Ten Years of AI in Review

From image classification to chatbot therapy

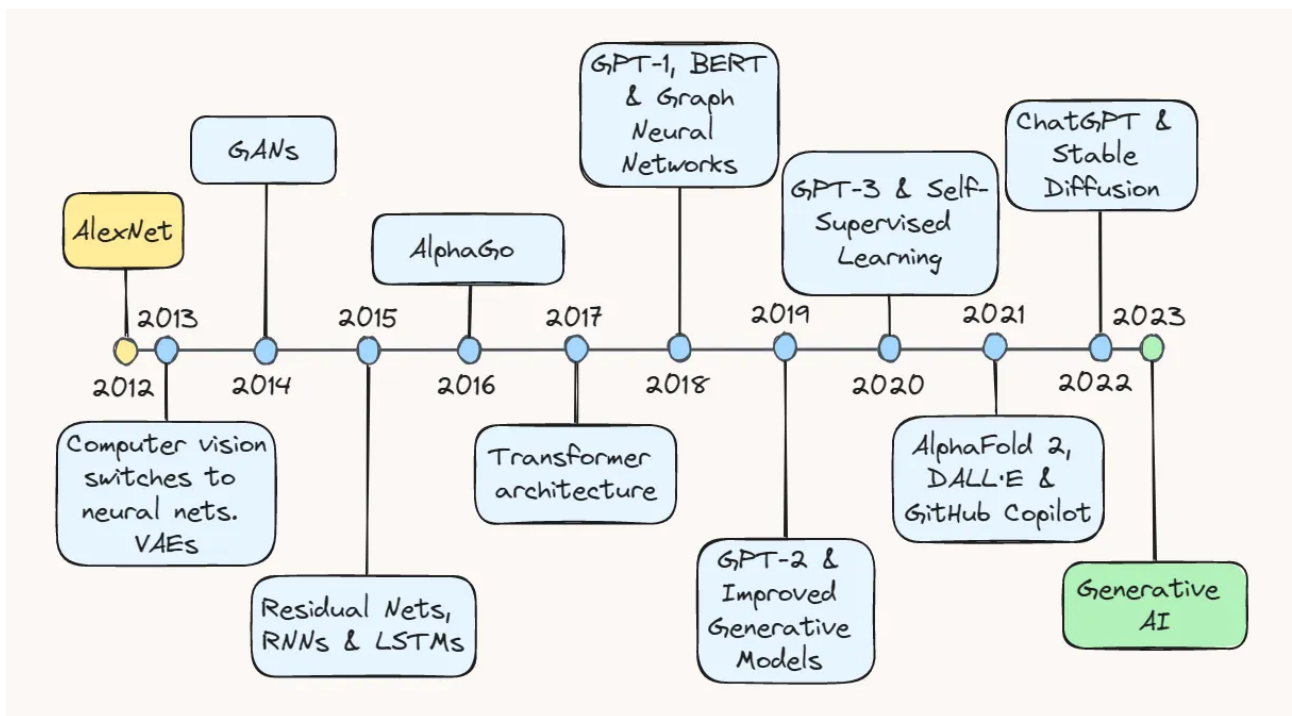Thomas A Dorfer Published in Towards Data Science
·15 min read ·May 23

Image by the Author.

The last decade has been a thrilling and eventful ride for the field of artificial intelligence (AI). Modest explorations of the potential of deep learning turned into an explosive proliferation of a field that now includes everything from recommender systems in e-commerce to object detection for autonomous vehicles and generative models that can create everything from realistic images to coherent text.

In this article, we'll take a walk down memory lane and revisit some of the key breakthroughs that got us to where we are today. Whether you are a seasoned AI practitioner or simply interested in the latest developments in the field, this article will

provide you with a comprehensive overview of the remarkable progress that led AI to become a household name.

2013: AlexNet and Variational Autoencoders

The year 2013 is widely regarded as the "coming-of-age" of deep learning, initiated by major advances in computer vision. According to a recent underline interview of Geoffrey Hinton, by 2013 *"pretty much all the computer vision research had switched to neural nets"*. This boom was primarily fueled by a rather surprising breakthrough in image recognition one year earlier.

In September 2012, AlexNet, a deep convolutional neural network (CNN), pulled off a record-breaking performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), demonstrating the potential of deep learning for image recognition tasks. It achieved a top-5 error of 15.3%, which was 10.9% lower than that of its nearest competitor.
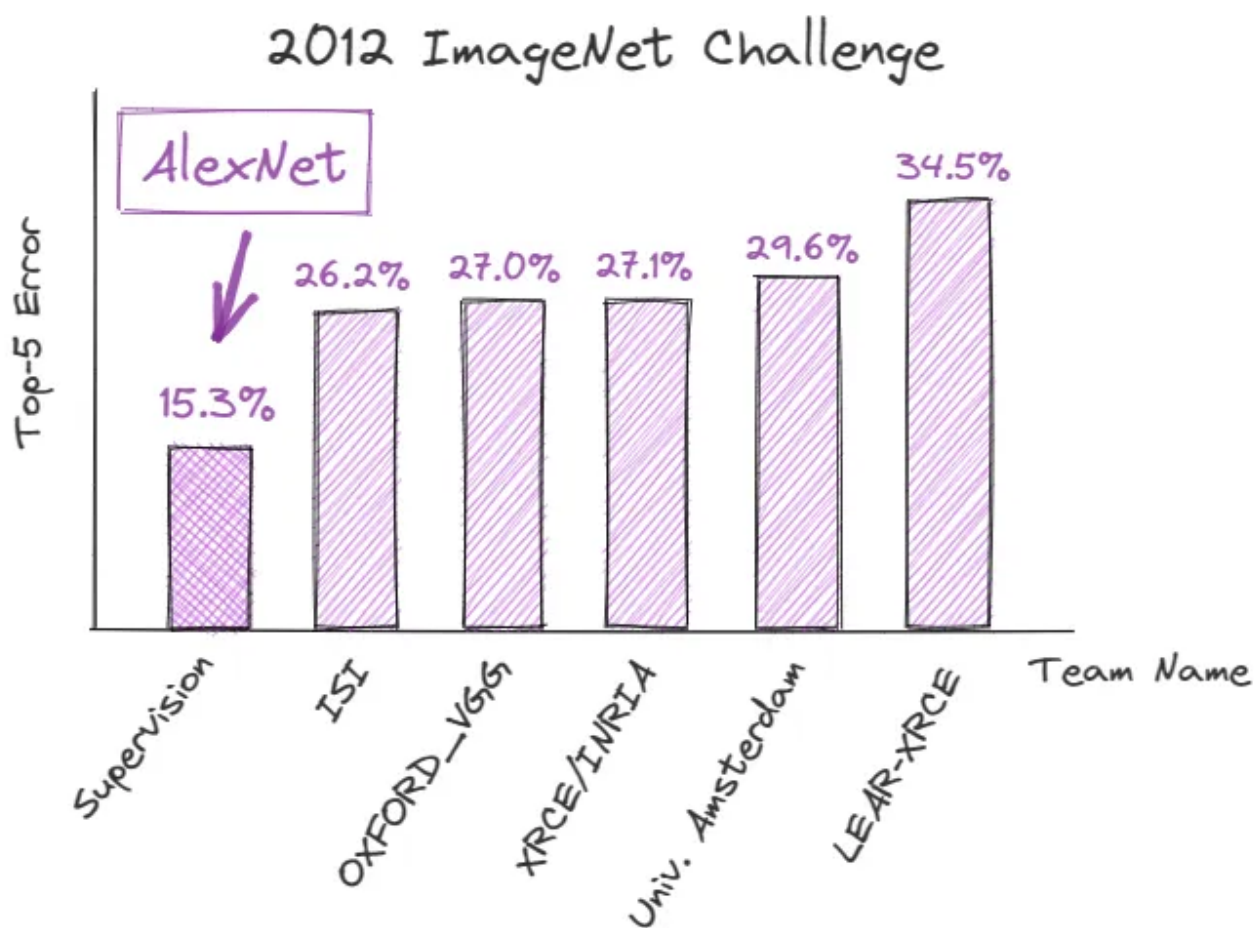

Image by the Author.

The technical improvements behind this success were instrumental for the future trajectory of AI and dramatically changed the way *deep* learning was perceived.

First, the authors applied a deep CNN consisting of five convolutional layers and three fully-connected linear layers — an architectural design dismissed by many as impractical at the time. Moreover, due to the large number of parameters produced by the network's

depth, training was done in parallel on two graphics processing units (GPUs), demonstrating the ability to significantly accelerate training on large datasets. Training time was further reduced by swapping traditional activation functions, such as sigmoid and tanh, for the more efficient rectified linear unit (ReLU).
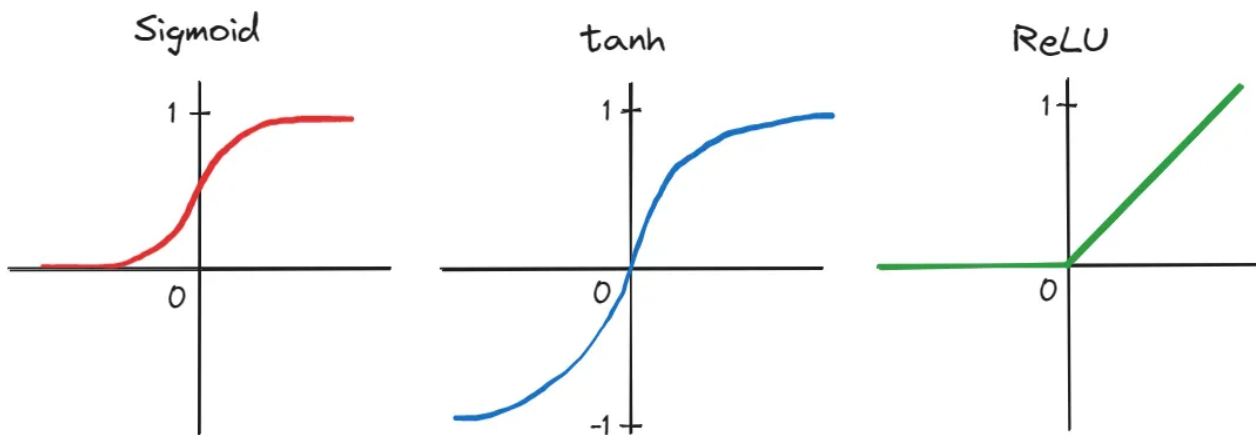

Image by the Author.

These advances that collectively led to the success of AlexNet marked a turning point in the history of AI and sparked a surge of interest in deep learning among both academics and the tech community. As a result, 2013 is considered by many as the inflection point after which deep learning truly began to take off.

Also happening in 2013, albeit a little drowned out by the noise of AlexNet, was the development of variational autoencoders, or VAEs — generative models that can learn to represent and generate data such as images and sounds. They work by learning a compressed representation of the input data in a lower-dimensional space, known as *latent space*. This allows them to generate new data by sampling from this learned latent space. VAEs, later on, turned out to open up new avenues for generative modeling and data generation, with applications in fields like art, design, and gaming.

2014: Generative Adversarial Networks

The following year, in June 2014, the field of deep learning witnessed another serious advance with the introduction of generative adversarial networks, or GANs, by Ian Goodfellow and colleagues.

GANs are a type of neural network capable of generating new data samples that are similar to a training set. Essentially, two networks are trained simultaneously: (1) a generator network generates fake, or synthetic, samples, and (2) a discriminator network evaluates their authenticity. This training is performed in a game-like setup, with the generator trying to create samples that fool the discriminator, and the discriminator trying to correctly call out the fake samples.

At that time, GANs represented a powerful and novel tool for data generation, being used not only for generating images and videos, but also music and art. They also contributed to the advance of unsupervised learning, a domain largely regarded as underdeveloped and

challenging, by demonstrating the possibility to generate high-quality data samples without relying on explicit labels.

2015: ResNets and NLP Breakthroughs

In 2015, the field of AI made considerable advances in both computer vision and natural language processing, or NLP.

Kaiming He and colleagues published a paper titled "Deep Residual Learning for Image Recognition", in which they introduced the concept of residual neural networks, or ResNets — architectures that allow information to flow more easily through the network by adding shortcuts. Unlike in a regular neural network, where each layer takes the output of the previous layer as input, in a ResNet, additional *residual* connections are added that skip one or more layers and directly connect to deeper layers in the network.

As a result, ResNets were able to solve the problem of vanishing gradients, which enabled the training of much deeper neural networks beyond what was thought to be possible at the time. This, in turn, led to significant improvements in image classification and object recognition tasks.

At around the same time, researchers made considerable progress with the development of recurrent neural networks (RNNs) and long short-term memory (LSTM) models. Despite having been around since the 1990s, these models only started to generate some buzz around 2015, mainly due to factors such as (1) the availability of larger and more diverse datasets for training, (2) improvements in computational power and hardware, which enabled the training of deeper and more complex models, and (3) modifications made along the way, such as more sophisticated gating mechanisms.

As a result, these architectures made it possible for language models to better understand the context and meaning of text, leading to vast improvements in tasks such as language translation, text generation, and sentiment analysis. The success of RNNs and LSTMs around that time paved the way for the development of large language models (LLMs) we see today.

2016: AlphaGo

After Garry Kasparov's defeat by IBM's Deep Blue in 1997, another *human vs. machine* battle sent shockwaves through the gaming world in 2016: Google's AlphaGo defeated the world champion of Go, Lee Sedol.
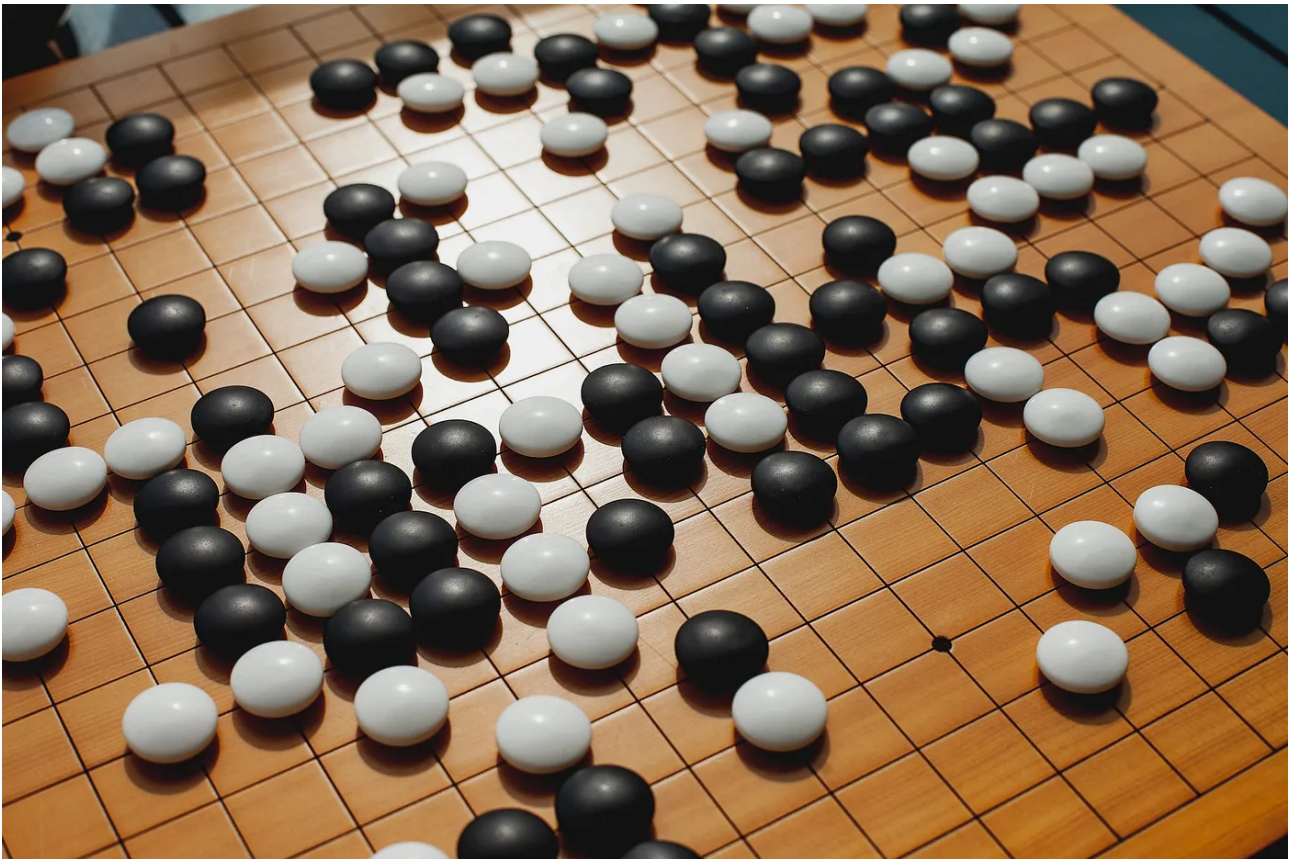
Photo by [Elena Popova](#) on [Unsplash](#).

Sedol's defeat marked another major milestone in the trajectory of AI advancement: it demonstrated that machines could outsmart even the most skilled human players in a game that was once considered too complex for computers to handle. Using a combination of deep reinforcement learning and Monte Carlo tree search, AlphaGo analyzes millions of positions from previous games and evaluates the best possible moves — a strategy that far surpasses human decision making in this context.

2017: Transformer Architecture and Language Models

Arguably, 2017 was the most pivotal year that laid the foundation for the breakthroughs in generative AI that we are witnessing today.

In December 2017, Vaswani and colleagues released the foundational paper "Attention is all you need", which introduced the transformer architecture that leverages the concept of self-attention to process sequential input data. This allowed for more efficient processing of long-range dependencies, which had previously been a challenge for traditional RNN architectures.

Transformers are comprised of two essential components: encoders and decoders. The encoder is responsible for encoding the input data, which, for example, can be a sequence of words. It then takes the input sequence and applies multiple layers of self-attention and feed-forward neural nets to capture the relationships and features within the sentence and learn meaningful representations.

Essentially, self-attention allows the model to understand relationships between different words in a sentence. Unlike traditional models, which would process words in a fixed order, transformers actually examine all the words at once. They assign something called *attention* scores to each word based on its relevance to other words in the sentence.

The decoder, on the other hand, takes the encoded representation from the encoder and produces an output sequence. In tasks such as machine translation or text generation, the decoder generates the translated sequence based on the input received from the encoder. Similar to the encoder, the decoder also consists of multiple layers of self-attention and feed-forward neural nets. However, it includes an additional attention mechanism that enables it to focus on the encoder's output. This then allows the decoder to take into account relevant information from the input sequence while generating the output.

The transformer architecture has since become a key component in the development of LLMs and has led to significant improvements across the domain of NLP, such as machine translation, language modeling, and question answering.

2018: GPT-1, BERT and Graph Neural Networks

A few months after Vaswani et al. published their foundational paper, the **G**enerative **P**retrained **T**ransformer, or GPT-1, was introduced by OpenAI in June 2018, which utilized the transformer architecture to effectively capture long-range dependencies in text. GPT-1

was one of the first models to demonstrate the effectiveness of unsupervised pre-training followed by fine-tuning on specific NLP tasks.

Also taking advantage of the still quite novel transformer architecture was Google, who, in late 2018, released and open-sourced their own pre-training method called **B**idirectional **E**ncoder **R**epresentations from **T**ransformers, or <u>BERT</u>. Unlike previous models that process text in a unidirectional manner (including GPT-1), BERT considers the context of each word in both directions simultaneously. To illustrate this, the authors provide a very intuitive example:

… in the sentence *"I accessed the bank account"*, a unidirectional contextual model would represent *"bank"* based on *"I accessed the"* but not *"account"*. However, BERT represents *"bank"* using both its previous and next context — *"I accessed the … account"* — starting from the very bottom of a deep neural network, making it deeply bidirectional.

The concept of bidirectionality was so powerful that it led BERT to outperform state-of-the-art NLP systems on a variety of benchmark tasks.

Apart from GPT-1 and BERT, graph neural networks, or <u>GNNs</u>, also made some noise that year. They belong to a category of neural networks that are specifically designed to work with graph data. GNNs utilize a message passing algorithm to propagate information across the nodes and edges of a graph. This enables the network to learn the structure and relationships of the data in a much more intuitive way.

This work allowed for the extraction of much deeper insights from data and, consequently, broadened the range of problems that deep learning could be applied to. With GNNs, major advances were made possible in areas like social network analysis, recommendation systems, and drug discovery.

2019: GPT-2 and Improved Generative Models

The year 2019 marked several notable advancements in generative models, particularly the introduction of <u>GPT-2</u>. This model really left its peers in the dust by achieving state-of-the-art performance in many NLP tasks and, in addition, was capable to generate highly realistic text, which, in hindsight, gave us a teaser of what was about to come in this arena. Other improvements in this domain included DeepMind's <u>BigGAN</u>, which generated high-quality images that were almost indistinguishable from real images, and NVIDIA's <u>StyleGAN</u>, which allowed for better control over the appearance of those generated images.

Collectively, these advancements in what's now known as generative AI pushed the boundaries of this domain even further, and…

2020: GPT-3 and Self-Supervised Learning

… not soon thereafter, another model was born, which has become a household name even outside of the tech community: <u>GPT-3</u>. This model represented a major leap forward in the scale and capabilities of LLMs. To put things into context, GPT-1 sported a measly

117 million parameters. That number went up to 1.5 billion for GPT-2, and 175 billion for GPT-3.

This vast amount of parameter space enables GPT-3 to generate remarkably coherent text across a wide range of prompts and tasks. It also demonstrated impressive performance in a variety of NLP tasks, such as text completion, question answering, and even creative writing.

Moreover, GPT-3 highlighted again the potential of using self-supervised learning, which allows models to be trained on large amounts of unlabeled data. This has the advantage that these models can acquire a broad understanding of language without the need for extensive task-specific training, which makes it far more economical.

The NYT writes about my favorite topic: Self-Supervised Learning.

It is the latest craze in AI/Deep Learning.
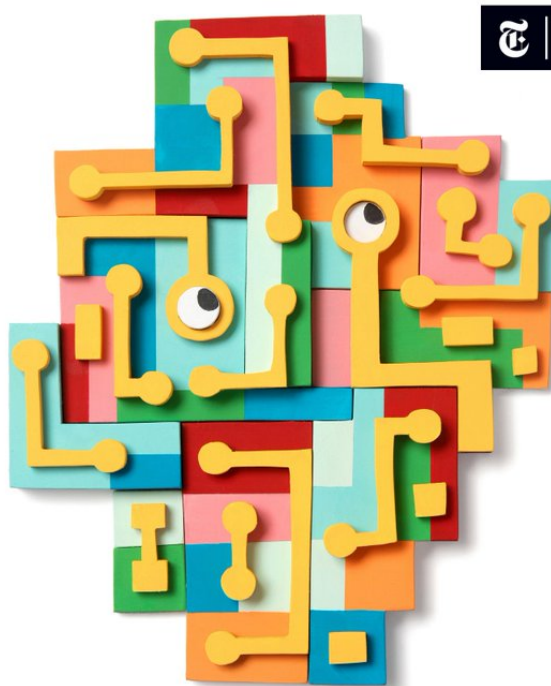
nytimes.com

Computers Already Learn From Us. But Can They Teach Themselves? (Published 2020)
Scientists are exploring approaches that would help machines develop their own sort of common sense.
7:24 AM · Apr 9, 2020

2021: AlphaFold 2, DALL·E and GitHub Copilot

From protein folding to image generation and automated coding assistance, the year of 2021 was an eventful one thanks to the releases of AlphaFold 2, DALL·E, and GitHub Copilot.

AlphaFold 2 was hailed as a long-awaited solution to the decades-old protein folding problem. DeepMind's researchers extended the transformer architecture to create *evoformer blocks* — architectures that leverage evolutionary strategies for model optimization — to build a model capable of predicting a protein's 3D structure based on its 1D amino acid sequence. This breakthrough has enormous potential to revolutionize areas like drug discovery, bioengineering, as well as our understanding of biological systems.

OpenAI also made it in the news again this year with their release of DALL·E. Essentially, this model combines the concepts of GPT-style language models and image generation to enable the creation of high-quality images from textual descriptions.

To illustrate how powerful this model is, consider the image below, which was generated with the prompt "Oil painting of a futuristic world with flying cars".

Image produced by DALL·E.

Lastly, GitHub released what would later become every developers best friend: Copilot. This was achieved in collaboration with OpenAI, which provided the underlying language model, Codex, that was trained on a large corpus of publicly available code and, in turn, learned to understand and generate code in various programming languages. Developers can use Copilot by simply providing a code comment stating the problem they are trying to solve, and the model would then suggest code to implement the solution. Other features include the ability to describe input code in natural language and translate code between programming languages.

2022: ChatGPT and Stable Diffusion

The rapid development of AI over the past decade has culminated in a groundbreaking advancement: OpenAI's ChatGPT, a chatbot that was released into the wild in November 2022. The tool represents a cutting-edge achievement in NLP, capable of generating coherent and contextually relevant responses to a wide range of queries and prompts. Furthermore, it can engage in conversations, provide explanations, offer creative suggestions, assist with problem-solving, write and explain code, and even simulate different personalities or writing styles.

Can you explain what the following function does?

```python
def updateDates(directory):
    for filename in os.listdir(directory):
        f = os.path.join(directory, filename)
        if os.path.splitext(f)[-1] == '.wav':
            pattern = "\d{4}-\d{2}-\d{2}"
            today = date.today().strftime("%Y-%m-%d")
            f_new = re.sub(pattern, today, f)
            os.rename(f, f_new)
```

The function `updateDates` takes in a single argument, `directory`, which is a file path to a directory on the computer. The function then iterates over all of the files in that directory using `os.listdir(directory)`. For each file, it checks if the file extension is '.wav' using `os.path.splitext(f)[-1]`. If the file extension is '.wav', it then uses the python `re` module to search for a pattern of 4 digits, a dash, 2 digits, another dash, and 2 more digits (\d{4}-\d{2}-\d{2}), which represent a date in the format of YYYY-MM-DD. It then gets the current date using `date.today().strftime("%Y-%m-%d")` and replaces the found date pattern with the current date. It then renames the file using the os.rename(f, f_new) method, with the new name being the original file name with the replaced date.

Image by the Author.

The simple and intuitive interface through which one can interact with the bot also stimulated a sharp rise in usability. Previously, it was mostly the tech community that would play around with the latest AI-based inventions. However, these days, AI tools have penetrated almost every professional domain, from software engineers to writers, musicians, and advertisers. Many companies are also using the model to automate services such as customer support, language translation, or answering FAQs. In fact, the wave of automation we're seeing has rekindled some worries and stimulated discussions on which jobs might be at risk of being automated.

Even though ChatGPT was taking up much of the limelight in 2022, there was also a significant advancement made in image generation. Stable diffusion, a latent text-to-image diffusion model capable of generating photo-realistic images from text descriptions, was released by Stability AI.

Stable diffusion is an extension of the traditional diffusion models, which work by iteratively adding noise to images and then reversing the process to recover the data. It was designed to speed up this process by operating not directly on the input images, but instead on a lower-dimensional representation, or latent space, of them. In addition, the diffusion process is modified by adding the transformer-embedded text prompt from the user to the network, allowing it to guide the image generation process throughout each iteration.

Overall, the release of both ChatGPT and Stable Diffusion in 2022 highlighted the potential of multimodal, generative AI and sparked a massive boost of further development and investment in this area.

2023: LLMs and Bots

The current year has undoubtedly emerged as the year of LLMs and chatbots. More and more models are being developed and released at a rapidly increasing pace.
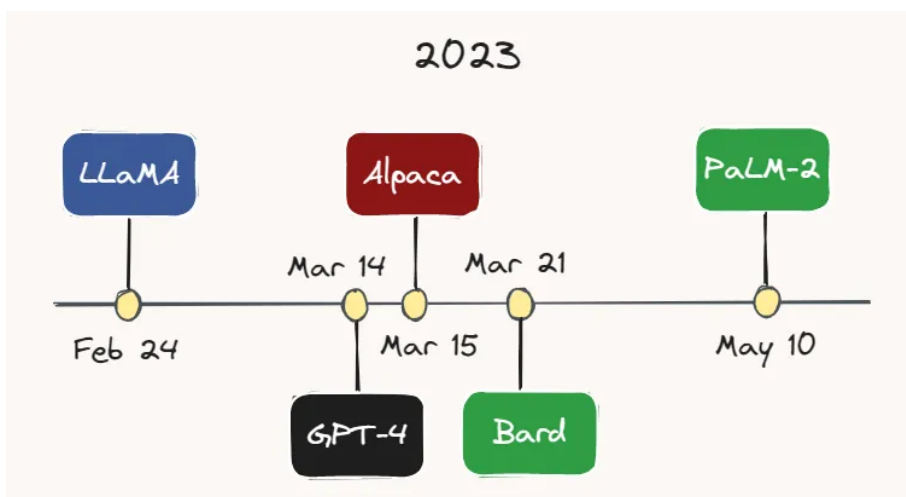

Image by the Author.

For instance, on February 24th, Meta AI released LLaMA — an LLM that outperforms GPT-3 on most benchmarks, despite having a considerably smaller number of parameters. Less than a month later, on March 14th, OpenAI released GPT-4 — a bigger, more capable, and multimodal version of GPT-3. While the exact number of parameters of GPT-4 is unknown, it is speculated to be in the trillions.

On March 15, researchers at Stanford University released Alpaca, a light-weight language model that was fine-tuned from LLaMA on instruction-following demonstrations. A couple days later, on March 21st, Google launched its ChatGPT rival: Bard. Google also just released its latest LLM, PaLM-2, earlier this month on May 10th. With the relentless pace of development in this area, it is highly likely that yet another model will have emerged by the time you're reading this.

We are also seeing more and more companies incorporating these models into their products. For example, Duolingo announced its GPT-4-powered Duolingo Max, a new subscription tier with the aim of providing tailored language lessons to each individual. Slack has also rolled out an AI-powered assistant called Slack GPT, which can do things like draft replies or summarize threads. Furthermore, Shopify introduced a ChatGPT-powered assistant to the company's Shop app, which can help customers identify desired products using a variety of prompts.

Say 'hello' to your new shopping BFF: the @Shop app's ChatGPT powered assistant is ready to talk products, trends, and maybe even the meaning of life

Shopify announcing its ChatGPT-powered assistant on Twitter.

Interestingly, AI chatbots are nowadays even considered as an alternative to human therapists. For example, Replika, a US chatbot app, is offering users an "*AI companion who cares, always here to listen and talk, always on your side*". Its founder, Eugenia Kuyda, says that the app has a wide variety of customers, ranging from autistic children, who turn to it as a way to "*warm up before human interactions*", to lonely adults who simply are in the need of a friend.

Before we conclude, I'd like to highlight what may well be the climax of the last decade of AI development: people are actually using Bing! Earlier this year, Microsoft introduced its GPT-4-powered "*copilot for the web*" that has been customized for search and, for the first time in… forever (?), has emerged as a serious contender to Google's long-standing dominance in the search business.

Looking back and looking forward

As we reflect on the last ten years of AI development, it becomes evident that we have been witnessing a transformation that has had profound impact on how we work, do business, and interact with one another. Most of the considerable progress that has lately been achieved with generative models, particularly LLMs, seems to be adhering to the common belief that "bigger is better", referring to the parameter space of the models. This has been especially noticeable with the GPT series, which started out with 117 million parameters (GPT-1) and, after each successive model increasing by approximately an order of magnitude, culminated in GPT-4 with potentially trillions of parameters.

However, based on a recent interview, OpenAI CEO Sam Altman believes that we have reached the end of the "bigger is better" era. Going forward, he still thinks that the parameter count will trend up, but the main focus of future model improvements will be on increasing the model's capability, utility, and safety.

The latter is of particular importance. Considering that these powerful AI tools are now in the hands of the general public and no longer confined to the controlled environment of research labs, it is now more critical than ever that we tread with caution and ensure that these tools are safe and align with humanity's best interests. Hopefully we'll see as much development and investment in AI safety as we've seen in other areas.

PS: In case I've missed a core AI concept or breakthrough that you think should have been included in this article, please let me know in the comments below!