Invesco

In Silico, Part III

The rise of Generative AI and how it could change our future

In Silico is a multi-part series discussing artificial intelligence, its economic and financial impact, and its role as a driver of change.

In this article we explore Generative AI technologies with a view to understanding their commercial uses and economic impact. For an introduction to the series please refer to 'In Silico, Part I: Understanding AI's "Hello, World!" Moment'. For an introduction to Machine Learning and Deep Learning, which underpin Generative AI, we strongly recommend reading 'In Silico, Part II: AI and the Quiet Revolution of Machine Learning'.

Ashley Oerth, CFA®
Senior Investment Strategy Analyst
Investment Thought Leadership

Cyril Birks
Global Thought Leadership Intern
Investment Thought Leadership

What is Generative AI?

In previous editions of this series, we've explored exactly what AI means—but generative AI is a relatively new idea to most readers. While enthusiasts might know "generative AI" from image generators (like DALL-E, Stable Diffusion, or Midjourney), ChatGPT was the first time most of the world encountered "generative AI." As a freely accessible chatbot with uncanny language capabilities, ChatGPT took the world by storm, reaching one million users in just five days.

Machine Learning is separated broadly into one of two broad categories: Discriminative AI and Generative AI. Discriminative AI is already in use across a wide variety of functions in today's economy, a phenomena we referred to as the 'Quiet Revolution' in our second installment of In Silico.[1] Discriminative AI is used to classify, categorize, and analyze data, helping users and analysts understand the relationships between variables, make predictions about possible outcomes given some set of variables, and understand whether some new data point is more like known entity A or known entity B. In Figure 1, we show how a discriminative model may be used to understand the separation of categories of data.

Generative AI, meanwhile, is used with the goal of creating (generating) new data that looks like its training data. A generative algorithm learns the relationships between variables or similarities between datasets not to analyze and categorize, but to infer about the general nature of the data and "generate" output data based on the model's understanding of the data. Models trained on human language can recreate new, plausible language data. Models trained on image data can recreate new, plausible image data. In Figure 2, we show how a generative model may be used to learn from an existing dataset and infer the next datapoint given an input.

Figure 1: Discriminative AI

Goal: Understand how different data types are seperated based on the qualities of each data type.
Example task: Differentiating between two 'types of users'

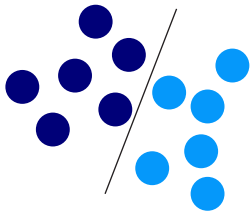
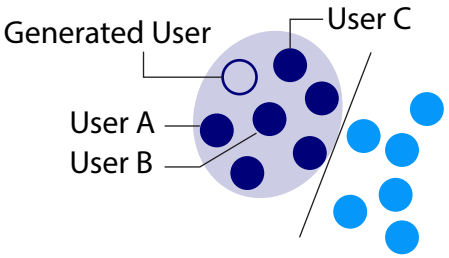


Figure 2: Genarative AI

Goal: Produce new datapoints related to a given input by using relationships learned in the training data.
Example task: Generating behavior like users A, B, and C



Source: Invesco. For illustrative purposes only.

1 See In Silico Part II: AI and the quiet revolution of machine learning

Overnight Success, Decade in the Making

Generative AI has been around for some time, but only recently have models become sophisticated enough to compete with human performance.

Most notable for today's excitement are Generative Pretrained Transformers (GPTs), which are the foundation of OpenAI's ChatGPT. The technology behind GPTs evolved over many decades, but perhaps one of the most important steps along the way was a 2017 paper from Google[2] that pioneered the concept of transformers. This technology allows machine learning models to better interpret text by relating each word in a given dataset back to previously seen words, enabling Deep Learning (covered in Part 2 of this series) to demonstrate better understanding of text data.

At their core, GPTs are models built on the same technology and trained with large amounts of unlabeled text data. The first model was launched in 2018 by OpenAI in the form of GPT-1, with successive improvements released through GPT-2 in 2019, GPT-3 in 2020, GPT-3.5 in 2022, and the latest GPT-4 in March 2023. Google built on a similar transformer-based technology through its Pathways Language Model (PaLM) announced in 2022 and its Language Model for Dialogue Applications (LaMDA) which was first launched in 2020 but entered the limelight in February 2023 as the foundation to Google's Bard chatbot, a ChatGPT competitor.

As may be evident already, there are a wide range of techniques for approaching generative AI models. Collectively, these are referred to today as large language models—language models that are trained on huge volumes of text data. Colloquially, we may see these simply called "GPTs," even if this misses some technical accuracy.

Major Generative AI Models

	Generative Pretrained Transformers (GPTs)	Generative Adversarial Networks (GANs)	Diffusion Models
Introduced	Circa 2018	Circa 2014	Circa 2015
Uses	Understanding the logical connections between words ("tokens").	Mimicking real data, such as for image or video generation	Making images clearer, filling in the blanks of missing data, upscaling content (from 1080p to 4K, for example) or generating images, inpainting, upscaling and super-resolution, generation.
Examples	Summarizing texts, creating text, copy-editing, reviewing code, creating code, performing logical and mathematical calculations, generating text to explain data.	Deepfakes	Text-to-image generation, photo editing, image enhancement, restoration.
Summary	GPTs, like ChatGPT, are a type of transformer-based model that use unsupervised learning to generate sentences that are both coherent and relevant to a given input. Think of GPTs as creative writers. They can produce coherent, contextually appropriate sentences by considering the relationships between all the words in each input.	GANs consist of two models: A generator and a discriminator. The generator creates increasingly convincing output, while the discriminator evaluates the output against real data. Over time, the generator gets better at generating data as it becomes better able to fool the discriminator. The result is the ability to generate outputs that look like the real thing.	In essence, diffusion models learn by adding random values and changes ("noise") to training data and then de-noising it to recover the original input. From this process they learn how to generate a desired output from randomness. Where other generative models build from the ground up, brick by brick, diffusion models are more like chiseling away at a block of material.

Generative AI, for Every Modality

Following successive advancements in computing power, data availability and capture, and more sophisticated models and techniques, generative AI has now reached a level of performance that rivals human capabilities. In a very short period, there has been an explosion in AI tools for businesses and ordinary consumers alike. While we are still in the very early days of AI for the masses, it appears that two loosely overlapping trends are emerging which can broadly be classified as B2C (AI as a Service, where users plug-and-play with AI through apps) and B2B (AI as a Tool, with Foundation Models leading the way).

While ChatGPT and its GPT-4 foundation model have stolen the limelight, there are already numerous generative AI systems for a variety of modalities. The most prevalent example is text, but there are also models for programming, images, audio, video, scientific research, and more.

2 Attention Is All You Need, Vaswani et al., June, 2017 (revised 6 December 2017)

Below we provide examples of popular modalities and where we see generative AI already making headway:

- **Text** is arguably the most exciting and prevalent modality approached currently. Examples of existing models used for text include OpenAI's GPT-4 (which powers ChatGPT), Google's LaMDA (which powers Bard), and Meta's LLaMA. Use cases can span across a variety of functions, from creativity and inspiration, to crafting personalized sales content, writing emails, paragraphs, or even whole papers. Such tools have also shown promise in summarizing documents, responding to customer inquiries, and more. However, these tools are imperfect and may generate false or misleading content, or may misrepresent company views when used as chatbots. In their current state, we view generative models as complementary tools for human creativity and analysis, not replacements.
- **Code** is perhaps the most transformative modality. Tools like CoPilot, GPT Engineer, and even ChatGPT are scaffolding the capabilities of everyday users and professionals alike. While generative coding systems are still in their infancy, they are still a capable tool in the hands of an experienced early-to-mid career professional or a hobbyist with time to spare. From debugging to outsourcing basic tasks to a quick and rudimentary mock-up of complex ideas, AI increasingly has utility as a coding companion.
- **Audio** In mid-June, Meta announced 'Voice Box' a generative AI capable of replicating human voices with comparatively tiny snippets of audio data. Meta hopes these tools could form the basis of live translation. Elsewhere, tools exist for all manner of audio tools, from enhancing functions that cut background noise to recognizing copyrighted music.
- **Image** AI image generation exploded onto the scene with the likes of Midjourney, Stable Diffusion, and DALL-E in 2022. However, multimodal image-focused AI tools are now being put together as composite offerings by the likes of Adobe. Users are able to use text-to-image AI tools to edit images with ease, highlighting backgrounds and describing new scenes, from direct replacements to image extensions. Such capabilities are likely to find their way into video, too.
- **Video** has already been treated by AI in the form of video upscaling and computer-generated imagery both in films and games. However, growing capabilities have opened the door to new opportunities for subtle video editing. For example: filmmakers are experimenting with language dubs that sync to reconfigured-, AI-generated mouth movements for actors, bringing a new level of realism for audiences who view in a language other than the original.

Figure 3: Generative AI capabilities by modality and supporting models

	Text	Code	Image	Audio	Video	Other
Applications	Note-taking	Code writing	Image generation	Voice synthesis	Video editing	3D Modelling
	Writing	Code documentation	Media & Advertising	Automated narration	Video generation	Gaming
	Customer support	Interpretation	Design	Accessibility	Animation	Biology & Chemistry
	Sales & Marketing	Debugging	Creative inspiration	Creative inspiration		
	Creative inspiration					
Models	OpenAI GPT -3 and -4	OpenAI GPT -3 and -4	OpenAI GPT -4	OpenAI Whishper	Microsoft X-CLIP	Galactica
	DeepMind Gopher	Tabnine	OpenAI DALL-E 2	AudioLM	Meta Make-A-Video	Minerva
	Meta LLaMA	Stability.ai	Craiyon	Jukebox	Phenaki	Alphatensor
	Hugging Face Bloom	Codex	Imagen		Soundify	GATO
	Cohere	Alphacode	Muse			Human Motion Diffusion Model
	Anthropic					
	Google PaLM					
	Google LaMDA					

Figure 4: Example Outputs from Generative AI

Image Generation

Prompt: "Teddy bears working on new AI research underwater with 1990s technology"

Sources: DALL-E 2 via OpenAI and Wikipedia, under public domain. For illustrative purposes only.

Text Generation

Prompt: "Explain wealth management as if you were talking to a high school student. Please use only four sentences."

ChatGPT: Wealth management is like having a personal coach for your money. It involves making smart decisions about how to save, invest, and grow your wealth over time. Wealth managers help you set financial goals, create a plan to achieve them, and provide guidance along the way. They also help you navigate through complex financial options and strategies to make sure your money works for you.

Sources: OpenAI's ChatGPT powered by GPT-3, and Invesco. For illustrative purposes only. Not intended as financial advice.

The Big Question: Will Generative AI Lead to Productivity Increases?

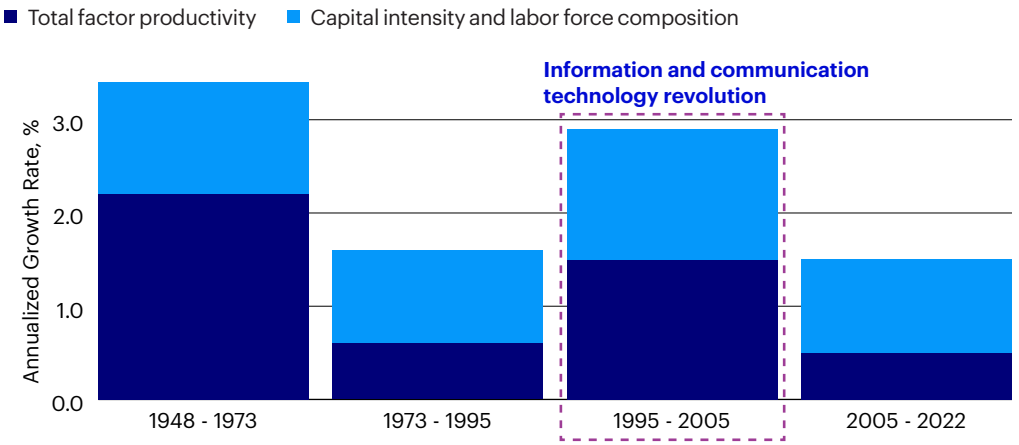
As businesses pursue generative AI for use in products, the broader economy may benefit from the effects of implementations. AI may augment or eventually substitute current labor practices. Productivity, as measured by total factor productivity (TFP) growth, may rise as workers are paired with AI tools that make workers more efficient in their day-to-day tasks.

In one example, Github—a code collaboration and repository website—paired developers with an AI coding assistant branded as CoPilot. The study found that those using the AI system performed a given software development task 56% faster than the control group.[3]

The hope underlying this is for a productivity increase similar to the information and communication technology revolution, which saw productivity rises at a rate three times that of recent decades. Indeed, AI may meaningfully uplift economic growth and help offset the headwinds from slowing population growth. We will treat this topic with greater detail in Part 4 of this series, where we will explore the macroeconomic implications of recent advancements in AI.

Figure 5: US Labor Productivity Growth, 1948 to 2022

Components of labor productivity growth



Sources: US Bureau of Labor Statistics and Brookings Institution, as of 31 December 2022.

3 Source: Github: Research: quantifying GitHub Copilot's impact on developer productivity and happiness from 7 September 2022.

Treading Carefully: Speedbumps and Shortcomings

While we see meaningful opportunity from recent generative AI tools, we must also recognize the weaknesses and challenges that remain in such tools. We expect—and hope for—continuous improvements that minimize or solve the shortcomings noted below, but for now these remain considerable roadblocks to further adoption and integration. In its current form, we view AI as a tool and a fascination, not as a replacement for information workers and services.

- **Hallucinations** – Since large language models are essentially clever parrots, they can produce wrong (untrue, misleading, or fictitious) information and present it with confidence We should be prudent when relying on outputs for critical work.
- **Bias** – Models are only as good as the data that we feed them. Datasets may have ingrained biases. In one example, an AI recruiting system found that playing lacrosse in high school was an indicator of success.[4] Users of such tools must be vigilant to their shortcomings.
- **Safeguards** – Generative models may have access to potentially harmful content that may otherwise be difficult to access. For example, how to assemble dangerous weapons. Users are very clever at breaching safeguards.
- **Intellectual Property** – Since AI is built on previous content, it can use work from others—oftentimes without authorization.
- **Cybersecurity** – Generative AI may provide a greater userbase with coding tools that enable malicious users to create malware programs more easily. Generative tools could presumably be used for repeated and differentiated cyber attacks as well. Furthermore, AI may well empower a new generation of phishing approaches. Encouragingly, AI could also be an effective response to such threats through more sophisticated threat detection and resolution.
- **Fake Content** – Generative tools may be used for propagation of fake or misleading stories, images, videos, voiceovers, and more. In May 2023, a manufactured photo—possibly from a generative AI tool—briefly moved markets when it circulated on Twitter.[5] Other threats exist, such as AI that allows users to take on the likeness and voice of real persons, colloquially referred to as "deepfakes." We expect more such examples as familiarity with and sophistication of AI increases.

Long-term Differentiation

Going forward, we expect that large, multimodal generative AI systems are likely to be packaged as foundation models on which companies can build custom solutions and users can interact with apps. In essence, these are pre-packaged platforms which companies can use as a "foundation" for more specialized solutions.

For now, Foundation Models (whether OpenAI's, Meta's, Google's, or others) have differences. But they all use natural language, and they are increasingly trained on the same or very similar data. Trillions of parameters for training represents a mind-boggling amount of data, but the world is surprisingly repetitive. How many songs sound alike? How many sentences are used over and over? In a universe where a handful of variables usually have enormous explanatory power, there is every chance that a large but finite and entirely reasonably sized data set is a sufficient model of the world that a data set of equal size is functionally indistinguishable. In that world, network effects and user experience might be the only meaningful differentiating factors left.

4 Source: Harvard Business Review, All the Ways Hiring Algorithms Can Introduce Bias, 6 May 2019.

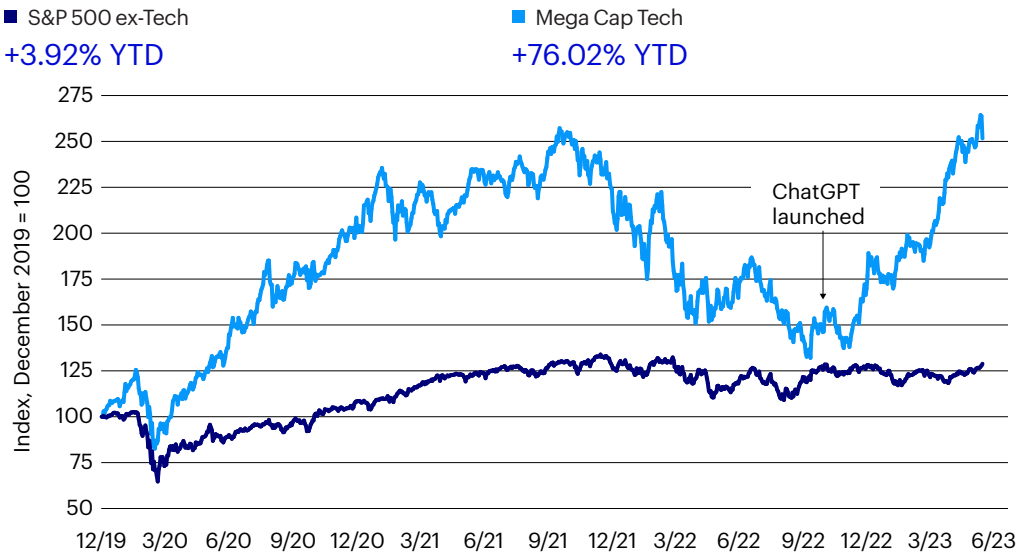
5 Source: Bloomberg, How Fake AI Photo of a Pentagon Blast Went Viral and Briefly Spooked Stocks, 23 May 2023.

Conclude

Generative AI is the overnight success almost a decade in the making. The earliest models appeared around 2014. 2022 was the year of Generative AI's debut, but only with the release of ChatGPT did Generative AI capture our collective imaginations, quickly leading to an explosion in AI tool development and use. Now, half-way through 2023, AI has caused a veritable frenzy in markets, with "AI" companies outperforming in the tech sector and outperforming across a variety of indices. In our view, we are likely entering (or already in) an AI bubble. We also believe that AI-related plays might pay off in the long-run, but we would strongly caution investors that, while some AI technologies have an established history as tools in tech companies, business models like 'AI as a Service' and dedicated AI tools for business and consumer use are yet to be proven. We would also caution investors that many AI technologies, especially generative technologies, exist and operate in a changing regulatory landscape which may further impact any expected returns.

Figure 6: AI leaders dominate US large-cap performance

Equity performance, December 2019 = 100



Sources: Bloomberg and Macrobond, as of 20 July 2023.

Notes: Data has been re-based to 31 December 2019. The 'S&P 500' Index is a market capitalization-weighted index of the 500 largest US stocks. 'S&P 500 ex-Tech' subtracts tech stock performance from the S&P 500's performance. 'Mega Cap Tech' refers to the NYSE FANG+ Index, a rules-based, equal-weighted equity benchmark that tracks the performance of 10 highly-traded growth stocks technology and tech-enabled companies across the technology, media & communications and consumer discretionary sectors.

Past performance does not guarantee future returns. An investment cannot be made directly in an index.

Investment risks

The value of investments and any income will fluctuate (this may partly be the result of exchange rate fluctuations) and investors may not get back the full amount invested.

Important information

This document is intended only for Professional Investors in Hong Kong, for Institutional Investors and/or Accredited Investors in Singapore, for certain specific sovereign wealth funds and/or Qualified Domestic Institutional Investors approved by local regulators only in the People's Republic of China, for certain specific Qualified Institutions and/or Sophisticated Investors only in Taiwan, for Qualified Professional Investors in Korea, for certain specific institutional investors in Brunei, for Qualified Institutional Investors and/or certain specific institutional investors in Thailand, for certain specific institutional investors in Malaysia upon request, for certain specific institutional investors in Indonesia and for qualified buyers in Philippines for informational purposes only. This document is not an offering of a financial product and should not be distributed to retail clients who are resident in jurisdiction where its distribution is not authorized or is unlawful. Circulation, disclosure, or dissemination of all or any part of this document to any unauthorized person is prohibited.

This document may contain statements that are not purely historical in nature but are "forward-looking statements," which are based on certain assumptions of future events. Forward-looking statements are based on information available on the date hereof, and Invesco does not assume any duty to update any forward-looking statement. Actual events may differ from those assumed. There can be no assurance that forward-looking statements, including any projected returns, will materialize or that actual market conditions and/or performance results will not be materially different or worse than those presented.
All material presented is compiled from sources believed to be reliable and current, but accuracy cannot be guaranteed. Investment involves risk. Please review all financial material carefully before investing. The opinions expressed are based on current market conditions and are subject to change without notice. These opinions may differ from those of other Invesco investment professionals.

The distribution and offering of this document in certain jurisdictions may be restricted by law. Persons into whose possession this marketing material may come are required to inform themselves about and to comply with any relevant restrictions. This does not constitute an offer or solicitation by anyone in any jurisdiction in which such an offer is not authorised or to any person to whom it is unlawful to make such an offer or solicitation.

This document is issued in the following countries:

- in Hong Kong by Invesco Hong Kong Limited景順投資管理有限公司, 45/F, Jardine House, 1 Connaught Place, Central, Hong Kong. This document has not been reviewed by the Securities and Futures Commission.

- in Singapore by Invesco Asset Management Singapore Ltd, 9 Raffles Place, #18-01 Republic Plaza, Singapore 048619.

- in Taiwan by Invesco Taiwan Limited, 22F, No.1, Songzhi Road, Taipei 11047, Taiwan (0800-045-066). **Invesco Taiwan Limited is operated and managed independently.**