451 Research

S&P Global
Market Intelligence

A primer on generative AI: What is it, who are the players — and did it write this headline?

Analysts - Nick Patience

Publication date: Monday, December 19 2022

Introduction

Generative AI is having a moment. The recent uptick of interest — at least among those in the technology industry — in OpenAI's ChatGPT conversational AI tool has shone a light on an area that has been growing, research project by research project, in the background for a few years now. ChatGPT is fun to play with, but its flaws were exposed quite quickly. Some were serious in terms of it writing untruths; some less so, such as its tendency for seemingly every poem it produced to rely heavily on iambic pentameter. But supporters would point out that at least it can write poetry, produce imaginary dialogues between people who were not alive at the same time, and even write occasional somewhat creditable analyses of the technology industry (full disclosure: None of this report was produced using ChatGPT).

Generative AI refers to AI models that are taught on large training sets that aim to predict the probability of a sequence of words and generate text. The first examples were large language models (LLMs), which emerged with Google's Bert in 2018, based on transformer architectures. They can do things such as summarizations, language translation and even write poetry (of a sort). They are deemed large mainly because of the number of parameters they have, but also the amount of compute resources they cost to run and the volumes of data needed to train them. Now there are models that can generate text, some that can generate images from text descriptions, some that can generate software code, and others that can generate sequences of proteins and even 3D models of those proteins' structure.

The Take

It is too soon to tell if the hype around ChatGPT is language models' "ImageNet moment." ImageNet was the set of annotated images created to test classification accuracy (e.g., a cat/not a cat) that

formed the basis of a competition from 2009-2017; some of the winners of which went on to create well-known AI companies. Early results with ImageNet were not promising, with low accuracy rates; they increased suddenly when competitors started using neural networks. Similarly, earlier LLMs have promised much, but the fact that ChatGPT is so easily accessible has created a buzz around it, greater even than that of DALL.E a few months ago with its image generation.

One of the key takeaways from ImageNet was that large training data sets are crucial to progress in machine learning, so it is worth remembering that these models are not producing anything from scratch — they are generating predictions based on their training sets. Upgrades are promised for all the major models in coming months too, so those that are basing their businesses on them are sure to show improvement in underlying performance; their challenge will be turning those into applications that solve genuine problems. AI has driven much over-promising many times in the past — take driverless cars for a recent example — and checks and balances will be needed to avoid releasing content that has no basis in fact.

The underlying models

Companies in this space can be categorized in various ways. We have chosen to highlight the underlying models first, then the companies generating text from text, those generating images from text and those generating code from text. Some of those providing models are also providing applications on top. There are many other ways they could be categorized, and this report is a snapshot of what is going on. It is not meant to be exhaustive, but we will update it through 2023.

The underlying models come from a mix of commercial firms focused on models, those — like Meta Platforms Inc. — that build them as part of their business, and finally, open-source projects. This is a partial list.

Table 1: Underlying Models

AI21 Labs	Israel-based company building its own LLMs and writing assistant apps.
Aleph Alpha	Germany-based research company pushing sparse models needing fewer parameters.
Anthropic	A public-benefit and AI "safety and research" company that has raised $704M, including a $580M round in April, led by Sam Bankman-Fried of FTX notoriety.
Cohere AI	Canadian company with its own LLM, which uses Google Cloud's Tensor Processing Units (TPUs) for training purposes.
DeepMind Gopher	A 280-billion-parameter language model from Alphabet Inc.'s Google.
Google	Its LaMDA and PaLM LLMs are still at the research stage. LaMDA, a dialog model, has been more visible and is the one that a Google engineer claimed was sentient in June 2022, which led to his dismissal. PaLM is multi-modal, able to process text, images and speech.
Hugging Face Bloom	An open-source 174-billion-parameter language model, built by the BigScience "open science" project and trained on infrastructure paid for by the French government.
Meta OPT-175B	A 175-billion-parameter language model, open-sourced by the owner of Facebook.
OpenAI GPT-3, Codex, DALL.E	Text, code and image models from the Microsoft-backed research organization originally founded by Elon Musk and Sam Altman.

Source: 451 Research

Text generators — these produce text when prompted by instructions or questions. Early use cases are centered on general-purpose writing assistants and automated marketing copy writers. We

presume that is because those are the least controversial use cases; few will be annoyed if marketing collateral is written by a computer rather than a person — except if that were their job, of course. But contrast that with models designed to write academic essays or news articles or legal contracts, where the potential for abuse is far higher.

Table 2: Text Generators

Compose AI	Chrome extension writing tool using its own LLM.
Copysmith	Content-generation tool that leverages GPT-3. Acquired similar company called Rytr in October 2022.
Grammarly	Writing assistant that "meets humans where they are" and corrects their writing, with plans for a more personalized offering.
Jasper.AI	A writing assistant focused on business use cases in sales, marketing and social media use cases that works in 26 languages and leverages OpenAI's GPT-3.
Kaizan	Extracts tasks and actions from client communications to enable customer access teams to work more efficiently.
Lavender	Helps salespeople write better emails so they get more replies and offers tools to onboard and train salespeople.
NovelAI	Storytelling and image creation tool from Anlatan that leverages Stable Diffusion.
Textio	Writing assistant aimed at the recruiting and performance management sectors, focused on producing bias-free text.
TypeWise	Text generator aimed at customer service and sales teams.
Wordtune	A Chrome extension writing tool from AI21 Labs.

Source: 451 Research

Image generators — these produce images when prompted by a text description.

Table 3: Image Generators

DALL.E	OpenAI's tool for generating images from text, now in public beta.
Mage.Space	For developing art based on DALL.E and Stable Diffusion.
Midjourney	Research company that enables images to be created from Discord.
OpenArt	For developing art based on DALL.E and Stable Diffusion.
Rosebud.AI	Smartphone apps for developing stock images, enhancing existing ones and making portraits "come alive."
Stable Diffusion	A text-to-image model from Stability AI that can also do image modification. The code has been made source-available.

Source: 451 Research

Code generators — these produce software code from natural language questions. For example, "show me an example of a support vector machine (SVM) algorithm implemented in Python" with the output being the Python code.

Table 4: Code Generators

AI2SQL	Produces SQL code from natural language.
Debuild	Builds web applications based on text descriptions.
GitHub Copilot	From Microsoft and OpenAI, it turns natural language into code in various languages; leverages OpenAI's Codex.
MutableAI	Autocompletes and documents Visual Studio code in various languages and can turn Jupyter notebooks into production-ready code.

Source: 451 Research

Outlook

The explosion of interest in generative AI driven by DALL.E and now ChatGPT has opened many potential new use cases and exposed pitfalls to overcome, such as blurring the boundaries between truth and facts and even getting simple arithmetic wrong. However, users need to realize that these are not finished products, but experimental tools, although judging by the avalanche of tweets about how ChatGPT shows why we are only a short distance from achieving artificial general intelligence (AGI) on the one hand or that it is the most dangerous thing unleashed on the internet on the other perhaps demonstrates how risky such experiments can be. However, it seems clear already that the earliest use cases will be focused on creative tasks, including copy writing and art generation; areas where accuracy is not essential. Beyond these, generative AI — text generation in particular — may be integrated into all sorts of applications and it has major implications in areas such a chatbots and enterprise search, which have labored under the weight of unmet expectations for years.