

Introduction to Generative AI

Hello! Welcome to *Modern Generative AI with ChatGPT and OpenAI Models!* In this book, we will explore the fascinating world of generative **Artificial Intelligence (AI)** and its groundbreaking applications. Generative AI has transformed the way we interact with machines, enabling computers to create, predict, and learn without explicit human instruction. With ChatGPT and OpenAI, we have witnessed unprecedented advances in natural language processing, image and video synthesis, and many other fields. Whether you are a curious beginner or an experienced practitioner, this guide will equip you with the knowledge and skills to navigate the exciting landscape of generative AI. So, let's dive in and start with some definitions of the context we are moving in.

This chapter provides an overview of the field of generative AI, which consists of creating new and unique data or content using **machine learning (ML)** algorithms.

It focuses on the applications of generative AI to various fields, such as image synthesis, text generation, and music composition, highlighting the potential of generative AI to revolutionize various industries. This introduction to generative AI will provide context for where this technology lives, as well as the knowledge to collocate it within the wide world of AI, ML, and **Deep Learning (DL)**. Then, we will dwell on the main areas of applications of generative AI with concrete examples and recent developments so that you can get familiar with the impact it may have on businesses and society in general.

Also, being aware of the research journey toward the current state of the art of generative AI will give you a better understanding of the foundations of recent developments and state-of-the-art models.

All this, we will cover with the following topics:

Understanding generative AI

Exploring the domains of generative AI

The history and current status of research on generative AI

By the end of this chapter, you will be familiar with the exciting world of generative AI, its applications, the research history behind it, and the current developments, which could have – and are currently having – a disruptive impact on businesses.

Introducing generative AI

AI has been making significant strides in recent years, and one of the areas that has seen considerable growth is generative AI. Generative AI is a subfield of AI and DL that focuses on generating new content, such as images, text, music, and video, by using algorithms and models that have been trained on existing data using ML techniques.

In order to better understand the relationship between AI, ML, DL, and generative AI, consider AI as the foundation, while ML, DL, and generative AI represent increasingly specialized and focused areas of study and application:

AI represents the broad field of creating systems that can perform tasks, showing human intelligence and ability and being able to interact with the ecosystem.

ML is a branch that focuses on creating algorithms and models that enable those systems to learn and improve themselves with time and training. ML models learn from existing data and automatically update their parameters as they *grow*.

DL is a sub-branch of ML, in the sense that it encompasses deep ML models. Those deep models are called **neural networks** and are particularly suitable in domains such as **computer vision** or **Natural Language Processing (NLP)**. When we talk about ML and DL models, we typically refer to discriminative models, whose aim is that of making predictions or inferencing patterns on top of data.

And finally, we get to generative AI, a further sub-branch of DL, which doesn't use deep Neural Networks to cluster, classify, or make predictions on existing data: it uses those powerful Neural Network models to generate brand new content, from images to natural language, from music to video.

The following figure shows how these areas of research are related to each other:

Artificial Intelligence

The ability of a machine to show human ability like reasoning, learning, such as creativity.



Machine Learning

The set of algorithms that make intelligent machines capable of improving with time and experience.



Deep Learning

A type of ML based on *deep* neural networks made of multiple layers of processing.



Generative AI



Figure 1.1 – Relationship between AI, ML, DL, and generative AI

Generative AI models can be trained on vast amounts of data and then they can generate new examples from scratch using patterns in that data. This generative process is different from discriminative models, which are trained to predict the class or label of a given example.

Domains of generative AI

In recent years, generative AI has made significant advancements and has expanded its applications to a wide range of domains, such as art, music, fashion, architecture, and many more. In some of them, it is indeed transforming the way we create, design, and understand the world around us. In others, it is improving and making existing processes and operations more efficient.

The fact that generative AI is used in many domains also implies that its models can deal with different kinds of data, from natural language to audio or images. Let us understand how generative AI models address different types of data and domains.

Text generation

One of the greatest applications of generative AI—and the one we are going to cover the most throughout this book—is its capability to produce new content in natural language. Indeed, generative AI algorithms can be used to generate new text, such as articles, poetry, and product descriptions.

For example, a language model such as GPT-3, developed by OpenAI, can be trained on large amounts of text data and then used to generate new, coherent, and grammatically correct text in different languages (both in terms of input and output), as well as extracting relevant features from text such as keywords, topics, or full summaries.

Here is an example of working with GPT-3:

V

Explain to a 5-year-old what a proton is with examples. Share also some references to existing documentation.



Okay, let's imagine that you have an apple. An apple is made up of tiny, tiny pieces called atoms. And in each atom, there is something called a proton. Protons are like the boss of the atom and they help to keep the atom together.



Think of the protons as little balls that stick together to make the apple. Without them, the apple wouldn't exist! And that's why protons are so important.

If you want to learn more about protons and atoms, you can check out this website called Kids Konnect. It has lots of information and pictures to help you understand:

<https://www.kidskonnect.com/science-facts/atoms-and-elements/>

You can also check out a video on YouTube called "The Incredible World of Atoms and Molecules for Kids" to learn about protons and other parts of an atom:

<https://www.youtube.com/watch?v=Kgc2HbK6QOQ>

⟳ Regenerate response



Figure 1.2 – Example of ChatGPT responding to a user prompt, also adding references

Next, we will move on to image generation.

Image generation

One of the earliest and most well-known examples of generative AI in image synthesis is the **Generative Adversarial Network (GAN)** architecture introduced in the 2014 paper by I. Goodfellow et al., *Generative Adversarial Networks*. The purpose of GANs is to generate realistic images that are indistinguishable from real images. This capability had several interesting business applications, such as generating synthetic

datasets for training computer vision models, generating realistic product images, and generating realistic images for virtual reality and augmented reality applications.

Here is an example of faces of people who do not exist since they are entirely generated by AI:



Figure 1.3 – Imaginary faces generated by GAN StyleGAN2 at <https://this-person-does-not-exist.com/en>

Then, in 2021, a new generative AI model was introduced in this field by OpenAI, **DALL-E**. Different from GANs, the DALL-E model is designed to generate images from descriptions in natural language (GANs take a random noise vector as input) and can generate a wide range of images, which may not look realistic but still depict the desired concepts.

DALL-E has great potential in creative industries such as advertising, product design, and fashion, among others, to create unique and creative images.

Here, you can see an example of DALL-E generating four images starting from a request in natural language:

Edit the detailed description

Surprise me

Upload



A picture of the philosopher Socrates as he had lived at the present time. He rides an electric scooter while messaging on his iPhone

Generate



Figure 1.4 – Images generated by DALL-E with a natural language prompt as input

Note that text and image generation can be combined to produce brand new materials. In recent years, widespread new AI tools have used this combination.

An example is Tome AI, a generative storytelling format that, among its capabilities, is also able to create slide shows from scratch, leveraging models such as DALL-E and GPT-3.

What is Generative AI?

Generative AI is a form of artificial intelligence that focuses on creating and developing data and models that can generate new content and ideas. This type of AI is used to create new solutions and products, allowing for more creative and innovative solutions to be developed.

Generative AI can be used to create new products and services, and to develop new technologies. It can also be used to improve existing products and services, and to create new ones. Generative AI can help businesses to increase their efficiency, reduce costs, and improve customer experience.

1 Unlocking the Potential of Generative AI

2 What Generative AI Can Do For Your Business

3 Generative AI Examples

4 Generative AI in Action

5 Generative AI Benefits

6 Generative AI Use Cases

7 Generative AI Applications

8 Themed Questions?

+

v Share

Figure 1.5 – A presentation about generative AI entirely generated by Tome, using an input in natural language

As you can see, the preceding AI tool was perfectly able to create a draft presentation just based on my short input in natural language.

Music generation

The first approaches to generative AI for music generation trace back to the 50s, with research in the field of algorithmic composition, a technique that uses algorithms to generate musical compositions. In fact, in 1957, Lejaren Hiller and Leonard Isaacson created the Illiac Suite for String Quartet (<https://www.youtube.com/watch?v=n0njBFLQSk8>), the first piece of music entirely composed by AI. Since then, the field of generative AI for music has been the subject of ongoing research for several decades. Among recent years' developments, new architectures and frameworks have become widespread among the general public, such as the WaveNet architecture introduced by Google in 2016, which has been able to generate high-quality audio samples, or the Magenta project, also developed by Google, which uses **Recurrent Neural Networks (RNNs)** and other ML techniques to generate music and other forms of art. Then, in 2020, OpenAI also announced Jukebox, a neural network that generates music, with the possibility to customize the output in terms of musical and vocal style, genre, reference artist, and so on.

Those and other frameworks became the foundations of many AI composer assistants for music generation. An example is Flow Machines, developed by Sony CSL Research. This generative AI system was trained on a large database of musical pieces to create new music in a variety of styles. It was used by French composer Benoît Carré to compose an album called *Hello World* (<https://www.helloworldalbum.net/>), which features collaborations with several human musicians.

Here, you can see an example of a track generated entirely by Music Transformer, one of the models within the Magenta project:

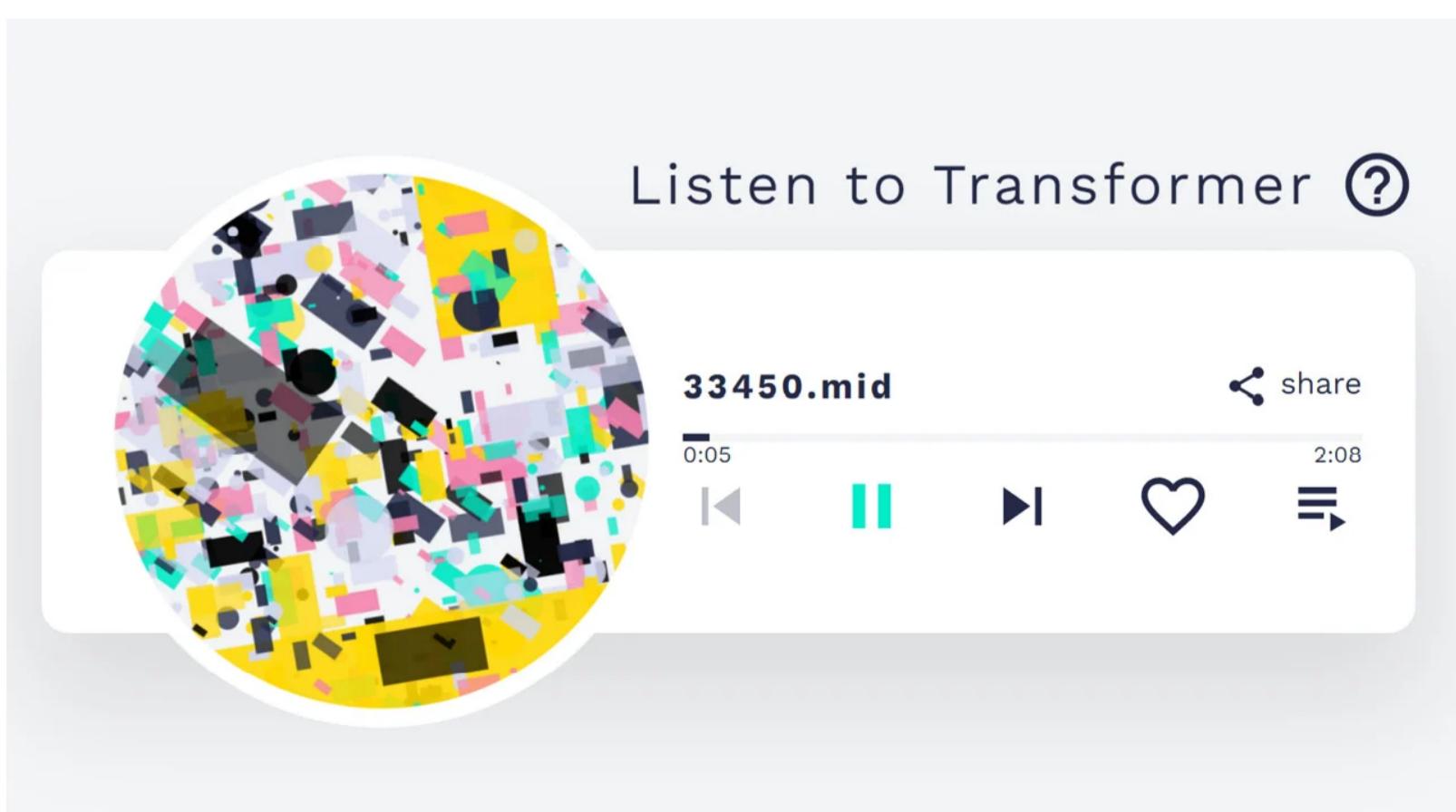


Figure 1.6 – Music Transformer allows users to listen to musical performances generated by AI

Another incredible application of generative AI within the music domain is speech synthesis. It is indeed possible to find many AI tools that can create audio based on text inputs in the voices of well-known singers.

For example, if you have always wondered how your songs would sound if Kanye West performed them, well, you can now fulfill your dreams with tools such as FakeYou.com (<https://fakeyou.com/>), Deep Fake Text to Speech, or UberDuck.ai (<https://uberduck.ai/>).

Selected voice

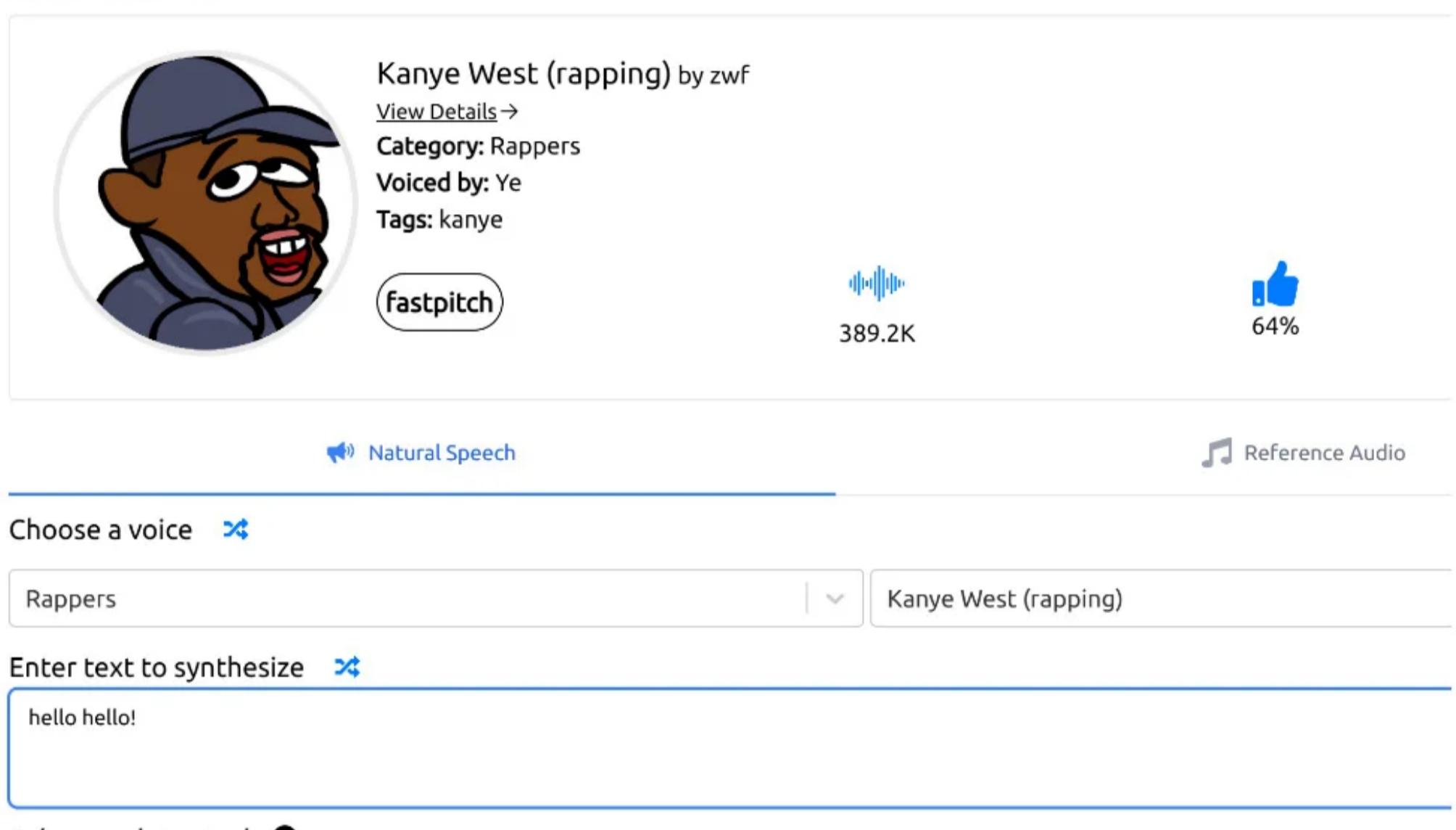


Figure 1.7 – Text-to-speech synthesis with UberDuck.ai

I have to say, the result is really impressive. If you want to have fun, you can also try voices from your all your favorite cartoons as well, such as Winnie The Pooh...

Next, we move to see generative AI for videos.

Video generation

Generative AI for video generation shares a similar timeline of development with image generation. In fact, one of the key developments in the field of video generation has been the development of GANs. Thanks to their accuracy in producing realistic images, researchers have started to apply these techniques to video generation as well. One of the most notable examples of GAN-based video

generation is DeepMind's **Motion to Video**, which generated high-quality videos from a single image and a sequence of motions. Another great example is NVIDIA's **Video-to-Video Synthesis (Vid2Vid)** DL-based framework, which uses GANs to synthesize high-quality videos from input videos.

The Vid2Vid system can generate temporally consistent videos, meaning that they maintain smooth and realistic motion over time. The technology can be used to perform a variety of video synthesis tasks, such as the following:

Converting videos from one domain into another (for example, converting a daytime video into a nighttime video or a sketch into a realistic image)

Modifying existing videos (for example, changing the style or appearance of objects in a video)

Creating new videos from static images (for example, animating a sequence of still images)

In September 2022, Meta's researchers announced the general availability of **Make-A-Video** (<https://makeavideo.studio/>), a new AI system that allows users to convert their natural language prompts into video clips. Behind such technology, you can recognize many of the models we mentioned for other domains so far – language understanding for the prompt, image and motion generation with image generation, and background music made by AI composers.

Overall, generative AI has impacted many domains for years, and some AI tools already consistently support artists, organizations, and general users. The future seems very promising; however, before jumping to the ultimate models available on the market today, we first need to have a deeper understanding of the roots of generative AI, its research history, and the recent developments that eventually lead to the current OpenAI models.

The history and current status of research

In previous sections, we had an overview of the most recent and cutting-edge technologies in the field of generative AI, all developed in recent years. However, the research in this field can be traced back decades ago.

We can mark the beginning of research in the field of generative AI in the 1960s, when Joseph Weizenbaum developed the chatbot ELIZA, one of the first examples of an NLP system. It was a simple rules-based interaction system aimed at entertaining users with responses based on text input, and it paved the way for further developments in both NLP and generative AI. However, we know that modern generative AI is a subfield of DL and, although the first **Artificial Neural Networks (ANNs)** were first introduced in the 1940s, researchers faced several challenges, including limited computing power and a lack of understanding of the biological basis of the brain. As a result, ANNs hadn't gained much attention until the 1980s when, in addition to new hardware and neuroscience developments, the advent of the **backpropagation** algorithm facilitated the training phase of ANNs. Indeed, before the advent of backpropagation, training Neural Networks was difficult because it was not possible to efficiently calculate the gradient of the error with respect to the parameters or weights associated with each neuron, while backpropagation made it possible to automate the training process and enabled the application of ANNs.

Then, by the 2000s and 2010s, the advancement in computational capabilities, together with the huge amount of available data for training, yielded the possibility of making DL more practical and available to the general public, with a consequent boost in research.

In 2013, Kingma and Welling introduced a new model architecture in their paper *Auto-Encoding Variational Bayes*, called **Variational Autoencoders (VAEs)**. VAEs are generative models that are based on the concept of variational inference. They provide a way of learning with a compact representation of data by encoding it into a lower-dimensional space called **latent space** (with the *encoder* component) and then decoding it back into the original data space (with the *decoder* component).

The key innovation of VAEs is the introduction of a probabilistic interpretation of the latent space. Instead of learning a deterministic mapping of the input to the latent space, the encoder maps the input to a probability distribution over the latent space. This allows VAEs to generate new samples by sampling from the latent space and decoding the samples into the input space.

For example, let's say we want to train a VAE that can create new pictures of cats and dogs that look like they could be real.

To do this, the VAE first takes in a picture of a cat or a dog and compresses it down into a smaller set of numbers into the latent space, which represent the most important features of the picture. These numbers are called **latent variables**.

Then, the VAE takes these latent variables and uses them to create a new picture that looks like it could be a real cat or dog picture. This new picture may have some differences from the original pictures, but it should still look like it belongs in the same group of pictures.

The VAE gets better at creating realistic pictures over time by comparing its generated pictures to the real pictures and adjusting its latent variables to make the generated pictures look more like the real ones.

VAEs paved the way toward fast development within the field of generative AI. In fact, only 1 year later, GANs were introduced by Ian Goodfellow. Differently from VAEs architecture, whose main elements are the encoder and the decoder, GANs consist of two Neural Networks – a generator and a discriminator – which work against each other in a zero-sum game.

The generator creates fake data (in the case of images, it creates a new image) that is meant to look like real data (for example, an image of a cat). The discriminator takes in both real and fake data, and tries to distinguish between them – it's the *critic* in our art forger example.

During training, the generator tries to create data that can fool the discriminator into thinking it's real, while the discriminator tries to become better at distinguishing between real and fake data. The two parts are trained together in a process called **adversarial training**.

Over time, the generator gets better at creating fake data that looks like real data, while the discriminator gets better at distinguishing between real and fake data. Eventually, the generator becomes so good at creating fake data that even the discriminator can't tell the difference between real and fake data.

Here is an example of human faces entirely generated by a GAN:



Figure 1.8 – Examples of photorealistic GAN-generated faces (taken from Progressive Growing of GANs for Improved Quality, Stability, and Variation, 2017: <https://arxiv.org/pdf/1710.10196.pdf>)

Both models – VAEs and GANs – are meant to generate brand new data that is indistinguishable from original samples, and their architecture has improved since their conception, side by side with the development of new models such as PixelCNNs, proposed by Van den Oord and his team, and WaveNet, developed by Google DeepMind, leading to advances in audio and speech generation.

Another great milestone was achieved in 2017 when a new architecture, called *Transformer*, was introduced by Google researchers in the paper, – *Attention Is All You Need*, was introduced in a paper by Google researchers. It was revolutionary in the field of language generation since it allowed for parallel processing while retaining memory about the context of language, outperforming the previous attempts of language models founded on RNNs or **Long Short-Term Memory (LSTM)** frameworks.

Transformers were indeed the foundations for massive language models called **Bidirectional Encoder Representations from Transformers (BERT)**, introduced by Google in 2018, and they soon become the baseline in NLP experiments.

Transformers are also the foundations of all the **Generative Pre-Trained (GPT)** models introduced by OpenAI, including GPT-3, the model behind ChatGPT.

Although there was a significant amount of research and achievements in those years, it was not until the second half of 2022 that the general attention of the public shifted toward the field of generative AI.

Not by chance, 2022 has been dubbed the *year of generative AI*. This was the year when powerful AI models and tools became widespread among the general public: diffusion-based image services (MidJourney, DALL-E 2, and Stable Diffusion), OpenAI's ChatGPT, text-to-video (Make-a-Video and Imagen Video), and text-to-3D (DreamFusion, Magic3D, and Get3D) tools were all made available to individual users, sometimes also for free.

This had a disruptive impact for two main reasons:

Once generative AI models have been widespread to the public, every individual user or organization had the possibility to experiment with and appreciate its potential, even without being a data scientist or ML engineer.

The output of those new models and their embedded creativity were objectively stunning and often concerning. An urgent call for adaptation—both for individuals and governments—rose.

Henceforth, in the very near future, we will probably witness a spike in the adoption of AI systems for both individual usage and enterprise-level projects.

Summary

In this chapter, we explored the exciting world of generative AI and its various domains of application, including image generation, text generation, music generation, and video generation. We learned how generative AI models such as ChatGPT and DALL-E, trained by OpenAI, use DL techniques to learn patterns in large datasets and generate new content that is both novel and coherent. We also discussed the history of generative AI, its origins, and the current status of research on it.

The goal of this chapter was to provide a solid foundation in the basics of generative AI and to inspire you to explore this fascinating field further.

In the next chapter, we will focus on one of the most promising technologies available on the market today, ChatGPT: we will go through the research behind it and its development by OpenAI, the architecture of its model, and the main use cases it can address as of today.

References

<https://arxiv.org/abs/1406.2661>

https://www.youtube.com/watch?v=ly9vRvyRf_E

<https://arxiv.org/abs/1912.04958>

This person does not exist: this-person-does-not-exist.com

<https://arxiv.org/abs/1808.06601>

<https://www.microsoft.com/en-us/research/blog/a-deep-generative-model-trifecta-three-advances-that-work-towards-harnessing-large-scale-power/>

<https://tcwang0509.github.io/vid2vid/>

[Previous Chapter](#)

[Next Chapter](#)