AI applications in this field. Concerning the human brain, Generative AI technologies in the brain category focus on the application of generative AI to help people communicate. This includes brain-to-text models and brain-to-images models. For multimodal technologies, a section was included in order to fit those models that combined several of the other categories, such as ChatGPT-4, that are able to produce both text and image outputs. Finally, we include a computational efficiency category, made in order to fit Alphatensor, a groundbreaking AI technology for the discovery of algorithms (Kauers and Moosbauer, 2022).

This systematic categorization is anticipated to hold significant value in the future. By offering a structured exploration of diverse generative AI applications, the survey serves as an essential reference for researchers, academics, and professionals, enabling them to comprehend better the evolving landscape of generative

AI and its far-reaching implications. For instance, a 3D game designer seeking generative AI solutions for a project may benefit from the categorized information, obtaining specific results under both 3D and gaming categories. Similarly, users with varied business needs can explore solutions under both business and text categories. This survey aims to provide users with a comprehensive understanding of how Generative AI is shaping up, assisting them in identifying the most suitable models for their specific projects in the future.

Generative AI Applications

In this section, we will introduce a broad overview of generative AI applications divided into subsections according to every different topic.

Text

Text models especially those centered around conversational chatbots have revolutionized AI since the launch of ChatGPT. Helped by natural language processing and large language models, these models have many very useful capabilities like summarization, writing assistance, code generation, language translation, and sentiment analysis. They have been the main focus in Generative AI because of the capabilities of ChatGPT, an application that millions of users are already taking advantage of in fact, ChatGPT is thought to increase productivity by 40% in basic writing tasks (Noy and Zhang, 2023), something very promising for this technology. GPT-2's output is statistically undistinguishable from human-written text (Kreps *et al*., 2022). If the earlier GPT model is indifferent to human output, this means the technology has the capacity to replace human output at least in some forms of writing.

Conversational AI Conversational AI has been one of the most talked-about topics in AI. These services act as chatbots capable of a wide variety of tasks converting text prompts into text outputs. The most famous example is ChatGPT, which now has integrated capabilities to look around the internet for information. Other models with updated information include Bing AI, Duck Assist, Google's BARD powered by LaMDA, the Beta version of ChatGPT, Metaphor and perplexity AI. Other chatbots that do not include updated information include Claude or Stanford Alpaca.

They are powered by Large Language Models or LLMs. Large Language Models (LLMs) refer to Transformer language models that contain hundreds of billions (or more) of parameters, which are trained on massive text data, such as GPT-3, Galactica, LLaMA, and PaLM (Zhao *et al*., 2023). Some of their capabilities are common-sense reasoning, mathematical reasoning, programming assistance, spatial reasoning, and text generation (Joublin *et al*., 2023; Tian *et al*., 2023; Frieder *et al*., 2024). In terms of business operations, there are many applications such as demand forecasting, inventory optimization, and risk management (Bahrini *et al*., 2023). Many of the capabilities are being researched at the time of writing these articles, just as the capabilities of LLMs are being discovered.

Text-to-Author Simulation The models have recently shown abilities to recreate certain styles of writing. Examples recently show LLMs being able to write as Daniel C. Dennett (Schwitzgebel *et al*., 2023) or as Lovecraft (Garrido-Merchán *et al*., 2023). Dennett's paper shows that experts on Dennett's work were successful at a 51% rate in distinguishing between the philosopher's work and the large language model's work. Lovecraft's paper shows that human readers without prior exposure to Lovecraft are unable to distinguish between texts written by the author and those written by ChatGPT. These are remarkable achievements and show the great capabilities of language models in imitating writing through finetuning.

Text-to-medical advice large language models have also been proven useful for preliminary medical advice through fine-tuning. We need to state that these models are still not completely safe for this use and they should not be used for replacing a human at the moment. Some of these models are Chatdoctor, GlassAI, Med-PaLM 2, (Singhal *et al*., 2023), and your doctor AI. They have shown promising capabilities to retrieve medical knowledge, reason over it and answer medical questions comparably to physicians. Med-PaLM 2 scored up to 86.5% on the MedQA dataset. These models again show a remarkable ability to create accurate responses through fine-tuning. The biggest startup in this space is Hippocratic AI (Hippocratic AI-hippocraticai.com, 2023) which has developed LLMs that outperform GPT-4 on medical datasets. At the moment, this is very significant as developments have already been found in fields that are able to directly impact human livelihood.

Text-to-Science other applications can be seen in science, where Galactica (Taylor *et al*., 2022) and Minerva (Lewkowycz *et al*., 2022) have emerged. Galactica is a large language model that can store, combine, and reason about scientific language. Minerva is a Large language model focused on quantitative reasoning tasks such as mathematics, science, and engineering problems at the college level. Although the models are not at all able to replace human reasoning on these tasks, they show promising results in a very challenging field of exploration. In fact, recent studies suggest these models.

Writing Assistance Generative AI can also be used for live writing assistance. Previously mentioned chatbots such as ChatGPT can be used for this purpose, but specific applications have been created such as GrammarlyGO (Grammarly, 2023) and PEER (Schick *et al*., 2022). GrammarlyGO is a writing assistant created by Grammarly, able to write drafts, outlines, replies, and

revisions. PEER is similar to Grammarly's software but it provides an explanation for its actions and it's fine-tuned for an academic article.

Images

Image generative AI has only grown since the launch of DALL-E 2 back in 2022. With both artistic and professional purposes, this technology has proved very useful in creating images from text prompts as well as image editing. In terms of art creation, it has pushed creative boundaries and has been revolutionary. In image creation, photorealism seems nearer with cutting-edge applications such as Midjourney which have provided very realistic images.

Artistic images in terms of artistic images, many platforms have been created for the creation of artistic images through text prompts. Some examples include creating images from text prompts, two platforms that are available for public use in order to use said models are Mage. Space which uses stable diffusion for art generation Night Cafe, which uses stable diffusion, and DALL-E 2 for artistic image generation. Other platforms include Wonder, which is a mobile application for artistic image creation and Neural. Love, is an AI-powered platform for audio, video, and image editing and enhancement which has an Art generator, from which you can select from many styles such as Fantasy or Sci-Fi. In contrast to these other platforms, DALL-E and Midjourney use their own for image generation.

These models have also been proven useful for other artistic image tasks. Tattoo creation can be helped via Tattoos AI. Moreover, meme creation is achieved through Supermeme AI. Also, artistic avatars can be generated through profile picture AI, using samples of yourself.

Image editing generative AI has proven useful in terms of image editing. Some useful applications include Alpaca AI, I2SB (Liu *et al*., 2023b), and Facet AI. Some capabilities of these applications are inpainting, outpainiting, upscaling, super-resolution, deblurring, and depth map generation. An example of an application using generative AI for image editing is Photoroom AI, which is able to erase backgrounds and remove objects in images through this software. Editing through natural language prompts can make photo editions much more accessible to non-professionals.

Even face restoration can be achieved through generative AI, as the Tencent Face Restoration tool shows. They achieve this through GANs, one of the pillars behind generative AI and deep learning. For the purpose of creativity, stable diffusion reimagining allows users to generate multiple variations of a single image.

Realistic images in terms of realistic image creation, there has been a plethora of models which enable realistic image generation. They include Bing AI Image Creator, Booth AI, Craiyon, DALLE 2, GLIGEN (Li *et al*., 2023a-b), Imagen, Midjourney, Muse (Chang *et al*., 2023), Parti (Yu *et al*., 2022), PrimeProfile, Runway ML text-to-image and stable diffusion ML. Through text inputs, they attempt photorealistic generations. Outside of simple text-to-image creation, there are many more uses for Generative AI. Through image samples, Generative AI can create photorealistic images. Aragon AI, Avatar AI, and PrimeProfile can create headshots through sample images.

The process of design can be optimized through generative AI. Play (Cheng *et al*., 2023) shows how text can be converted into layouts using latent diffusion. As well, Autodraw is a drawing-to-shapes model that turns simple drawings into shapes. Both of these applications can quickly optimize the design process.

Video

Video generative AI helps producers with storytelling. Although still a developing field because of the complexity that video generation poses, listed use cases such as digital human videos, human motion capture, and video dubbing are revolutionary uses that can quickly lead to productivity improvements.

Image-to-video generative AI can also be used for Image-to-Video generation, very useful for Virtual Reality. Two models that have been created through generative AI are GeoGPT (Ren and Wang, 2022) and SE3DS (Koh *et al*., 2023). GeoGPT provided a novel approach to synthesizing a consistent long-term video given a single scene image and a trajectory of large camera motions. SE3D is a method for high-resolution images and video generation from novel viewpoints, including viewpoints that extrapolate far beyond the input images while maintaining 3D consistency.

Text-to-video text-to-video models are still at an early stage. Among them, there is the Imagen video, meta Make a video, Phenaki, and Runway Gen-2. Imagen video uses a cascade of diffusion models for the creation of video outputs. Meta make a video is a video generation model created by meta-research that can do text-to-video, image-to-video, and video-editing. Although they are far from creating realistic outputs, they have shown promising signs and they have been useful from simple videos. Phenaki creates multiple-minute-long videos through text prompts. Moreover, Runway Gen-2 can generate videos through text, video, and image inputs. Shorter videos in GIF form can be generated through CogVideo, trained by inheriting a pre-trained text-to-image model, CogView2.

These video models have many applications, for example, the creation of videos with digital humans which could lead to cost reductions. Applications like Colossyan AI, Elai AI, Heygen AI, HourOne AI, Rephrase AI and Synthesia can create professional videos through diverse avatars. Some of them such as Synthesia combine this technology with 120 different languages for speech creation.

As well, you can use generative AI to transform articles into video outputs. SuperCreator is a mobile app that generates short videos for TikTok, Reels, and Shorts through Generative AI with an article input. As well synths video transforms articles into YouTube videos.

Generative AI can lead to deeper personalization of the videos, very useful for businesses. A very good example of this is Tavus AI, a video generation platform that personalizes videos of you to each audience member, automatically. As well, D-ID uses generative AI technologies to create real-time video to create an immersive human-like experience.

They can also be useful for artistic video generation. An example is Kaiber, an application that creates artistic videos through text and image prompts. Even for movie creation, with Opus AI, a text-to-video generator focused on everything from scenes, characters, dialogue, and visual effects.

Other notable video generation methods are Riverside AI which is an AI-powered video production site with edition capabilities, and Scenescape (Fridman *et al*., 2024), a method for text-driven perpetual view generation and human motion diffusion model.

As has been stated this is still a developing field because of the dimensionality of videos. Despite this fact, use cases have already been proven in education, where there was no significant difference in learning between showing learners traditional and synthetic videos (Leiker *et al*., 2023). This shows significant promise in the technology which could allow for much faster learning. It needs to be stated that educational videos are not the most complex kind. Nevertheless, if we can automatize simple video creation, we could see boosts in productivity in industries such as tourism, education, and news where videos are not too complex.

3D

3D modeling can work just with having a text prompt, an image, or a video input. They have varied applications such as game creation, the metaverse, and urban planning for which 3D designs are fundamental. Artificial intelligence shows signs of being able to produce 3D models even when there is a weak configuration of cameras (Pepe *et al*., 2023) which are very promising results for the automatization of 3D modeling, something costly for businesses.

Image-to-3D models in terms of image inputs, we can create 3D models with both a single image and with many images. For single-image inputs, popular models are GeNVS (Chan *et al*., 2023), Kaedim, MakeIt-3D (Tang *et al*., 2023a) and Real Fusion (Melas-Kyriazi *et al*., 2023). For many image inputs, we have NVIDIA Lion (Vahdat *et al*., 2022), EVA3D (Hong *et al*., 2022), Neural-Lift-360 (Xu *et al*., 2023), and Scenedreamer (Chen *et al*., 2023). Particularly for persons, we have PersoNeRF (Weng *et al*., 2023), which takes sample human images and generates a 3D model. We can also generate a 3D model through 2D images. We can also transform video inputs into 3D models through Deepmotion and Plask AI. Lastly, we can also create a 3d model through geometric points through NVIDIA LION (Vahdat *et al*., 2022).

A visible impact of this technology can be observed in the metaverse. Two companies that have combined both Generative AI and the metaverse are Metaphysic AI and Versy AI. Metaphysic AI as an example allows creators to transform image inputs into the production camera feed with minimal latency through generative AI.

Text-to-3D models regarding text inputs, some of the most important models are Adobe Firefly, dream fusion, GET3D (Gao *et al*., 2022), Magic3D (Lin *et al*., 2023), Synthesis AI and Text2Room. They created 3D textured shapes through text inputs. For animated 3D inputs, Mirage is a 3D tool to generate animated 3D pieces. We can even be able to generate 4D models through generative AI, as MAV3D (Singer *et al*., 2023) shows with a dynamic scene generator.

Although these technologies cannot directly substitute current workflows, the capabilities allow for unlimited possibilities. Google is already using generative AI for Google Maps which makes us hopeful of the possible capabilities of this technology (Li *et al*., 2023c).

Code and Software

Developers have been greatly assisted by these technologies by both GitHub Copilot and ChatGPT since this technology's inception. Through Natural Language, these models can help the user program and build websites. They can also help with more repetitive tasks of the programmer such as documentation. The most ambitious app, Adept, even says that generative AI can make humans communicate with a computer through just language. The democratization of code can help many professionals without a technical background be able to move around these programs with ease, which could be a major technological advance. In fact, the most famous coding Generative AI application, GitHub Copilot leads to increases of 55% in productivity, from a recent study (Peng *et al*., 2023).

Text-to-Application Creation Concerning app creation, there are many Generative AI technologies that can be used for app generation. With reference to apps, Berri AI, Debuild AI, Flutterflow, Google Generative App Builder, Imagica AI, Literally Anything IO, Scale Spellbook, Second AI and ZBrain are examples of Generative AI technologies with which users can easily create web apps through text prompts. Even LLM app creation has become easily available to non-technical professionals through text and data inputs.

Text-to-Cibersecurity measures even cybersecurity can be boosted through natural language, with Microsoft security copilot, an AI-powered security analysis tool that enables responses to threats quickly, processes signals, and assesses risk exposure.

Text-to-coding documentation for coding documentation, both Mintlify and Stenography have emerged as great ways to use generative AI for code documentation. Recent studies have shown that documentation takes up 5% of a programmer's time to code, thus advances towards automating documentation production seem very useful.

Text-to-coding translation code translation has also been made possible through generative AI, with Vercel AI code translator being one of the most useful tools. This could make companies shifting from one coding language to another much more productive.

Text-to-excel coding in terms of specific languages, Excel has been widely explored for spreadsheet code generation through generative AI. Some applications are AI Office Bot, Data Sheets GPT, Excel Formula Bot (Formula Bot AI Excel formula generator, google workspace AI-Sheets and sheets AI. They generate formulas quickly through text prompts and AI Office Bot even explains them. This can be very useful as Excel is one of the most widely used tools in office work and workflows could see improvements in productivity.

Text-to-multilingual coding there are several software for multilingual code generation through just text inputs. Although ChatGPT is widely used for coding, there are many more generative AI applications that are being created for that purpose. While most of them work as coding assistants, they are also able to generate code through text prompts. Some of them are Alphacode (Li *et al*., 2023a-c), Amazon Codewhisperer, BlackBox AI, code complete, CodeGeeX, Codeium, GitHub Copilot, GitHub Copilot X, GhostWriter Replit, Mutable AI (mutable.ai. They are used to complete, explain, transform, and generate code. They generate new lines based on context and syntax. As we can observe, it is one of the fields with the biggest amount of applications. They can be personalized to your writing style; Codex (Chen *et al*., 2023) is the model behind GitHub Copilot, the most famous coding assistant.

Text-to-SQL coding There are applications for SQL code generation like AI2SQL and seek AI.

Text-to-website creation regarding website creation, Durable, and Mutiny. Both applications generate a website with images and text through text prompts. Specifically for user interface generation, we have three applications, Diagram AI, Galileo AI, and Uizard AI, which use Generative AI for generating good user interfaces and optimizing the customer's experience. The.com even automates web page generation, so companies create personalized pages for each of their customers.

Unrelated generative AI Code technologies other technologies have emerged in the field of coding. An example is design-to-code technologies, through Locofy, a tool that turns designs into code for mobile apps and the web. Furthermore, text-to-automation tools through Drafter AI, a platform to automate even the most advanced analytical tasks, and Lasso AI which builds any robotic process automation using natural language. Even Adept has emerged with the project of making users able to interact with everything on their computer through text prompts.

Speech

Speech technologies try to imitate human speech. Text-to-speech technologies have made it easier to develop speeches. Other speech-to-speech technologies have made voice cloning very easy through generative AI. This technology has endless future possibilities in podcasts, YouTube videos, and helping mutes to communicate.

Speech-to-speech models, ACE-VC (Hussain *et al*., 2023), and VALL-E are the most important models. VALL-E specifically can take a three-second recording of someone's voice and replicate that voice, turning written words into speech, with realistic intonation and emotion depending on the context of the text.

Supertone AI provides speech editing similar to the other presented models.

Text-to-speech in terms of speech generation, generative AI has made it easy to create speech recordings through a text prompt. A plethora of platforms have been created including Coqui, Descript Overdub, ElevenLabs Listnr, Lovo AI, Resemble AI, Replica Studios, Voicemod and Wellsaid. The most important model is AudioLM (Borsos *et al*., 2023), Google's framework for high-quality audio generation with long-term consistency.

Video Recordings-to-Speech Dubverse, turns video recordings into speech, very useful for video dubbing tasks.

AI Understanding

AI has reached a good level in terms of translating different types of information in texts, videos, speech, and more into natural language. This is very useful because of

AI's ability to communicate and transform complex forms of communication into easier text. If we can transform any input into text, then we can easily understand it and we can even use that output as an input in other technologies, making for much more complete AI models. In Speech-to-Text tasks as an example, AI is very close to human performance, reaching an average error of 1.15% (Radford *et al*., 2023). These results are very promising because if computers are able to convert all inputs into text, they will be able to enable quick understanding of large and complex sets of data. In the future, this field shows the promise to convert a report composed of text, images, and audio into a text output which would be greatly helpful for users.

Generative region-to-text This technology has also been applied to generative region-to-text modeling. GriT (Wu *et al*., 2022) is a transformer that aims for object understanding with the region, and text pairs, where the region locates objects and the text describes objects. This can be useful for object detection tasks.

Images-to-text other technologies even turn images into text. These technologies can be used in many fields such as computer vision and help AI better understand human-generated content. As for these technologies, some examples of applications are Flamingo (Alayrac *et al*., 2022), Segment Anything (Kirillov *et al*., 2023), and VisualGPT (Chen *et al*., 2022).

Multiple inputs-to-Text There are even platforms in which we can transform multiple forms of input into text. Primer AI is a tool to understand and act on vast amounts of text, images, audio, and videos in real time. It helps in understanding and acting on this information to protect security and democracy. As for speak AI, it helps marketing and research teams turn unstructured audio, video, and text into competitive insights using transcription and NLP. Through both technologies, we can see how generative AI can help us quickly analyze big and unstructured sets of data. We can even get to understand and act on it through Primer and quickly obtain insights through Speak AI.

Pdf-to-text generative AI can also use natural language to retrieve information from documents. Two applications are ChatDOC and MapDeduce. They are able to quickly extract, locate, and summarize information from PDFs through natural language queries. Since late October 2023, ChatGPT has also integrated this capability into its model.

Speech-to-text one of the main fields has been speech-to-text technologies, as subtitles and transcriptions are very useful. Applications include Cogram AI, Deepgram AI, Dialpad AI, Fathom Video, Fireflies AI, GoogleUSM (Zhang *et al*., 2023a-e), Papercup, Reduct Video, Whisper (Radford *et al*., 2023) and Zoom IQ. These technologies do not only perform speech-to-text tasks, as some of them do much more. Deepgram AI identifies the speaker, the language, and the keywords. Dialpad AI includes real-time recommendations, call summaries, and the automation of customer touchpoints. Papercup even translates and creates human-sounding voices. Lastly, zoom has integrated AI into its systems, including features such as chat summaries and email drafts. By combining many generative AI technologies, we can observe how workflows can be optimized.

Table data-to-text generative AI has also been found useful in transforming tables of data into text. Some applications of Generative AI for this purpose are Defog AI, MURMUR (Saha *et al*., 2022), and TabT5. Murmur specifically is capable of understanding unstructured data. If we are able to perfect this technology, this could have major effects on optimizing business decision-making, through quickly understanding table data.

Videos-to-text flamingo is even able to achieve this task on video inputs. For video inputs, we have found TwelveLabs and Minotaur (Goyal *et al*., 2023). TwelveLabs extracts key features from a video input such as action, object, text on the screen, speech, and people and it transforms all of that into vector representations. These vectors enable quick search. Minotaur tackles query-based video understanding in long-form videos. In this space, another model called MOVIECLIP (Bose *et al*., 2023) was found very useful as it models the accurate recognition of visual scenes in movies. Through this technology, we can observe how computers are starting to understand unstructured sets of data effectively.

Business

Generative AI has clear implications, using many of the listed technologies such as text, image, and video in order to apply it to business. It can help businesses cut costs by reducing repetitive tasks or even automating other more creative, costly processes such as designs, marketing documents, or slide decks. It can even make new types of AI-powered businesses appear such as Harvey, which automates law, or Truewind, which automates accounting. Although young, we can only imagine how much generative AI will change the way in which businesses operate through the manners listed below. To show the relevance and impact of this technology, Goldman Sachs predicts that Generative AI could drive GDP growth by 7%.

Accounting documentation Other industries have also been affected. An example is accounting, with Truewind which applies AI to bookkeeping in order to make fewer errors and empower transparency.

Advertisements Advertisements can also be created through generative AI, as we can see through many apps such as Ad Creative AI, Clickable, Omneky, Pencil and Waymark. The last of them Waymark is very useful as it generates videos based on a scan of the web for local

business data. As well, LensAI is also useful as it fine-tunes ads by targeting through identifying objects, logos, actions, and context and matching them with relevant ads. Storytelling in these ads can also be powered by Generative AI, with applications such as AI 21 Labs and Subtxt that help in this matter.

Architecture Another example of an industry that has had some generative AI applications is architecture with SWAPP AI and Autodesk Space maker. SWAPP applies intelligent, advanced algorithms to deliver accurate, detailed, and complete Architectural construction documents and BIM models. Autodesk Spacemaker is a cloud-based AI software that empowers architects, urban planners, and real estate developers to design high-quality site proposals. In fact, Generative AI can also help in the Real Estate part of the process, which is shown by Zuma, AI-powered real estate assistance that automates lead generation.

Business visualizations in terms of research, there are already applications in which you can generate real-world data-backed answers through simple natural language queries. These answers come in the form of charts and visualizations to integrate processes and make market search even quicker. Several companies of this type are Alphawatch, Dataherald, OpenAxis AI, and Maya.

The company designs visual content that can be powered through generative AI. Designs can be quickly created through only text prompts as seen with Microsoft Designer which creates invitations, digital postcards, graphics, and more. Even logos are able to be created through generative AI, as can be observed through Brandmark and Looka AI. Brandmark also creates other business-related content such as business cards. For name ideas, you can use Namelix, brandinition, and Brandsnap in order to come up with business names.

Customer Assistance Generative AI can also be used to automate communication with the customer. A series of apps achieve personalized chatbots for your business: One Reach AI, OpenSight AI, Brainfish, and Yuma AI. E-mails can also be automatized through Generative AI with tools like InboxPro, Lavender, Smartwriter, and Twain. Some of these technologies even include social media data and e-mail analytics that can optimize operations. Even platforms with voice assistance such as Poly AI have been created.

Dubbing Lastly, dubbing can be helped by Generative AI through Sonantic, a Text-to-voice acting platform that provides editing and direction.

Education moreover, education has been affected, firstly with the advent of ChatGPT in students' work (Li *et al*., 2023a-c) and with companies such as Broadn, which uses language models and generative AI to help you create your own private learning course, unique to your learning style.

Feedback-generative AI can also be useful for receiving immediate feedback on our business ideas through the applications Venturus AI and Mixo AI which analyze business ideas.

Finance Generative AI can massively help the finance industry's tedious process. BloombergGPT a Large Language Model built from scratch for finance. It can be used for sentiment analysis, named entity recognition, news classification, and question answering, among others. We can see how this could be of massive help for finance professionals. Even for modeling, Quilt Labs AI is an AI-powered tool for the transformation of financial data into financial models. Finance can be seen as a great example of how applying generative AI to an industry can help automate processes.

Internal communications Applications can help massively in employee management through applications like Albus ChatGPT, ChatGPT in Slack, and Moveworks through conversation summaries and employee support automation.

Legal documentation Specific industries have been deeply affected by Generative AI. A great example of an industry that has been affected is law. Some applications and companies are Casetext CoCounsel, Darrow AI, Harvey AI and Spellbook Legal. Harvey AI assists with contract analysis, due diligence, litigation, and regulatory compliance and can help generate insights, recommendations, and predictions based on data. As for Darrow AI, it does case sourcing and due diligence to get law cases for your firm. Regarding Spellbook Legal, it uses GPT-4 to review and suggest the terms of your contract. We can observe how Generative AI is occupying many spaces in law that have the potential to be automated. In this space, there is an application called TaxGPT that takes advantage of OpenAI's GPT to fill out tax documents.

Marketing tasks for marketing, generative AI has had a huge effect, as it is able to make creative region and image generation easier. In terms of copywriting, a plethora of applications have already been developed, including Anyword, Copy AI, Google Workspace-Gmail and Docs, Hyperwrite, Jasper, Letterdrop, Regie AI, Simplified AI, Type AI and Writesonic. Some of the capabilities are writing emails, website content, drafts, replies marketing content, and product descriptions. We can easily see how the optimization of these processes would be very useful for many businesses. In fact, Regie AI even adapts the tone of the LLM to your company's tone, adapting even more to the business's needs. Here we again observe how businesses can combine many generative AI technologies to optimize their processes with Jasper, which does social media posts, emails, blogs, and reports.

Modeling even modeling could be affected by Generative AI with companies such as LA LA LAND providing an AI-powered digital model studio to show your 3D designs as lifelike models.

Multi-category business content we can observe how companies are investing resources into AI with EinsteinGPT, a platform created by Salesforce that creates personalized content across every Salesforce cloud. It will generate content across every sales, service, marketing, commerce, and IT interaction, transforming customer experience. An application that integrates all processes into one is AI Intern IO, which offers AI for most operations around a business: Text, reports, code, marketing, HR documents, legal documents, documentation, and translations.

Product design product creation can also be optimized through generative AI through an application like Cohere AI that offers LLM to retrieve, generate, and classify text to create the best products.

Repetitive tasks Generative AI can also help companies automate repetitive tasks. This can be achieved through many applications like Bardeen AI, Magical AI, and Notion AI. These applications specifically designed for repetitive tasks are especially useful for companies that want to automate relatively simple processes through machine learning.

Sales can be powered by generative AI through the plethora of applications that have already been created. Contact centers can be optimized through applications like Cresta, Forethought AI, Grain AI, and Replicant which transform the customer experience. Replicant can solve customer service over the phone, text, and chat. Others such as Cresta and Grain provide live help to contact centers. Cresta transforms real-time insights into real-time actions and Grain AI automates note-taking, record-keeping, and insight-capture for customer conversations. As for Forethought, it aims to automate the customer experience. For the sales preparation, an application Tennr was created to generate the perfect meeting prep before every sales call. There is even an app, Copy Monkey AI, created for optimizing Amazon listings and your product's ranking organically.

Scientific research It can also help tedious processes in scientific research. We can see this through applications such as Agolo AI, ArxivGPT, ConsensusNLP, Elicit AI and Koala. Some capabilities of these applications are finding papers, extracting key claims, and spotlighting insights. Specifically, ConsensusNLP and Koala are chatbots personalized to scientific research. Fact-checking has also been explored through generative AI through Golden.

Slide generation of the analyst's workflows can also be made easier through generative AI. This is achieved through helping both in slide generation and market research. In terms of slide generation, there are several apps that can create presentations through natural language. Some of them are Autoslide AI, Canva Docs to Decks, ChatBA, Decktopus AI, Gamma AI, google workspace AI-slides, Tome AI and Slide AI. Some of them work with just a small text prompt, like Tome AI and others work by introducing long texts like Canva Docs, which converts documents into slide presentations. As well, Decktopus even generates slide notes which can be quite useful.

Social media content More specifically for social media content creation, there are some applications like Clips AI, Pictory AI, Predis AI, Tweethunter and Tweetmonk. Clips AI and Pictory AI repurpose long-form content into social media posting. Predis AI generates video and image posts in your brand language. Both Tweethunter and Tweetmonk generate tweets of your brand's content. We can observe how Generative AI adapts to your brand and quickly automates these processes. Enterprises can also use Generative AI to generate podcasts through Bytepods.

Strategic decision-making Generative AI can also be helpful for more strategic, high-level departments of a company. Applications like Rationale AI can help in the creation of several business analyses through GPT.

Synthetic data generation Lastly, generative AI can be used to create realistic synthetic data for testing environments. This can be achieved through websites such as Hazy, Mostly AI, Octopize, and Tonic.

Travel Planning Other capabilities include travel itinerary creation, with application examples such as Roam Around, TripNotes, or ChatGPT's Kayak plug-in. The first two show the ability to create a visiting schedule while the Kayak plugin is able to look for hotels, flights, and more through natural language.

Gaming

The gaming industry will be greatly helped by Generative AI technologies, because of being able to use it from images, text, and 3D models. 3D models can help with creation and text models with storytelling and characters. We can view gaming as a very clear case study as to how Generative AI can be used through all parts of the value chain in a certain industry. These seem like very significant advances as studies point out that the implementation of Generative AI in the industry will lead to more realistic gaming environments as well as more engaging and personalized video games (Wu *et al*., 2023). I appreciate your patience. Here is the reordered information.

General Video Game Creation Generative AI can be used for video game creation. This can be seen through applications like CSM, Illiad AI, and Latitude. Explicitly for game assets, PixelVibe helps in the creation of them through generative AI.

Video game Characters Specifically for game characters, we have found Character AI, ConvAI, in world AI and RCT AI Chaos Box. ConvAI and World AI craft characters through natural language. Just by inserting character settings, you are able to obtain full characters.

As for RCT AI Chaos Box, this engine uses the Chaos Box algorithm to analyze real-time player inputs and dynamically generate NPC responses and new storylines based on Deep Reinforcement Learning.

The video game Levels MarioGPT (Sudhakaran *et al*., 2024) is designed for an open-ended text-to-level generation with LLMs.

Video game textures for game textures, Armorlab is a software designed for AI-powered texture authoring.

Music

Music creation can also be greatly helped by generative AI. You have the option to achieve this goal either by inputting a basic text prompt or by using music tracks as input. Some studies suggest that humans are unable to distinguish between AI-generated music and human-made music. This can be observed through an AI-generated track by Drake and The Weeknd, reaching more than 600,000 views on Spotify in 24 h. It only lasted 24 h because of the Intellectual Property issues that arose from using another individual's music samples for your own profit. This is a really clear example that if we don't use generative AI responsibly, problems will arise because of the use of public information. Other than purely creating songs, this technology also has a wide array of use cases in order to boost musician's creativity. When artists encounter a creative block, they can use these technologies in order to get ideas through text prompts, music samples, and even gestures.

Text-to-music in terms of music generation through natural language prompts, there have been many applications to do so. They include Aiva, ERNIE-music (Zhu *et al*., 2023), Harmonai, Infinite Album, Jukebox, Mubert, Musico, Noise2Music (Huang *et al*., 2023), Sonify, soundful and Splash AI. They have the capability of generating music through simple natural language. Splash AI is able to search and combine samples based on text prompts or use a powerful generative model trained on this dataset to create unique new production-quality music. Users are able to create music in varied genres such as hyperpop, EDM, Glitch, Phonk, Trap, Lo-Fi, and Hiphop. Despite tracks only being fifty seconds or less in length, results are still impressive. All the other models are similar to Splash AI in the sense of creating music tracks through a natural language prompt, while results can differ.

Other inputs-to-music models integrate other types of input into their technology. Musico reacts to gestures, movements, codes, and other sounds in order to create music. Even dance is starting to be able to be transformed into music through a model called EDGE.

Musical editing Musical editing can also be powered by Generative AI with applications such as Moises AI and SingSong (Donahue *et al*., 2023). Moises AI helps musical editing practitioners by eliminating voices and modifying speed and tone.

Biotechnology

Biotechnology is helped by generative AI technologies helping in the process of molecule modeling. This can help with both drug discovery and protein modeling advancing the field. As these technologies advance, biotechnology could as well see their advancements made much easier. Absci Corporation, a listed company in the NASDAQ, already uses generative AI in its drug creation process. Protein modeling for example is a field with very low success rates (Syrlybaeva and Strauch, 2023) and Generative AI is able to bring success rates up to 20%, quickly developing realistic structures (Winnifrith *et al*., 2023). We can see how these technologies can make workflows much quicker.

Drug Discovery Regarding drug discovery, NVIDIA Bionemo is a cloud service for generative AI in drug discovery researchers are provided with generative and predictive biomolecular AI models at scale. There are a plethora of companies that use Generative AI for drug creation including Atomic AI, BigHat AI, Exscientia, Menten AI and ProteinQure.

Using BigHat AI as an example to demonstrate capabilities in the field of drug discovery, they design antibody therapies for patients using machine learning and synthetic biology. Molecules are iteratively improved through AI models.

They own proprietary machine-learning models in order to design hundreds of different molecules at once. Molecular properties are then measured and this information is incorporated into the model. This is how through Generative AI and feedback from testing, drug discovery companies make their processes more efficient. They both improve efficiency and generate high-quality data. In November 2022, it announced a collaboration with Merck in order to collaborate in up to three drug discovery programs.

Protein modeling in terms of protein modeling, found models include BARTSmiles (Chilingaryan *et al*., 2022), a generative language model for molecular representation, and Alphafold (Ren *et al*., 2023), a computer program that predicts protein structures for the whole human genome. Also, two companies that have been found that center their business operations around protein design with Generative AI are Cradle and Profluent.

In order to observe a clear use case of AI in Protein design, we can look at Cradle's business operations. Cradle uses Alphafold's model in order to predict the structure of an individual sequence or multimer. Cradle optimizes the thermostability of your proteins with models trained on large datasets. Similar to BigHat AI, through generative AI you can design your proteins,

predict performance, and visualize your results. Then you can introduce your results in order to further train your company models.

Brain

Brain models convert brain waves into text and audio forms. Through their use, we could be able to better understand the meaning of one's brain waves. One of their most promising use cases is that they can help mute people communicate through generative AI. Although it's still a developing technology, some promising results already can be seen in this field, as Meta reached 41% accuracy results in their models. These are promising results as they can change how medical practitioners look at brains. As well, education can be bolstered through generative AI. Recent research (Darmawan, 2012) has shown that technology can help learning. If these technologies allow researchers to convert brain signals into text, we will better understand how brains communicate, thus empowering more efficient communication.

Brain signals-to-images Using Stable Diffusion for Brain Images (Tang *et al*., 2023b) is a new method that transforms brain waves into images based on a Diffusion Model (DM) called stable diffusion. This model was able to reconstruct images that participants had seen in the past. These are very impressive results but in order to recreate the experiment, the model would have to be re-trained, which makes it very resource-consuming. This marks the first time stable diffusion has been used in order to recreate brain images.

Brain signals-to-text regarding models that have been created to transform brain signals into text, we have found Meta AI's speech from the brain and Non-Invasive Brain Recordings (Défossez *et al*., 2023). They both try to decode speech from non-invasive brain recordings.

Multimodal

Models can take advantage of many of the listed technologies and combine them into one application. These listed applications take multiple inputs which can greatly help AI advancements. Also, projects of multi-tasking agents, such as GATO, could be the future of Generative AI. Although some models specifically text-to-slides do take advantage of many generative AI technologies, these models have been selected because of not fit anywhere else.

Multimodal chatbots The fourth version of GPT, GPT-4, can accept image, audio, and text inputs and produce text and image outputs. The most capable chatbot becoming multimodal is very significant for the field. Being able to communicate with chatbots through many inputs of data expands the possibilities of the chatbots and can help users understand more forms of data. Expanding forms of input can expand the technology's autonomy as Artificial Intelligence can then be integrated into more workflows than they could before.

Regarding multimodal language models, Kosmos-1 (Huang *et al*., 2024), is a multimodal language model with several capabilities. Some capabilities come in the form of language understanding and generation, perception-language tasks, including multimodal dialogue, image captioning, visual question answering, and vision tasks, such as image recognition with descriptions. Regarding Prismer (Liu *et al*., 2023a), it is a vision language model with multi-modal experts. Some tasks include image captioning, question answering, object detection, and segmentation. This model is competitive with current state-of-the-art vision models, whilst requiring up to two orders of magnitude less training data. As for PaLM-E (Driess *et al*., 2023), it is an embodied multimodal language model. On the one hand, PaLM-E was primarily developed to be a model for robotics and it solves a variety of tasks on multiple types of robots and for multiple modalities (images, robot states, and neural scene representations). At the same time, PaLM-E is a generally capable vision-and-language model. It can perform visual tasks, such as describing images, detecting objects, or classifying scenes, and is also proficient at language tasks, like quoting poetry, solving math equations, or generating code. In the realm of chatbots that can take many forms of data as inputs, the ERNIE bot can generate text, images, audio, and video given a text prompt and can deliver voice in several local Chinese dialects, including Sichuan. Ernie's user turnout did not align with initial expectations. Nevertheless, its CEO has claimed that the chatbot matches ChatGPT's capabilities.

Integrated AI models frameworks as for multimodal cloud services in generative AI, NVIDIA Picasso is a cloud service for building and deploying generative AI-powered image, video, and 3D applications. It integrates text-to-image, text-to-video, and text-to-3D models.

There even is a framework called HuggingGPT (Shen *et al*., 2024) that leverages LLMs (e.g., ChatGPT) to connect various AI models in machine learning communities (e.g., hugging face) to solve AI tasks. It makes LLMs act as a controller to manage existing AI models to solve complicated AI tasks and language could be a generic interface to empower this. It achieves impressive results in language, vision, speech, and other challenging tasks, which paves a new way towards advanced artificial intelligence. Combining different models through the same framework can produce impressive results. You can generate audio through text written in an image. Combining these models will provide endless possibilities, using AI to optimize many production processes all at once. Just combining text and image models, a graphic design could very quickly achieve higher efficiency.

Concerning adobe firefly it is a family of Adobe models that uses text to create images, vectors, videos, and 3D models out of text prompts. It comprises several AI models to generate images and other assets through text prompts. The generative fill option is now in Photoshop, where it allows users to add, extend, and remove content from their images with simple text prompts. For example, you can quickly remove people, add puddles, and add animals. You can also generate new illustrations from a single one. By adding all these features into the same application, workflows become quicker.

Generalist agents as for attempts at generalist agents, GATO is a single agent beyond text outputs. It follows a multi-modal, multi-task, multi-embodiment generalist policy. The same network can play games, chat, and press buttons at the same time. Regarding Imbue (formerly generally intelligent), it is a company in charge of the development of generally capable agents. Valued at 1 billion dollars, they focus on imbuing computers with intelligence and human values. Their aim is to deploy aligned human-level AI systems that can generalize to a wide range of economically useful tasks and assist with scientific research. In order to achieve so, they prototype agents for internal use with the help of their big pre-trained models focusing on coding tasks at the moment. To our knowledge, this is the only company that focuses on AI Agents. As well, we believe that their project is significant as Imbue's focus is on making more explainable models.

Steps towards AGI or Artificial General Intelligence are very significant since they bring us closer to human-like intelligence through computer systems (Latif *et al*., 2023). AutoGPT became the first approach to AGI (Talebirad and Nadiri, 2023) by acting autonomously from human instructions. This made the model earn over 150,000 stars on GitHub.

Computational Efficiency

Alphatensor Fawzi *et al*. (2022) is the first-ever Artificial Intelligence (AI) system for developing new, effective, and indubitably correct algorithms for essential operations like matrix multiplication based on reinforcement learning. The task given to Alphatensor was to improve the efficiency of matrix multiplications, which occur in many fundamental computations. Automating the algorithm discovery procedure is intricate, as the space of possible algorithms is enormous. That is why this model uses AlphaTensor, which is trained to play a single-player game where the objective is finding tensor decompositions within a finite factor space. AlphaTensor discovered algorithms that outperform the state-of-the-art complexity for many matrix sizes. Matrix multiplication is crucial for many areas from neural networks to scientific computing routines thus this study will surely have an effect on computing efficiency.

Conclusion

This article provides an up-to-date outlook on generative AI applications. The biggest finding from the article is the compilation of technologies that had not been done before. We have found players in each of the selected categories and some examples of the capabilities of the technologies they are using.

Main Takeaways

In conclusion, generative AI has already demonstrated its immense potential to revolutionize various industries and reshape our interactions with digital content. As these models continue to advance, they offer businesses and individuals unprecedented capabilities in content creation, problem-solving, and decision-making. Their capacity to generate realistic images, audio, text, and other data modalities unlocks novel opportunities for innovation and growth, while also enabling more personalized and efficient experiences. This survey describes how technologies can reshape particular sectors and details models in several output categories such as text, image, and video. Examples and capabilities are included in each section, clearly detailing what the impact of Generative AI is in each category. This survey helps users better navigate the very complex and changing landscape of generative AI.

Ethical Challenges in Generative AI

By fostering responsible development and adoption of generative AI, we can harness its transformative potential to shape a more creative, efficient, and prosperous future for businesses and individuals alike. The most famous technology, ChatGPT, has been proven to generate factual statements that cannot be verified from the source. This can make mislead individuals who are over-confident in ChatGPT. Thus, some good recommendations are using LLMs for creative writing in tasks such as brainstorming and ideas generation and for non-creative writing in tasks such as spelling and grammar check, summarization, copywriting copy editing, and coding assistance (Sison *et al*., 2023). For the correct adoption of this technology, individuals must understand what the use of each should be. Users should never share training data as models such as ChatGPT use responses to further train their models through Reinforcement Learning from Human Feedback. Users should also be careful with intellectual property rights as models are trained with public data. As said before in the text, the text output that GPT-2 produces is indifferent to human output thus we have to be very careful when trying to understand who is the decision-maker behind a certain piece of text. Other ethical implications arise through applications such as ChatDoctor which can deliver a medical diagnosis. Ethical challenges will arise from handling private data and managing AI responsibility for critical situations such as cancer diagnosis (La Salvia *et al*., 2022).

Future Work

For future work, this survey is to be updated. The launch of ChatGPT-3 propelled the launch of many of the mentioned applications. The venture capital industry is investing a big amount of money into these technologies thus it is highly likely that many more examples will appear shortly after this article is published. This constantly changing landscape makes updating this body of text necessary. Only a year before the article was published, generative AI was mostly kept in academia. Throughout this article, we have seen a plethora of applications. This exponential growth highlights the need to be updated in this industry.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

The authors have not received any financial support or funding to report.

Authors' Contributions

Roberto Gozalo-Brizuela: Contributed to the manuscript through drafting the text and conducting extensive research.

Eduardo Garrido Merchan**:** Provided the initial concept and oversaw the project, ensuring scholarly standards and integrity.

Ethics

The presented research study represents an original contribution and has not been previously submitted or published. The authors have thoroughly reviewed and complied with the submission requirements for this original paper, as well as adhered to the research ethics code.

References

Alayrac, J. B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., ... & Simonyan, K. (2022). Flamingo: A visual language model for few-shot learning. *Advances in Neural Information Processing Systems*, *35*, 23716-23736. https://doi.org/10.48550/arXiv.2204.14198

Aydin, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations?. *Academic Platform Journal of Engineering and Smart Systems*, *11*(3), 118-134. https://doi.org/10.21541/apjess.1293702

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023, April). ChatGPT: Applications, opportunities and threats. *In 2023 Systems and Information Engineering Design Symposium* (*SIEDS*), (pp. 274-279). IEEE. https://doi.org/10.1109/SIEDS58326.2023.10137850

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination and interactivity. *ArXiv Preprint ArXiv:2302.04023*. https://doi.org/10.48550/arXiv.2302.04023

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., ... & Zeghidour, N. (2023). Audiolm: a language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech and Language Processing*. https://doi.org/10.1109/TASLP.2023.3288409

Bose, D., Hebbar, R., Somandepalli, K., Zhang, H., Cui, Y., Cole-McLaughlin, K., ... & Narayanan, S. (2023). Movieclip: Visual scene recognition in movies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (pp. 2082-2091). https://doi.org/10.1109/WACV56688.2023.00212

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (aigc): A history of generative AI from gan to chatGPT. *ArXiv Preprint ArXiv:2303.04226*. https://doi.org/10.48550/arXiv.2303.04226

Chan, E. R., Nagano, K., Chan, M. A., Bergman, A. W., Park, J. J., Levy, A., ... & Wetzstein, G. (2023). Generative novel view synthesis with 3D-aware diffusion models. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, (pp. 4217-4229). https://doi.org/10.48550/arXiv.2304.02602

Chang, H., Zhang, H., Barber, J., Maschinot, A. J., Lezama, J., Jiang, L., ... & Krishnan, D. (2023). Muse: *Text-to-Image Generation Via Masked Generative Transformers*. *ArXiv Preprint ArXiv: 2301.00704*. https://doi.org/10.48550/arXiv.2301.00704

Chen, J., Guo, H., Yi, K., Li, B., & Elhoseiny, M. (2022). Visualgpt: Data-efficient adaptation of pretrained language models for image captioning. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 18030-18040). https://doi.org/10.48550/arXiv.2102.10407

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., ... & Zaremba, W. (2021). Evaluating Large Language Models Trained on Code. *arXiv preprint arXiv:2107.03374*. https://doi.org/10.48550/arXiv.2107.03374

Chen, Z., Wang, G., & Liu, Z. (2023). Scenedreamer: Unbounded 3D Scene Generation from 2D Image Collections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2023.3321857

Cheng, C. Y., Huang, F., Li, G., & Li, Y. (2023). Play: Parametrically conditioned layout generation using latent diffusion. *arXiv preprint arXiv:2301.11529*. https://doi.org/10.48550/arXiv.2301.11529

Chilingaryan, G., Tamoyan, H., Tevosyan, A., Babayan, N., Khondkaryan, L., Hambardzumyan, K., ... & Aghajanyan, A. (2022). Bartsmiles: Generative masked language models for molecular representations. *arXiv preprint arXiv:2211.16349*. https://doi.org/10.48550/arXiv.2211.16349

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An overview. *IEEE Signal Processing Magazine*, *35*(1), 53-65. https://doi.org/10.1109/MSP.2017.2765202

Darmawan, D. (2012). Biological Communication Behavior through Information Technology Implementation in Learning Accelerated. https://doi.org/10.4236/ijcns.2012.58056

Donahue, C., Caillon, A., Roberts, A., Manilow, E., Esling, P., Agostinelli, A., ... & Engel, J. (2023). SingSong: Generating Musical Accompaniments from Singing. *arXiv preprint arXiv:2301.12662*. https://doi.org/10.48550/arXiv.2301.12662

Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Ichter, B., ... & Florence, P. (2023). Palm-e: An Embodied Multimodal Language Model. *arXiv preprint arXiv:2303.03378*. https://doi.org/10.48550/arXiv.2303.03378

Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., & King, J. R. (2023). Decoding Speech Perception from non-invasive Brain Recordings. *Nature Machine Intelligence*, *5*(10), 1097-1107. https://doi.org/10.1038/s42256-023-00714-5

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An Early Look at the Labor Market Impact Potential of Large Language Models. *arXiv preprint arXiv:2303.10130*. https://doi.org/10.48550/arXiv.2303.10130

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., ... & Kohli, P. (2022). Discovering Faster Matrix Multiplication Algorithms with Reinforcement Learning. *Nature*, *610*(7930), 47-53. https://doi.org/10.1038/s41586-022-05172-4

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business and Information Systems Engineering*, *66*(1), 111-126. https://doi.org/10.1007/s12599-023-00834-7

Fridman, R., Abecasis, A., Kasten, Y., & Dekel, T. (2024). Scenescape: Text-Driven Consistent Scene Generation. *Advances in Neural Information Processing Systems*, *36*. pp. 39897-39914.

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P., & Berner, J. (2024). Mathematical capabilities of chatgpt. *Advances in Neural Information Processing Systems*, *36*. pp. 27699-27744.

Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., ... & Fidler, S. (2022). Get 3D: A generative model of high quality 3D textured shapes learned from images. *Advances in Neural Information Processing Systems*, *35*, 31841-31854.

Garrido-Merchán, E. C., Arroyo-Barrigüete, J. L., & Gozalo-Brihuela, R. (2023). Simulating HP Lovecraft Horror Literature with the ChatGPT Large Language Model. *arXiv preprint arXiv:2305.03429*. https://doi.org/10.48550/arXiv.2305.03429

Goyal, R., Mavroudi, E., Yang, X., Sukhbaatar, S., Sigal, L., Feiszli, M., ... & Tran, D. (2023). Minotaur: Multi-task Video Grounding from Multimodal Queries. *arXiv preprint arXiv:2302.08063*. https://doi.org/10.48550/arXiv.2302.08063

Hong, F., Chen, Z., Lan, Y., Pan, L., & Liu, Z. (2022). Eva3D: Compositional 3D Human Generation from 2D Image Collections. *arXiv preprint ArXiv: 2210.04888*. https://doi.org/10.48550/arXiv.2210.04888

Huang, Q., Park, D. S., Wang, T., Denk, T. I., Ly, A., Chen, N., ... & Han, W. (2023). Noise2music: Text-Conditioned Music Generation with Diffusion Models. *arXiv preprint arXiv:2302.03917*. https://doi.org/10.48550/arXiv.2302.03917

Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., ... & Wei, F. (2024). Language is not all you need: Aligning Perception with Language Models. *Advances in Neural Information Processing Systems*, *36*. pp. 72096-72109.

Hussain, S., Neekhara, P., Huang, J., Li, J., & Ginsburg, B. (2023, June). Ace-vc: Adaptive and Controllable Voice Conversion using Explicitly Disentangled self-Supervised Speech Representations. *In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), (pp. 1-5). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10094850

Joublin, F., Ceravola, A., Deigmoeller, J., Gienger, M., Franzius, M., & Eggert, J. (2023). A Glimpse in ChatGPT Capabilities and its Impact for AI Research. *arXiv preprint arXiv:2305.06087*. https://doi.org/10.48550/arXiv.2305.06087

Kauers, M., & Moosbauer, J. (2022). The FBHHRBNRSSSHK-Algorithm for Multiplication in $\mathbb{Z}_2^{5\times5}$ is Still not the End of the story. *arXiv preprint arXiv:2210.04045*. https://doi.org/10.48550/arXiv.2210.04045

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment Anything. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, (pp. 4015-4026). https://doi.org/10.48550/arXiv.2304.02643

Koh, J. Y., Agrawal, H., Batra, D., Tucker, R., Waters, A., Lee, H., ... & Anderson, P. (2023). Simple and Effective Synthesis of Indoor 3D Scenes. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. *37*, No. 1, pp. 1169-1178). https://doi.org/10.1609/aaai.v37i1.25199

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the News that's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, *9*(1), 104-117. https://doi.org/10.1017/XPS.2020.37

Kulkarni, C., Druga, S., Chang, M., Fiannaca, A., Cai, C., & Terry, M. (2023). A Word is Worth a Thousand Pictures: Prompts as AI Design Material. *arXiv preprint arXiv:2303.12647*. https://doi.org/10.48550/arXiv.2303.12647

La Salvia, M., Torti, E., Leon, R., Fabelo, H., Ortega, S., Martinez-Vega, B., ... & Leporati, F. (2022). Deep Convolutional Generative Adversarial Networks to Enhance Artificial Intelligence in Healthcare: A skin cancer application. *Sensors*, *22*(16), 6145. https://doi.org/10.3390/s22166145

Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., ... & Zhai, X. (2023). Artificial General Intelligence (AGI) for Education. *arXiv preprint arXiv: 2304.12479*, *1*. https://doi.org/10.48550/arXiv.2304.12479

Leiker, D., Gyllen, A. R., Eldesouky, I., & Cukurova, M. (2023). Generative AI for Learning: Investigating the Potential of Synthetic Learning Videos. *arXiv preprint arXiv2304.03784*. https://doi.org/10.48550/arXiv.2304.03784

Lewkowycz, A. andreassen, A., Dohan, D., Dyer, E., Michalewski, H., Ramasesh, V., ... & Misra, V. (2022). Solving Quantitative Reasoning Problems with Language Models. *Advances in Neural Information Processing Systems*, *35*, 3843-3857.

Li, L., Ma, Z., Fan, L., Lee, S., Yu, H., & Hemphill, L. (2023a). ChatGPT in Education: A Discourse Analysis of Worries and Concerns on Social Media. *Education and Information Technologies*, 1-34. https://doi.org/10.1007/s10639-023-12256-9

Li, Y., Liu, H., Wu, Q., Mu, F., Yang, J., Gao, J., ... & Lee, Y. J. (2023b). Gligen: Open-set grounded text-to-image generation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 22511-22521). https://doi.org/10.48550/arXiv.2301.07093

Li, C., Zhang, C., Waghwase, A., Lee, L. H., Rameau, F., Yang, Y., ... & Hong, C. S. (2023c). Generative AI Meets 3D: A Survey on Text-to-3D in AIGC Era. *arXiv preprint arXiv:2305.06131*. https://doi.org/10.48550/arXiv.2305.06131

Lin, C. H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., ... & Lin, T. Y. (2023). Magic 3D: High-resolution text-to-3D content creation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 300-309). https://doi.org/10.48550/arXiv.2211.10440

Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G. N., & Togelius, J. (2021). Deep Learning for Procedural Content Generation. *Neural Computing and Applications*, *33*(1), 19-37. https://doi.org/10.1007/s00521-020-05383-8

Liu, S., Fan, L., Johns, E., Yu, Z., Xiao, C., & Anandkumar, A. (2023a). Prismer: A Vision-Language Model with an Ensemble of Experts. *arXiv preprint arXiv:2303.02506*, *3*. https://doi.org/10.48550/arXiv.2303.02506

Liu, G. H., Vahdat, A., Huang, D. A., Theodorou, E. A., Nie, W., & Anandkumar, A. (2023b). I^2SB: Image-to-IMAGE Schr\" Odinger Bridge. *arXiv preprint arXiv:2302.05872*. https://doi.org/10.48550/arXiv.2302.05872

Melas-Kyriazi, L., Laina, I., Rupprecht, C., & Vedaldi, A. (2023). Realfusion: 360° reconstruction of any object from a single image. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (pp. 8446-8455).

Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT press. ISBN-10: 0262046822.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*. https://doi.org/10.48550/arXiv.2302.06590

Pepe, M., Alfio, V. S., & Costantino, D. (2023). Assessment of 3D Model for Photogrammetric Purposes Using AI Tools Based on NeRF Algorithm. *Heritage*, *6*(8), 5719-5731. https://doi.org/10.3390/heritage6080301

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust Speech Recognition Via Large-Scale Weak Supervision. *In International Conference on Machine Learning*, (pp. 28492-28518). PMLR. https://proceedings.mlr.press/v202/radford23a.html

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, *1*(2), 3. https://3dvar.com/Ramesh2022Hierarchical.pdf

Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W., ... & Zhavoronkov, A. (2023). AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical Science*, *14*(6), 1443-1452. https://doi.org/10.1039/d2sc05709c

Ren, X., & Wang, X. (2022). Look outside the room: Synthesizing a consistent long-term 3D scene video from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 3563-3573). https://doi.org/10.48550/arXiv.2203.09457

Saha, S., Yu, X. V., Bansal, M., Pasunuru, R., & Celikyilmaz, A. (2022). Murmur: Modular multi-step reasoning for semi-structured data-to-text generation. *arXiv preprint arXiv:2212.08607*. https://doi.org/10.48550/arXiv.2212.08607

Schick, T., Dwivedi-Yu, J., Jiang, Z., Petroni, F., Lewis, P., Izacard, G., ... & Riedel, S. (2022). Peer: A collaborative language model. *arXiv preprint arXiv:2208.11663*. https://doi.org/10.48550/arXiv.2208.11663

Schwitzgebel, E., Schwitzgebel, D., & Strasser, A. (2023). Creating a large language model of a philosopher *Mind and Language*. https://doi.org/10.1111/mila.12466

Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2024). Hugginggpt: Solving AI tasks with chatgpt and its friends in hugging face. *Advances in Neural Information Processing Systems*, *36*, pp. 38154-38180.

Singer, U., Sheynin, S., Polyak, A., Ashual, O., Makarov, I., Kokkinos, F., ... & Taigman, Y. (2023). Text-to-4D dynamic scene generation. *arXiv preprint arXiv:2301.11280*. https://doi.org/10.48550/arXiv.2301.11280

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., ... & Natarajan, V. (2023). Towards expert-level medical question answering with large language models. *arXiv preprint arXiv:2305.09617*. https://doi.org/10.48550/arXiv.2305.09617

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). ChatGPT: More than a "weapon of mass deception" ethical challenges and responses from the Human-Centered Artificial Intelligence (HCAI) perspective. *International Journal of Human–Computer Interaction*, 1-20. https://doi.org/10.1080/10447318.2023.2225931

Sudhakaran, S., González-Duque, M., Freiberger, M., Glanois, C., Najarro, E., & Risi, S. (2024). Mariogpt: Open-ended text2level generation through large language models. *Advances in Neural Information Processing Systems*, *36*. pp. 54213-54227.

Syrlybaeva, R., & Strauch, E. M. (2023). Deep learning of protein sequence design of protein–protein interactions. *Bioinformatics*, *39*(1), btac733. https://doi.org/10.1093/bioinformatics/btac733

Talebirad, Y., & Nadiri, A. (2023). Multi-agent collaboration: Harnessing the power of intelligent llm agents. *arXiv preprint arXiv:2306.03314*. https://doi.org/10.48550/arXiv.2306.03314

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023a). Semantic Reconstruction of Continuous Language from Non-Invasive Brain Recordings. *Nature Neuroscience*, *26*(5), 858-866. https://doi.org/10.1038/s41593-023-01304-9

Tang, J., Wang, T., Zhang, B., Zhang, T., Yi, R., Ma, L., & Chen, D. (2023b). Make-it-3D: High-fidelity 3D creation from a single image with diffusion prior. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 22819-22829). https://make-it-3d.github.io/

Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., ... & Stojnic, R. (2022). Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*. https://doi.org/10.48550/arXiv.2211.09085

Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., & Bermano, A. H. (2022). Human motion diffusion model. *arXiv preprint arXiv:2209.14916*. https://doi.org/10.48550/arXiv.2209.14916

Tian, H., Lu, W., Li, T. O., Tang, X., Cheung, S. C., Klein, J., & Bissyandé, T. F. (2023). Is ChatGPT the ultimate programming assistant-how far is it?. *arXiv preprint arXiv:2304.11938*. https://doi.org/10.48550/arXiv.2304.11938

Weng, C. Y., Srinivasan, P. P., Curless, B., & Kemelmacher-Shlizerman, I. (2023). Personnerf: Personalized reconstruction from photo collections. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 524-533). https://doi.org/10.48550/arXiv.2302.08504

Winnifrith, A., Outeiral, C., & Hie, B. (2023). Generative artificial intelligence for de novo protein design. *arXiv preprint arXiv:2310.09685*. https://doi.org/10.48550/arXiv.2310.09685

Wu, J., Wang, J., Yang, Z., Gan, Z., Liu, Z., Yuan, J., & Wang, L. (2022). Grit: A generative region-to-text transformer for object understanding. *arXiv preprint arXiv:2212.00280*. https://doi.org/10.48550/arXiv.2212.00280

Wu, Y., Yi, A., Ma, C., & Chen, L. (2023). Artificial intelligence for video game visualization, advancements, benefits and challenges. *Mathematical Biosciences and Engineering*, *20*(8), 15345-15373. https://doi.org/10.3934/mbe.2023686

Xu, D., Jiang, Y., Wang, P., Fan, Z., Wang, Y., & Wang, Z. (2023). Neurallift-360: Lifting an in-the-wild 2d photo to a 3D object with 360° views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 4479-4489).

Yu, J., Xu, Y., Koh, J. Y., Luong, T., Baid, G., Wang, Z., ... & Wu, Y. (2022). Scaling autoregressive models for content-rich text-to-image generation. *arXiv preprint arXiv:2206.10789*, *2*(3), 5. https://3dvar.com/Yu2022Scaling.pdf

Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., & Kreis, K. (2022). Lion: Latent point diffusion models for 3D shape generation. *Advances in Neural Information Processing Systems*, *35*, pp. 10021-10039.

Zhang, M., Qamar, M., Kang, T., Jung, Y., Zhang, C., Bae, S. H., & Zhang, C. (2023a). A survey on graph diffusion models: Generative AI in science for molecule, protein and material. *arXiv preprint arXiv:2304.01565*. https://doi.org/10.48550/arXiv.2304.01565

Zhang, Y., Han, W., Qin, J., Wang, Y., Bapna, A., Chen, Z., ... & Wu, Y. (2023b). Google USM: Scaling automatic speech recognition beyond 100 languages. *arXiv preprint arXiv:2303.01037*. https://doi.org/10.48550/arXiv.2303.01037

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. S. (2023c). Text-to-image diffusion model in generative Ai: A survey. *arXiv preprint arXiv:2303.07909*. https://doi.org/10.48550/arXiv.2303.07909

Zhang, C., Zhang, C., Zheng, S., Qiao, Y., Li, C., Zhang, M., ... & Hong, C. S. (2023D). A complete survey on generative AI (AIGC): Is chatGPT from GPT-4 to GPT-5 all you need? *arXiv preprint arXiv: 2303.11717*. https://doi.org/10.48550/arXiv.2303.11717

Zhang, C., Zhang, C., Zheng, S., Zhang, M., Qamar, M., Bae, S. H., & Kweon, I. S. (2023e). Audio diffusion model for speech synthesis: A survey on text to speech and speech enhancement in generative AI. *arXiv preprint arXiv:2303.13336*. https://doi.org/10.48550/arXiv.2303.13336

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*. https://doi.org/10.48550/arXiv.2303.18223

Zhu, P., Pang, C., Chai, Y., Li, L., Wang, S., Sun, Y., ... & Wu, H. (2023). ERNIE-music: Text-to-waveform music generation with diffusion models. *arXiv preprint arXiv:2302.04456*. https://doi.org/10.48550/arXiv.2302.04456