Generative Artificial Intelligence: Unveiling the Potential and Challenges

Brahmaleen Kaur Sidhu

Department of Computer Science and Engineering, Punjabi University, Patiala, Punjab, India
Email: *brahmaleen.ce[at]pbi.ac.in*

Abstract: ***Generative artificial intelligence (AI) is a subset of artificial intelligence that has emerged as a transformative technology with the potential to revolutionize various domains, including art, entertainment, research, healthcare, and finance. Generative AI models can generate answers, essays, poems, stories, product descriptions, and all manner of text. They can also produce images, music, audio, video, and synthetic training data. Among the beneficiaries are data scientists, application developers, marketers, sales teams, digital artists, designers in the media, educators, and researchers. On the flip side, generative AI has also heightened risks of potential copyright infringements, data privacy violations, discrimination, deep fakes, and other deceptive practices. This paper presents an in-depth exploration of the foundations of generative AI, its potential applications, and the challenges associated with developing and deploying generative AI models.***

Keywords: Diffusion models, Embeddings, Generative adversarial networks, Generative artificial intelligence, Prompt engineering, Transformers, Variational autoencoders

1. Introduction

Generative artificial intelligence enables users to generate new, realistic and high-quality content including text, images, sounds, animation, 3D models, etc., based on a variety of inputs. The field of generative AI has seen drastic growth in recent years, owing to advances in research, algorithms, computational resources, and applications. This is evident from the surge in number of patents filed in the subject in recent years as shown in Figure 1. The Lens reports over 20,000 patents records in generative AI [1]. Figure 2 lists some of the top patent owners in the field.

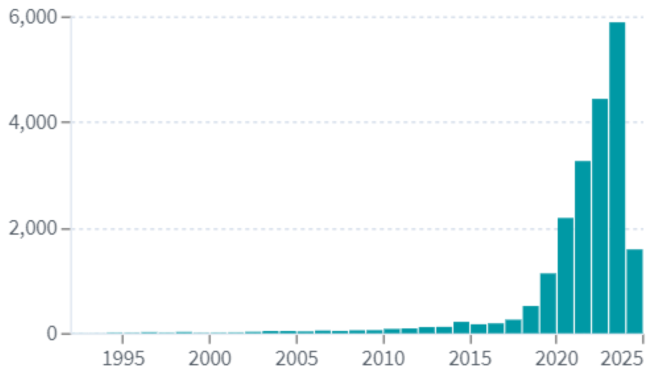

Figure 1: Number of patents published in the field of Generative AI in recent years

Generative AI has tremendous potential across various areas. In the field of education and writing, it can help by exploring ideas, brainstorming and providing real-time feedback. In healthcare, generative AI can lead to a better understanding of health and help in drug discovery, ultimately assisting medical practitioners in development of new treatments. In biology and genomics, it can help analyse large volumes of data. In agriculture, it can help achieve greater yield by improving remote sensing and robotics. In software development, generative AI can be used for faster and easier development by suggesting code. In geoscience, generative AI can derive a better understanding of Earth's ecosystems and enhance sensing and forecasting. A report by McKinsey Global Institute estimates that generative AI carries the potential to add equivalent of \$2.6 trillion to \$4.4 trillion annually to the global economy [2].

Figure 2: Top owners of patents in the field of Generative AI

Senior digital and analytics leaders at a McKinsey forum held in 2023 on generative AI said they believed that the technology will fundamentally change the way they do business. Economists at Goldman Sachs comment that despite significant uncertainty around the potential for generative AI, its ability to generate content that is indistinguishable from human-created output and to break down communication barriers between humans and machines reflects a major advancement with potentially large macroeconomic effects. These breakthroughs are estimated to drive a 7% (or almost \$7 trillion) increase in global GDP and lift productivity growth by 1.5 percentage points over a 10-year period [3].

Gartner believes that the field of generative AI will progress rapidly in both scientific discovery and technology commercialization. Gartner's 2022 predictions regarding generative AI claim that by 2025, generative AI will be producing 10 percent of all data with 20 percent of all test data for consumer-facing use cases and will be used by 50 percent of drug discovery and development initiatives; and by 2027, 30 percent of manufacturers will use generative AI to enhance their product development effectiveness. Quoting from a Gartner research excerpt of 2024 [4]: "Generative AI is spearheading the productivity revolution: improving existing features, adding new functionality and augmenting human potential to use unstructured data systems."

The productivity potential of generative AI is driving its widespread adoption across the enterprise, being applied for creating new materials to preserving data privacy. As with most of the technologies, some safety concerns prevail such as negative use of deepfakes. Nonetheless, the exploration of generative AI methods is growing and proving itself in a wide range of industries, including life sciences, healthcare, manufacturing, material science, media, entertainment, automotive, aerospace, defence and energy.

Generative AI is a powerful tool for streamlining the workflow of creatives, engineers, researchers, scientists, and more. The use cases and possibilities span all industries and individuals. This paper comprises seven sections. The second section defines generative AI. The third section discusses various architectures used in development of generative AI models. The fourth section lists some emerging trends in this subject. The fifth and sixth sections talk about the applications and challenges of generative AI, respectively. The last section concludes the paper.

2. What is Generative AI

Generative AI describes algorithms that create incredibly realistic content that reflects characteristics of the training data. These algorithms learn a representation of artifacts from the training data and use it to generate novel artifacts that preserve a likeness to original data. It can produce a variety of content including audio, code, images, text, simulations, and videos.

Generative AI starts with a prompt that could be in the form of a text, an image, a video, musical notes or any other input that the AI system can process. AI algorithms then return new content and response to the prompt. Prompts are the instructions or questions provided to the model, specifying what kind of output is to be generated. The process of carefully designing prompts, in order to influence the style, tone, topic, or characteristics of the generated content and ensure that it aligns with their preferences or objectives, is known as prompt engineering.

Unlike discriminative artificial intelligence that performs classification tasks, modern generative AI uses machine learning and deep neural networks to understand and conditionally generate new examples from complex data distributions. Once the training data is collected, the generative AI model analyses patterns and relationships within the data to understand the underlying rules governing the content. It continuously fine-tunes its parameters as it learns, improving its ability to generate content that appears authentic and human-like.

3. How Generative AI works

Developing a generative AI model is a resource intensive process. Various architectures are used in the development of generative AI, as discussed below.

3.1 Embeddings to represent high-dimensional data

Embeddings act like a translator, transforming complex data which can be text, images, audio, or any other format into a more concise and machine-readable format. This format is typically a vector of numbers, where each number captures specific aspects or features of the original data. Embeddings are used to represent high-dimensional and complex data in a lower-dimensional space while preserving important relationships and characteristics of the original data. The underlying structure and semantics of the data is preserved, which is essential for generating meaningful and coherent outputs in generative AI. Word embeddings are used to represent words as dense vectors in a lower-dimensional space. These embeddings capture semantic similarities between words, allowing generative models to understand and manipulate textual data more effectively. For example, in models like Word2Vec [4], GloVe [5], or fastText [6], words with similar meanings are represented by vectors that are closer together in the embedding space.

Image embeddings are used in computer vision to represent images as feature vectors in a lower-dimensional space. Techniques like convolutional neural networks (CNNs) are used to extract image embeddings that capture visual features and patterns. These embeddings enable generative models to generate realistic images or perform tasks like image captioning and image retrieval.

Graph based data such as social networks or knowledge graphs are represented as lower dimensional vectors using node embeddings. Graph embeddings capture the structural and relational information of the graph wherein nodes represent entities and edges represent relationships, enabling generative models to perform tasks like node classification and link prediction. Audio embeddings are used to represent audio signals as feature vectors in a lower dimensional space. Techniques like CNNs or RNNs are used to extract audio embeddings that capture spectral and temporal features. These embeddings enable generative models to generate audio, perform speech synthesis, or classify audio signals.

3.2 Variational Autoencoders to generate new data

Variational Autoencoders (VAEs) [7], use an encoder-decoder architecture to generate new data, particularly for image and video generation. VAEs consist of two neural networks typically referred to as the encoder and decoder. When given an input, an encoder converts it into a smaller, more dense latent space representation of the data. This is achieved through a series of convolutional or fully connected layers that progressively downsample the input data while extracting relevant features. The compressed representation preserves the information needed for a decoder to reconstruct the original input data, while discarding any irrelevant information.

Unlike traditional autoencoders, where the latent space is deterministic, VAEs model the latent space as a probability distribution, typically a Gaussian distribution. This is followed by the sampling step, where random samples are drawn from the learned probability distribution in the latent space. These samples represent different points in the latent space and correspond to different variations or attributes of the input data. The decoder takes the samples from the latent space and reconstructs them into output data, such as images or videos. Like the encoder, the decoder consists of a series of convolutional or fully connected layers that upsample the latent samples while reconstructing the original data. The encoder and decoder work together to learn an efficient and simple latent data representation. This allows generative AI to easily sample new latent representations that can be mapped through the decoder to generate novel data. While VAEs can generate outputs such as images faster, the images generated by them are not as detailed as those of diffusion models.

VAEs are trained using a variational inference approach, where the objective is to minimize the reconstruction error (reconstruction loss) while simultaneously maximizing the agreement between the learned latent distribution and a prior distribution (e.g., Gaussian distribution). This is typically achieved through the minimization of the Kullback-Leibler divergence between the learned distribution and the prior distribution.

By learning a continuous and probabilistic representation of the input data in the latent space, VAEs enable the generation of new data samples that capture the underlying structure and variability of the original data. This makes them particularly well-suited for tasks like image and video generation in generative AI.

3.3 Generative Adversarial Networks to generate new data

Generative Adversarial Networks (GANs) [8] are a class of deep learning models that pit two neural networks against each other: a generator that generates new examples and a discriminator that learns to distinguish the generated content as either real (from the domain) or fake (generated). These networks are trained simultaneously in a competitive manner such that they get smarter; the generator produces better content, and the discriminator gets better at spotting the generated content. This procedure repeats, pushing both to continually improve after every iteration until the generated content is indistinguishable from the existing content.

The generator takes random noise or a latent vector as input and learns to generate synthetic data samples, such as images or video frames. In the case of video generation, the generator produces sequences of frames that resemble real video footage. It uses CNNs or recurrent neural networks (RNNs) to transform the input noise into realistic data samples.

The discriminator is a binary classifier that learns to distinguish between real data samples (i.e., actual video frames) and fake data samples generated by the generator. It is trained on a combination of real and generated data samples and learns to assign high probabilities to real samples and low probabilities to generated samples. The discriminator is typically implemented as a convolutional neural network that outputs a probability score indicating the likelihood that a given sample is real.

During training, the generator and discriminator are trained simultaneously in a min-max game. The generator aims to produce data samples that are indistinguishable from real samples to fool the discriminator, while the discriminator aims to correctly classify real and fake samples. This adversarial training process encourages the generator to improve its ability to generate realistic data samples over time. The training of GANs is guided by a loss function that balances the objectives of the generator and discriminator. The generator seeks to minimize the discriminator's ability to distinguish between real and fake samples, while the discriminator aims to maximize its classification accuracy. This loss function is typically formulated as a binary cross-entropy loss or a variant thereof.

Once the GAN is trained, the generator can be used to generate new data samples by sampling random noise from a latent space and passing it through the generator network. In the case of video generation, the generator outputs sequences of frames that form a coherent and realistic video sequence. While GANs can provide high-quality samples and generate outputs quickly, the sample diversity is weak, therefore making GANs better suited for domain-specific data generation.

3.4 Diffusion Models to generate detailed quality images

Diffusion models [9], also known as denoising diffusion probabilistic models (DDPMs), are generative models that generate high-quality images with intricate details. These models operate by iteratively adding and removing noise levels to an initial image, gradually refining it to produce a realistic output. Vectors in latent space are determined through a two-step training process. The two steps are forward diffusion and reverse diffusion. The forward diffusion process slowly adds random noise to training data, while the reverse process reverses the noise to reconstruct the data samples. Novel data can be generated by running the reverse denoising process starting from entirely random noise.

The process begins with an initial image, often a noise image or a low-resolution version of the target image. The image undergoes a series of diffusion steps, where noise is gradually added to or removed from the image at each step. The diffusion process is controlled by a diffusion coefficient, which determines the rate at which noise is added or removed. To maintain image quality and resolution, the image is periodically upsampled during the diffusion process, allowing for finer details to be captured.

At each diffusion step, noise is applied to the image in a controlled manner. This noise can take various forms, such as Gaussian noise or structured noise patterns, and is typically added or removed using diffusion equations. As the diffusion process progresses, the image gradually evolves, with noise levels decreasing and image details becoming more refined. This refinement process helps to generate images with high levels of detail and realism.

Once the diffusion process is complete, the final image is obtained, representing a high-quality, detailed output that closely resembles the target image. A diffusion model can take longer to train than a VAE model or a GAN model, but due to its two-step process, hundreds of layers can be trained, resulting in highest-quality output that can rival those created by VAE or GAN.

Also, diffusion models are categorized as foundation models because they are large-scale, offer high-quality outputs, are flexible, and are considered best for generalized use cases. However, because of the reverse sampling process, running foundation models is a slow and lengthy process.

3.5 Transformers for Large Language Models

Transformers have emerged as a dominant architecture for large language models (LLMs) in generative AI. These models leverage the mechanisms of transformer architecture to process and generate text data with remarkable fluency and coherence. These mechanisms (self-attention and positional encodings) help represent time and allow for the algorithm to focus on how words relate to each other over long distances, making transformers particularly adept for text-based generative AI applications. A self-attention layer assigns a weight to each part of an input. The weight signifies the importance of that input in context to the rest of the input. Positional encoding is a representation of the order in which input words occur. A transformer is made up of multiple transformer blocks, also known as layers. For example, a transformer has self-attention layers, feed-forward layers, and normalization layers, all working together to decipher and predict streams of tokenized data, which could include text, protein sequences, or even patches of images.

A paper [10], presented at the Neural Information Processing Systems conference in 2017, introduced the transformer architecture as a novel approach for sequence transduction tasks, such as machine translation. Transformers, with their self-attention mechanism, have since become a cornerstone in the development of LLMs for generative AI, enabling models like Generative Pre-Trained Transformers (GPTs) to achieve remarkable performance in understanding and generating human-like text. GPTs undergo extensive pre-training on vast amounts of text data using unsupervised learning techniques. During pre-training, the model learns to understand the structure, semantics, and context of language by predicting the next word in a sequence given the preceding context. Transformers facilitate efficient and scalable pre-training, allowing models to learn rich representations of language.

After pre-training, the language model can be fine-tuned on specific downstream tasks in generative AI, such as text generation, summarization, translation, and dialogue systems. Fine-tuning involves adapting the pre-trained model to the nuances of the target task by fine-tuning its parameters on task-specific datasets. Transformers provide a flexible and adaptable framework for fine-tuning, enabling models to achieve state-of-the-art performance on various generative AI tasks.

Transformers excel at text generation tasks in generative AI, where they produce coherent and contextually relevant text based on a given prompt or input. By leveraging the transformer architecture's self-attention mechanism, LLMs can capture long-range dependencies in text data and generate fluent and diverse outputs. Transformers like GPT are widely used for text generation applications, including story generation, poem generation, code generation, and more.

Transformers are also employed in dialogue systems for generating conversational responses in human-machine interactions. Models like GPT and its variants, such as LaMDA (Language Model for Dialogue Applications), are specifically designed for dialogue generation tasks. These models can understand and generate contextually relevant responses in conversations, enabling more engaging and natural interactions with users.

Transformers are effective for summarization and translation tasks in generative AI, where they condense or translate text data while preserving its meaning and coherence. LLMs like GPT can generate concise summaries of long documents or articles, as well as translate text between different languages with high accuracy and fluency. The versatility, scalability, and effectiveness of transformers have made them indispensable tools for researchers and practitioners in the field of natural language processing and generative AI.

3.6 Neural Radiance Fields for generating 3D content from 2D images

Neural Radiance Fields (NeRF) represent a groundbreaking approach in generative AI for generating 3D content from 2D images [11]. These models leverage neural networks to learn a volumetric representation of a scene, capturing its geometry and appearance directly from a set of 2D images. NeRF models represent a scene as a continuous 3D volume, with each point in the volume associated with a radiance value. This radiance value represents the color and intensity of light emitted or reflected from that point in the scene. By learning the radiance field of the scene, NeRF models can generate photorealistic 3D renderings from novel viewpoints.

NeRF models typically consist of two neural networks: a volume rendering network and a positional encoding network. The volume rendering network takes a 3D point in space and predicts its radiance value based on the appearance of the scene observed in 2D images. The positional encoding network encodes the 3D coordinates of points in space to facilitate learning of the radiance field.

NeRF models are trained using a set of 2D images captured from different viewpoints of the scene. During training, the model learns to predict the radiance values of 3D points in the scene that correspond to the observed 2D image pixels. This involves optimizing the parameters of the volume rendering and positional encoding networks to minimize the discrepancy between predicted and observed radiance values.

Once trained, NeRF models can generate high-resolution 3D renderings of the scene from novel viewpoints. By querying the learned radiance field at different points in space along the rays cast from the camera, NeRF can produce detailed and realistic images that capture the geometry and appearance of the scene.

Neural Radiance Fields have demonstrated impressive capabilities in generating photorealistic 3D content from 2D images, including scenes with complex geometry, lighting, and textures. They have applications in various fields, including computer graphics, virtual reality, augmented reality, and robotics, and continue to advance the frontier of generative AI.

4. Emerging Trends in Generative AI

In generative AI, few-shot learning, zero-shot learning, and multimodal generation techniques play crucial roles in expanding the capabilities of models to generate diverse and high-quality outputs with limited training data or explicit supervision. These techniques have garnered significant attention in recent years as they enable models to generalize, adapt, and generate diverse outputs with limited supervision or input data, creating new possibilities for creative content generation, cross-modal understanding, and adaptive learning systems.

4.1 Few-shot learning

Few-shot learning techniques enable generative AI models to learn from a small number of training examples per class or task and generalize to new scenarios with similar characteristics. Instead of relying on large amounts of labelled data for each class, algorithms aim to learn a rich and generalizable representation of the underlying data distribution, enabling the model to adapt quickly to new tasks or classes. Few-shot learning techniques include meta-learning, transfer learning, and data augmentation, which help models generalize from limited data by leveraging knowledge from related tasks or domains.

Few-shot learning has applications in various domains where labelled data is scarce or costly to obtain, such as medical imaging, natural language processing, and computer vision. In image generation tasks, few-shot learning allows generative models to create new images in a particular style or category with only a few reference examples provided as input. Few-shot learning empowers generative AI systems to adapt quickly to new tasks, styles, or preferences, making them more versatile and applicable in real-world scenarios with limited labelled data.

4.2 Zero-shot learning

Zero-shot learning enables generative AI models to generate outputs for classes or concepts that were not seen during training. In zero-shot learning, models are trained to associate input data with auxiliary information or semantic embeddings, enabling them to generalize to unseen classes based on their semantic relationships with seen classes. Zero-shot learning methods often involve attribute-based classifiers, semantic embedding spaces, and generative models that infer relationships between classes and perform classification without direct supervision.

Zero-shot learning is useful in scenarios where the classes of interest may evolve over time, or where it is impractical to collect labelled examples for all possible classes. Applications include image recognition, natural language understanding, and cross-domain transfer learning.

4.3 Multimodal Generation

Multimodal generation refers to the generation of diverse and coherent outputs across multiple modalities, such as images, text, and audio. Multimodal generation models combine different types of data inputs and generate corresponding outputs that are consistent and complementary across modalities. Techniques for multimodal generation include encoder-decoder architectures, attention mechanisms, and adversarial training, which enable models to learn complex relationships between different modalities and produce coherent multimodal outputs.

Multimodal generation has applications in areas such as image captioning, visual storytelling, video synthesis, and cross-modal translation, where generating diverse and semantically meaningful outputs across multiple modalities is desired. In image captioning, multimodal generation models can generate descriptive captions for images, ensuring that the generated text aligns with the visual content.

5. Applications of Generative AI

Academic research in generative AI has been pursued for a very long time now, but its only recently that the applications of generative AI have broken into the mainstream. Generative AI models can take inputs such as text, image, audio, video, and code and generate new content into any of the modalities mentioned. For example, it can turn text inputs into an image, turn an image into a song, or turn video into text. Figure 3 represents some of the prominent use cases.

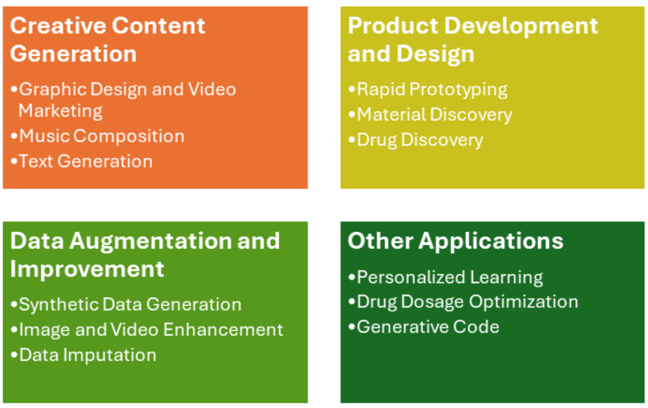

Figure 3: Generative AI use cases

5.1 Language

Language is compositional, which means that words can be combined to form new meanings and expressions. All human activities such as answering questions, solving math problems, providing instructions, using tools, and even playing games are described by language.

Advancement in language-based application domain of generative AI models is revolutionary. Language modelling exhibits the ability to encode, understand, and unify many different human, computer, and domain-specific languages. Language models can be trained to use tools such as web search, run programs, navigate robots, and write code. The most popular examples of language-based generative models are the LLMs, which are being leveraged for a wide variety of tasks, including essay generation, code development, translation, and even understanding genetic sequences. Language models are trained to understand text or sequence of words using GPTs. Figure 4 shows that the ability of the first GPT [10] was limited to performing simple text processing such as correctly identify which part of speech a particular word is in a sentence.

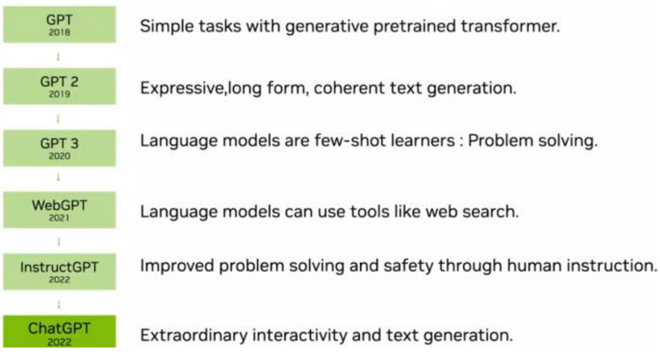

Figure 4: Evolution of GPT Language Models [12]

The successor to GPT, GPT 2 [13] could generate long form, coherent, and consistent sentences. GPT 3 [14] demonstrated that a language model can be used for more than just understanding text and became capable of problem solving. Following its success, a series of interfaces emerged to captivate the world: WebGPT [15] capable of accurately answering open-ended questions using a text-based web browser; InstructGPT [16] capable of following English instructions; and ChatGPT [17] capable of interaction in a conversational way, answering followup questions, admitting its mistakes, challenging incorrect premises, and rejecting inappropriate requests.

OpenAI’s latest milestone, GPT-4 [15], accepts image and text inputs, emits text outputs, that exhibit human-level performance on various professional and academic benchmarks. For example, it passes a simulated bar exam with a score around the top 10% of test takers.

Enterprises may either use available LLMs out-of-the-box or fine-tune them for specific use cases. An LLM fine-tuned on textbooks can assist teachers in a classroom setting. An LLM fine-tuned on a company’s call centre knowledge will provide answers to questions and perform basic troubleshooting. An LLM fine-tuned on legislation documents can assist lawmakers in policy making. An LLM fine-tuned on medical literature can assist healthcare professionals in patient health diagnosis.

In order to fine-tune language models to become less biased and better at solving specific tasks, enterprises need to collect and fine-tune an existing foundational model with a relatively small sample of demonstration data for specific tasks in supervised way. Reinforcement learning and reward system may be used to improve model outputs. Human feedback such as rating or ranking the model outputs may be used. In machine learning, reinforcement learning from human feedback (RLHF) or reinforcement learning from human preferences is a technique that trains a "reward model" directly from human feedback and uses the model as a reward function to optimize an agent's policy using reinforcement learning through an optimization algorithm like Proximal Policy Optimization [19].

5.2 Audio

Music, audio, and speech are emerging fields within generative AI that offer diverse opportunities for innovation and creativity. Generative AI techniques are increasingly being applied to music composition, enabling the creation of original melodies, harmonies, and rhythms. Deep learning models, such as RNNs and GANs, are used to generate music in various styles and genres. Music generation AI systems can compose entire songs or assist musicians by providing creative suggestions and variations. Applications include music production, automated composition, and personalized music recommendation systems.

Generative models are used to synthesize realistic audio signals, such as sounds of musical instruments, environmental noises, and human voices. Techniques like WaveGAN, SampleRNN, and WaveNet employ deep learning architectures to generate high-fidelity audio waveforms. Audio synthesis AI systems can create immersive soundscapes for virtual reality environments, video games, and multimedia applications. Applications include sound design, audio effects processing, and speech synthesis. Generative AI is used for both speech synthesis (text-to-speech) and speech recognition (speech-to-text) tasks. State-of-the-art models, such as Tacotron, Deep Voice, and Transformer-based architectures, can generate natural-sounding speech from text inputs. Speech recognition systems leverage deep learning techniques, including RNNs and CNNs, to transcribe spoken language into text. Applications include virtual assistants, voice-controlled devices, language translation services, and accessibility tools for individuals with speech impairments.

5.3 Visuals

One of the most popular applications of generative AI is within the realm of images. Generative AI techniques such as GANs and VAEs can create realistic 3D renderings from 2D images, synthesize novel 3D shapes, and generate 3D avatars for virtual environments and games. There’s flexibility in generating images with different aesthetic styles, as well as techniques for editing and modifying generated visuals. Generative AI techniques are used to generate video content, including deepfake videos, video synthesis, and video completion. These models can manipulate and transform existing videos, generate realistic human motions, and synthesize dynamic scenes based on textual descriptions or input images.

Generative AI models can create graphs, charts, diagrams, and infographics based on input data or textual descriptions, facilitating the creation of visually appealing and informative content for data visualization, presentations, and educational materials. Generative AI is employed to create diverse and customizable avatars and digital characters for various purposes, including gaming, virtual reality, and online communication platforms.

Generative AI techniques enable artistic style transfer, where the style of one image or artwork is applied to another image while preserving its content. Models can transfer the artistic style of famous paintings, photographs, or illustrations to user-provided images, creating visually stunning and unique compositions. Generative AI models are used for image editing, enhancement, and restoration tasks, such as super-resolution, denoising, and colorization.

5.4 Synthetic data

Synthetic data is extremely useful to train AI models when data doesn't exist, is restricted, or is simply unable to address corner cases with the highest accuracy. The development of synthetic data through generative models is perhaps one of the most impactful solutions for overcoming the data challenges of many enterprises. It spans all modalities and use cases and is possible through a process called label efficient learning. Generative AI models can reduce labelling costs by either automatically producing additional augmented training data or by learning an internal representation of the data that facilitates training AI models with less labelled data.

5.5 Other applications

The impact of generative models is wide-reaching, and its applications are only growing. There are numerous examples of how generative AI is helping to advance and transform the fields of transportation, natural sciences, and entertainment.

In the automotive industry, generative AI is expected to help create 3D worlds and models for simulations and car development. Synthetic data is also being used to train autonomous vehicles. Being able to road test the abilities of an autonomous vehicle in a realistic 3D world improves safety, efficiency, and flexibility while decreasing risk and overhead.

The field of natural sciences greatly benefits from generative AI. In the healthcare industry, generative models can aid in medical research by developing new protein sequences to aid in drug discovery. Practitioners can also benefit from the automation of tasks such as scribing, medical coding, medical imaging, and genomic analysis.

Meanwhile, in the weather industry, generative models can be used to create simulations of the planet and help with accurate weather forecasting and natural disaster prediction. These applications can help to create safer environments for the general population and allow scientists to predict and better prepare for natural disasters.

All aspects of the entertainment industry, from video games to film, animation, world building, and virtual reality, are able to leverage generative AI models to help streamline their content creation process. Creators are using generative models as a tool to help supplement their creativity and work.

6. Challenges of Generative AI

As an evolving space, generative models are still considered to be in their nascent stages. Nonetheless, the pace of change induced by a technology has never been faster. There are some major challenges, as well as some significant areas for growth.

For enterprises seeking to tap this valuable technology, scaling up to generative AI has unique talent-related challenges. In addition to learning new terminology such as reinforcement learning and CNN, generative AI demands advanced analytics capabilities and AI expertise such as prompt engineering and database curation. Thus, enterprises will need to enhance their capabilities through some mix of training and recruiting. An effective strategic road map for a generative AI scale-up must include [2]:

- Vision, alignment, and commitment from senior leadership and business-unit-level accountability for delivering results.
- A list of priority domains where several related use cases can be built as sometimes traditional analytical AI offers better solution than generative AI.
- Assessment of enabling capabilities, including talent, agile operating model, technology, and data.
- A thorough scale-up plan that sequences when and how to tackle each domain and build enabling capabilities by augmenting existing ones or acquiring new ones.

Generative AI models can boast billions of parameters and require fast and efficient data pipelines to train. Significant capital investment, technical expertise, and large-scale compute infrastructure are necessary to maintain and develop generative models. For example, diffusion models could require millions or billions of images to train. Moreover, to train such large datasets, massive compute power is needed, and AI practitioners must be able to procure and leverage hundreds of GPUs to train their models. Also, some domains do not have enough high-quality and unbiased data to train a model. As an example, few 3D assets exist and they're expensive to develop. Such areas will require significant resources to evolve and mature. Further compounding the issue of a lack of high-quality data, many organizations struggle to get a commercial license to use existing datasets or to build bespoke datasets to train generative models. Creating explainable generative AI is also a challenge as it relies on neural networks with billions of parameters, thereby complicating attempts to explain to users how any given answer is produced.

Thus, in addition to old-school change management skills, up-front senior leadership alignment and sponsorship, business unit accountability for results, value-centered use cases, and clear targets, generative AI scale=up requires massive scale-up of compute infrastructure as well. Many companies such as NVIDIA, Cohere, and Microsoft have a goal to support the continued growth and development of generative AI models with services and tools to help solve these issues. These products and platforms abstract away the complexities of setting up the models and running them at scale.

6.1 Impact on employment

Adoption of generative AI may also have a major impact on employment markets around the world exposing around 300 million full-time jobs to automation [3]. A Goldman Sachs report estimates that about two-thirds of U.S. occupations are exposed to some degree of automation by generative AI, and about half of the workload of a quarter of these occupations could be replaced (c.f. Figure 5). Further, the report assures that not all that automated work will translate into layoffs; workload is more likely to be complemented rather than substituted by AI.

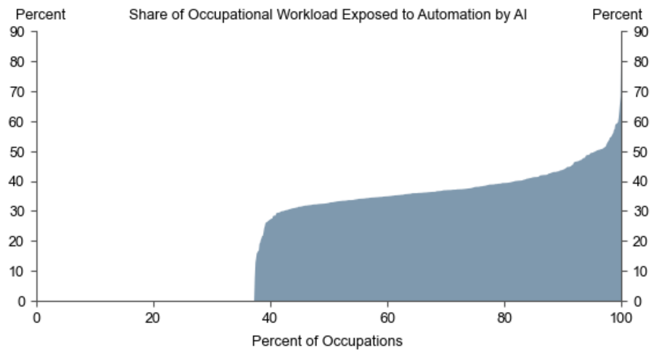

Figure 5: Share of occupational workload exposed to automation by AI [3]

6.2 Risks

Generative AI also brings along a unique set of risks. These relate to the quality of results, potential for misuse and abuse, and the potential to disrupt existing business models. Generative AI is prone to algorithmic bias caused by imperfect training data or engineering decisions during the development and deployment phases. Models may learn and perpetuate these biases in the generated content. Careful selection of training data to avoid including toxic or biased content is crucial. Prompt engineering can help mitigate bias and undesirable outputs by providing context, constraints, or guidelines that steer the model away from generating sensitive or harmful content. Instead of using off-the-shelf solutions, organizations should consider using smaller, specialized or customized models based on their own data to fit their needs and minimize biases.

Digitally forged images or videos called deepfakes have raised security concerns. Deepfake is a type of synthetic media, created using deep learning techniques, particularly GANs, which replace the likeness of one person with another, representing someone doing or saying something they didn't do or say. The greatest danger posed by deepfakes is their ability to spread false information that appears to come from trusted sources. Generative AI development community has been advocating for a harmonized international regulatory control over the technology's development.

Organizations deploying generative AI models should be mindful of the reputational and legal risks involved in unintentionally publishing biased, offensive, or copyrighted content. Training data and model outputs can generate infringement on copyrighted, trademarked, patented, or otherwise legally protected materials. Extensive use of personal or sensitive information during model training gives rise to privacy concerns. Potentially disproportionate and negative impacts on particular groups and local communities is also a risk in deploying biased generative AI models.

Generative AI may produce different answers to the same prompts, impeding users' ability to assess the accuracy and reliability of outputs. Also, there is a risk of generative AI 'hallucinations', wherein models produce illogical and weird answers or outputs. This calls for intervention by human subject matter experts for model output validation. Finally, the ESG (Environmental, Social, and Governance) fallout is also a concern. Training and deployment of foundation models may increase carbon emissions and exceed ESG expectations.

7. Conclusion

Generative AI is making inroads in business applications, streamlining workflows, automating routine tasks and leading to a new generation of applications. It is improving the routine efficiency of knowledge workers, helping scientists develop drugs faster and accelerating the development of software code. Generative AI's ability to create content that is indistinguishable from human created content, males it useful in areas like entertainment, advertising, and creative arts. Generative AI algorithms can be used to create synthetic data that can be used to train and evaluate other AI algorithms in natural language processing and computer vision, thereby improving their efficiency and accuracy. Generative AI algorithms can be used to explore and analyse complex data in new ways, allowing businesses and researchers to uncover hidden patterns and trends that may not be apparent from the raw data alone.

However, the speed of generative AI's emergence as a critical capability has left businesses little time to prepare for the effects on their people, and for preparing to upskill employees or attract the talent they'll need to keep pace. Also, responsible use of generative AI is a concern. Issues such as model interpretability and unbiased decision making must be comprehensively tackled before scaling any application to generative AI.

References

[1] "The Lens - Free & Open Patent and Scholarly Search," [Online]. Available: https://www.lens.org/. [Accessed 7 April 2024].

[2] V. Kamalnath, L. Lerner, J. Moon, G. Sari, V. Sohoni and S. Zhang, "Capturing the full value of generative AI in banking," McKinsey & Company, 2023.

[3] “Generative AI could raise global GDP by 7%,” Goldman Sachs, 2023.
[4] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and D. Jeffery, “Distributed Representations of Words and Phrases and their Compositionality,” in Advances in Neural Information Processing Systems, 2013.
[5] J. Pennington, R. Socher and C. D., “GloVe: Global Vectors for Word Representation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014.
[6] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, “Enriching Word Vectors with Subword Information,” Transactions of the Association for Computational Linguistics, vol. 5, p. 135–146, 2017.
[7] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” in Proceedings of the International Conference on Learning Representations, 2014.
[8] I. J. Goodfellow, J. Pouget-Abadie and M. Mirza, “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, 2014.
[9] J. Ho, A. Jain and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in Proceedings of the 38th International Conference on Machine Learning, 2021.
[10] A. Vaswani , N. Shazeer, N. Parmar, J. Uszkoreit and L. Jones, “Attention is All you Need,” in Advances in Neural Information Processing Systems, 2017.
[11] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[12] “Generative AI Explained,” NVIDIA, [Online]. Available: https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-07+V1. [Accessed 9 April 2024].
[13] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” OpenAI, 2019.
[14] T. B. Brown, B. Mann and N. Ryder, “Language Models are Few-Shot Learners,” OpenAI, 2020.
[15] J. Hilton, R. Nakano, S. Balaji and J. Schulman, “WebGPT: Improving the factual accuracy of language models through web browsing,” OpenAI, 2021.
[16] R. Lowe and J. Leike, “Aligning language models to follow instructions,” OpenAI, 2022.
[17] OpenAI, “Introducing ChatGPT,” OpenAI, 2022.
[18] OpenAI, “GPT-4,” OpenAI, 2023.
[19] W. B. Knox and P. Stone, “Interactively shaping agents via human reinforcement: the TAMER framework,” in Proceedings of the fifth international conference on Knowledge capture, 2009.

Author Profile

Dr. Brahmaleen Kaur Sidhu earned her Ph.D. degree in Faculty of Engineering and Technology from Punjabi University, Punjab, India, M.Tech. degree in Computer Science and Engineering from the Punjab Technical University, Punjab, India, and B.Tech. degree in Computer Science and Engineering from Punjabi University. She is currently working as Assistant Professor in the Department of Computer Science and Engineering, Punjabi University and has around 18 years of teaching experience. Her research interests include software architecture, software evolution, software quality, refactoring, model-driven development, data science and machine learning. She has around 80 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE, and a book titled "A Handbook of Reinforcement Learning" published in 2023. She has been awarded the “International Innovative Educator Award 2021" and is listed in “100 Eminent Academicians of 2021” by International Institute of Organized Research.