IBM ML/DL MOD4

Generative AI

WHAT IS GENERATIVE AI ?

Generative AI is a type of artificial intelligence that creates new, original content that people have never seen before.

- Most AI systems are discriminative AI models, which predict and classify data.

- In contrast, generative AI models are a type of deep learning AI system that uses algorithms to generate content based on a submitted prompt, thus the name of generative AI.

For example, a discriminative model could tell a bicycle from a truck and a generative model could generate a new image that looks like a bicycle.

So, generative AI's distinction from other AI systems is its ability to produce content that is new and considered creative, such as images, videos, music, synthetic data, essays, answers to questions, and more.

Here's a quick example of how easily generative AI can create new images. A person typed "butterflies, flowers, mountains" in the text prompt of the image generating tool called Free Image Generator. The tool instantly created the following four images!

Generative AI's output has gained global attention due to the uniqueness, high quality, and speed of content generation. Generative AI is a revolutionary change that's as radical to global humanity as the invention of electricity and automobiles. Let's see how generative AI works.

How does generative AI work?

Think of generative AI as a virtual artist. Just like a human artist, it needs inspiration and tools to create something unique. Instead of using paint and canvas, however, generative AI uses algorithms and data sets.

Here's the overall generative AI process.

- First, a person feeds the AI a large amount of data. This could be anything from images and sounds to text and numbers.

- Then, the AI analyzes this data, looking for patterns and relationships between the different pieces of information. The neural network is trained on a dataset of examples of the type of output it is intended to generate, such as images or text. During the training process, the neural network learns to identify patterns and relationships in the input data and use them to generate new outputs that are similar to, but not identical to, the examples it was trained on.

- Next, the AI uses what it has learned to create something new. The neural network generates new outputs by inputting a random seed value. The seed value serves as the starting point for the generation process. The neural network processes the seed value and generates a new output that is based on the patterns and relationships it learned during training. For example, if someone gave the AI a set of images of dogs, it might use its knowledge of different dog breeds to create an image of a new dog that doesn't exist in real life.

Generative AI can also complete more complex tasks, like writing stories or composing music. In these cases, the AI analyzes patterns in language or music to create something entirely new.

Of course, not all generative AI is perfect. Just like human artists, sometimes the results can be a bit strange or unexpected. However, as this technology improves, you can expect to see even more impressive creations from generative AI in the future!

Types of generative AI models

Let's explore the three primary types of generative AI models:

- Variational autoencoder (VAE)

Variational Autoencoder (VAE) −

Think of **variational autoencoder (VAE)** models as a skilled artist who can look at a painting, quickly sketch a simplified version of it, and then recreate a new painting using only that simplified sketch as a reference. The artist is capturing the essential elements of the painting and then using them to create a new work of art.

VAEs use a similar process. The "encoder" network compresses the input data into a lower-dimensional representation and the "decoder" network reconstructs the original data from this compressed representation. This allows VAEs to capture the underlying structure and patterns in the data, which can then generate new, similar data.


Encoder Decoder
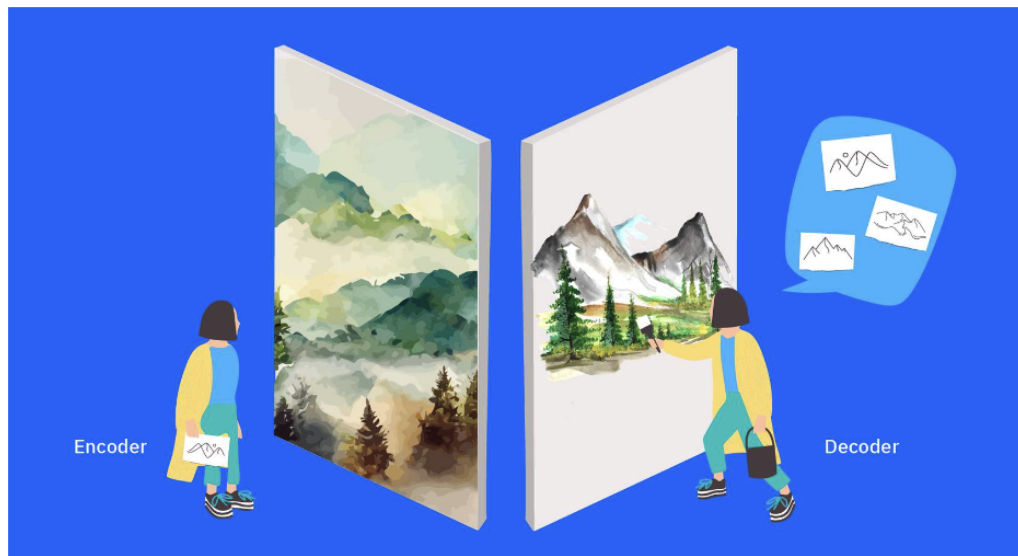
- Generative adversarial network (GAN)

Generative adversarial network (GAN) —

Think of a **generative adversarial network (GAN)** model as a competition between a skilled forger (the generator) and a talented art critic (the discriminator). The forger creates fake paintings, while the critic tries to determine whether each painting is genuine or a forgery. As the forger improves their technique, the critic becomes more discerning, and this cycle continues until the forger can create near-perfect forgeries.

In GANs, the generator creates new data, while the discriminator evaluates the quality of the generated data. The generator tries to create data that is realistic enough to fool the discriminator, while the discriminator learns to better distinguish between real and generated data. This competition leads to the generator creating increasingly realistic content.
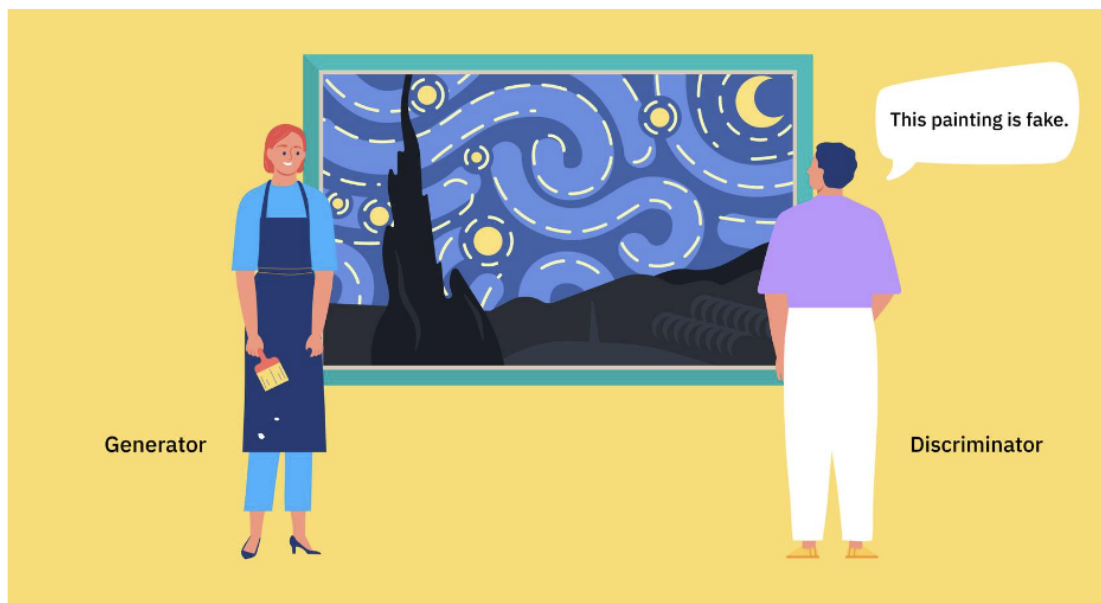
- Autoregressive

Autoregressive   —

Imagine an **autoregressive** model as a skilled storyteller who listens to the beginning of a story and then continues it by predicting what comes next based on the words and events that have occurred so far. The storyteller uses their knowledge of language, grammar, and storytelling conventions to create a coherent and engaging continuation of the story.

Autoregressive models generate new content by predicting the next element in a sequence based on the previous elements. They are particularly well-suited for generating text because they can model the conditional probabilities of words and characters in a sentence.

Examples of generative AI applications

CHATGPT	IBM WATSON DISCOVERY	DALL-E AND DALL-E 2	BARD

OpenAI launched ChatGPT, an AI chatbot, in November 2022. Able to interact by using conversational natural language, this AI tool goes beyond the traditional search engine responses of simply listing related results. Instead, ChatGPT follows instructions given in the prompt and provides a detailed response. For example, with ChatGPT, a person can enter a text prompt of "Write a poem about cats" and the result will be a poem about cats, rather than a listing of websites about cats.

CHATGPT	IBM WATSON DISCOVERY	DALL-E AND DALL-E 2	BARD

IBM Watson Discovery uses foundational technologies, such as large language models (LLMs), to obtain insights from what it is that we don't know that we don't know. It's used widely in genomic research discovering what amino acids might lurk inside the protein not known before and unearth the relationship of various actors or entities for government security related questions.

CHATGPT	IBM WATSON DISCOVERY	DALL-E AND DALL-E 2	BARD

Developed by OpenAI, these are generative AI models that use natural language text input to generate digital images. The first version of DALL-E could only render AI-created images in a cartoonish fashion, but the latest version can produce much more realistic images due to improved image processing algorithms.

CHATGPT	IBM WATSON DISCOVERY	DALL-E AND DALL-E 2	BARD

Bard is a generative AI tool that Google launched in an initial, limited capacity in March 2023. Bard is founded on Google's Bidirectional Encoder Representations from Transformers (BERT) model. It isn't generative AI; rather, Google developed it for natural language processing (NLP), especially for its capacity to interpret the nuances of a user's search words. Bard builds on BERT's capability of natural language interactions with Bard's generative AI capability to generate new content. For example, musicians can use Bard to compose music and lyrics.

Industry uses of generative AI

Generative AI is already impacting, and will increasingly impact, a wide range of industries, including sports, entertainment, healthcare, retail, banking, manufacturing, engineering, security, media, agriculture, and the list goes on and is continuing to expand.

Review these real-world examples of generative AI in practice.

Sports

In sports, generative AI can help improve athletic performance and enhance fan engagement.

One application of generative AI in sports is the creation of personalized workout plans. By analyzing an athlete's biometric data, generative AI can create customized workout plans tailored to the athlete's fitness level and goals. This technology can improve athletic performance by providing athletes with targeted and efficient training regimes.

Another application of generative AI in sports, around fan engagement, is the creation of realistic 3D models of athletes for use in sports video games and virtual reality experiences. By using generative AI, game developers can create highly realistic and personalized virtual athletes that can interact with users in real-time. This technology can enhance fan engagement by providing users with immersive and engaging sports experiences.

Generative AI can analyze sports data and identify patterns and trends that can inform coaching strategies and player selection.

Overall, generative AI has the potential to transform the way people play, coach, and experience sports, enabling new levels of personalization, efficiency, and engagement.

Use case

The Miami Dolphins, an American football team in the United States, are using generative AI to improve player performance and prevent injury. They are collaborating with a company called Blue River Technology to develop a computer vision system that uses generative AI to analyze video footage of players and identify areas for improvement. The system

analyzes players' body movements and identifies areas where they could be at risk for injury. By using generative AI in this way, the Miami Dolphins can identify potential issues before they become serious problems, enabling them to take proactive measures to prevent injuries and improve player performance.

This technology has the potential to transform the way that sports teams approach player health and performance, improving the overall quality of play and reducing the risk of injury for athletes.

Entertainment

In the entertainment industry, generative AI can help create new and immersive experiences for users.

One application of generative AI in entertainment is the creation of a wide variety of highly realistic and personalized virtual characters and environments for video games and virtual reality experiences. For example, generative AI can create realistic facial expressions and body movements for virtual characters, enhancing the user experience and immersion.

Another application of generative AI in entertainment is the generation of personalized music playlists. By analyzing a user's listening history and preferences, generative AI can create unique and personalized music playlists that cater to each individual user's tastes.

This technology can enhance the user experience by creating highly personalized and engaging music experiences.

Use case

Amper Music uses generative AI to create original music for video games, films, and other multimedia projects. Their platform allows users to input parameters such as mood, genre, and tempo, and then generates a unique and original piece of music in real-time. It generates music using a generative model that is trained on a large dataset of musical patterns and structures. By using generative AI to create music, Amper Music can offer highly customizable and original compositions tailored to each individual project's needs.

This technology has the potential to transform the creation of music for multimedia projects, enabling more efficient and personalized music production.

Healthcare

In healthcare, generative AI can help improve disease diagnosis, medical imaging, and personalized medicine.

One application of generative AI is in the generation of synthetic medical images, which can train machine learning algorithms and improve disease diagnosis. For example, generative AI can generate synthetic medical images of rare diseases, which can be difficult to obtain in real life.

Another application of generative AI in healthcare is the creation of medical simulations. These simulations can help train healthcare professionals and reduce medical errors. By generating synthetic patient data, generative AI can also train predictive models and improve personalized medicine.

Generative AI has the potential to transform healthcare by enabling the creation of realistic medical simulations, improving disease diagnosis, and enhancing personalized medicine.

Use case

PathAI uses generative AI to improve the accuracy of disease diagnosis through the analysis of medical images. Specifically, PathAI has developed a deep learning algorithm that can accurately detect cancer cells in digital pathology images. This algorithm was trained on a large dataset of annotated pathology images using a variational autoencoder (VAE) generative model. The VAE learned the underlying structure of the pathology images to generate synthetic images that were like the real images in the training set. By training their algorithm on both real and synthetic images, PathAI was able to improve the accuracy of their cancer detection algorithm.

This technology has the potential to improve the accuracy and speed of cancer diagnosis, to ultimately improve patient outcomes.

Business

In business, generative AI can help improve decision-making, personalize customer experiences, and enhance operational efficiency.

One application of generative AI in business is the generation of synthetic data. By generating synthetic data, companies can augment their existing datasets and improve the accuracy

of predictive models. For example, a financial institution could generate synthetic financial data to train their predictive models and improve risk management.

Another application of generative AI in business is the creation of personalized product recommendations. By analyzing a customer's purchase history and preferences, generative AI can create personalized product recommendations tailored to each individual customer. This technology can improve customer satisfaction and increase sales.

People are also using generative AI to improve operational efficiency by creating synthetic data for testing and validation.

Generative AI has the potential to transform the way businesses operate by improving decision-making, enhancing customer experiences, and increasing efficiency.

Use case

Syntiant uses generative AI to develop low-power, high-performance deep learning chips for a range of industries, including smart home devices, wearables, and voice-controlled devices. Their deep learning chips are designed to process data on the device itself, rather than relying on cloud-based processing, which can reduce latency and improve privacy. To train their deep learning models, Syntiant uses a generative adversarial network (GAN) to create synthetic data by generating new data samples that are similar to the training data. By training their models on both real and synthetic data, Syntiant was able to improve the accuracy of their deep learning chips, making them more reliable and efficient.

> This technology has the potential to transform the way that companies design and manufacture deep learning chips, enabling a new generation of low-power, high-performance devices.

Limitations of generative AI

As you continue to dive deeper into the world of generative AI, it's crucial to consider the limitations and ethical concerns surrounding these groundbreaking technologies. With models like GPT-4 transforming the way content is created, from text and images to music, it's essential to strike a balance between their extraordinary potential and the need for responsible, equitable usage.

Explore these key challenges and ethical dilemmas that come with harnessing the power of generative AI.

Limitations of generative AI

- **Lack of originality:** Generative AI models rely on large datasets to learn and generate content. As a result, they might not create entirely original content but rather mimic patterns from their training data, which can lead to a lack of creativity and innovation.

- **Incompleteness:** While generative AI models are becoming increasingly sophisticated, they still struggle to understand the nuanced contexts and might generate incomplete or nonsensical content.

- **Bias:** Generative AI models can perpetuate existing biases present in their training data, leading to the generation of biased content that might reinforce stereotypes and discriminatory behavior.

- **Computational resources:** Training and deploying generative AI models require significant computational power, which can be expensive and contribute to environmental concerns such as energy consumption and carbon emissions.

Ethical concerns of generative AI

- **Misinformation and fake content:** Generative AI can create convincing fake content, like deepfakes or falsified news articles, which can lead to the spread of misinformation and have severe consequences for individuals and societies.

- **Intellectual property and copyright:** Generative AI can produce content that resembles copyrighted material. This raises questions about intellectual property rights and potential infringements.

- **Privacy:** Generative AI can create realistic images and text about individuals, potentially violating their privacy and causing harm to their reputation.

- **Loss of human touch:** As generative AI becomes more prevalent, there is a risk that the human touch will be lost in various creative domains, potentially leading to a decline in the appreciation of human-created art and culture.

- **Unemployment and job displacement:** The rise of generative AI might lead to job displacement in creative industries, as machines take over tasks previously performed by humans.

Overall, while generative AI has many potential applications and benefits, it's important to be aware of its limitations and potential pitfalls to use it effectively and ethically.

Reflection: What do you think?

Should students in school be permitted to use generative AI tools to complete their class assignments? Why or why not?