



THE FUTURE OF INVESTING

The Growth Opportunity for Artificial Intelligence

Making sense of a technology poised to radically transform everyday life, business, and investment opportunities.

In just a few months, the newest generation of artificial intelligence (AI) technology has electrified computer science, industry, and capital markets.

And for good reason, say experts who discussed the growth potential of generative AI with Matt Orton, CFA, Chief Market Strategist at Raymond James Investment Management. They believe generative AI is poised to turbocharge innovation and reshape the economy for several reasons:

- Unlike anything that came before, they say, generative AI is truly transformative.
- That's because it delivers new tools – effectively, “an exoskeleton for the mind” – to improve efficiency, creativity, problem-solving, and critical thinking for both workers and companies.



Jules White, Ph.D.

Associate Professor of Computer Science and Associate Dean of Strategic Learning Programs
Vanderbilt University School of Engineering



Srin Pajjuri

Managing Director and Senior Analyst
Raymond James



Matt Orton, CFA

Head of Advisory Solutions and Market Strategy
Raymond James Investment Management

- Generative AI requires immense amounts of new computing power. This demand is stimulating rapid growth in semiconductor manufacturing, IT infrastructure, and many related industries.
- Watch for AI to become an accelerator for all industries. It will change business models and spur demand for AI-savvy workers in sectors across the economy.

Like nothing that has come before

Until last year, artificial intelligence had touched most people's lives only in incremental ways, via smartphones, automated driver-assist features in cars, or chatbots that guide service interactions on websites and apps.

That changed on Nov. 30, 2022 with the release of the fourth generation of ChatGPT, which stands for Chat Generative Pre-trained Transformer. GPT-4 attracted 1 million users within five days of its release.¹ It then hit the 100 million-user mark within two months, making it the fastest-growing consumer application in history.²

This unprecedented user growth “brought a moment of realization for many about the rapid progress that had been made in AI technologies over the past few years,” Orton said. Generative AI – a form of artificial intelligence that can create new content such as text, audio, video, images, and computer coding – “is poised to unleash the next wave of productivity and to replace many of the mundane tasks people must do on a daily basis,” he said, both in their personal and professional lives.

What makes generative AI truly different from anything that came before is the range of its capabilities, said Jules White, an Associate Professor of Computer Science and the Associate Dean for Strategic Learning Programs at Vanderbilt University's School of Engineering. At its heart, AI enables machines to interpret information they receive, either in the form of text or images. Until recently, the ability to interpret data tended to be very specialized. It was applicable only to a few specific use cases. It also often was brittle, breaking down if faced with tasks beyond those it was originally developed to perform.

As an example of how much more generative AI can do now, White described the work that a specific feature of ChatGPT – Advanced Data Analysis – can carry out using a Vanderbilt

report on student enrollment. The 20-page PDF is filled with tables that break down enrollment by category, including students' majors, genders and status as either full- or part-time students.

In the past, extracting data from the report would have required manual copying and pasting. Code could have been written to develop a program that would extract information to complete a specific task, like creating charts from tables of data. Moreover, that code would require specific instructions on matters like what colors to use for various categories of information or what labels should be assigned to columns and rows or the X and Y axes of charts.

But with ChatGPT's Advanced Data Analysis feature, White said, someone who doesn't know how to code could upload a PDF of the report into the application and make a request in simple conversational language, like “give me a visualization of all the data on page four.” The application would then produce the visualizations, making all the key decisions along the way, like what colors to use or how categories of information should be identified. If the application needed another tool to complete its work – like a chart-building function that would enable the data to be delivered in a PowerPoint presentation – it would then develop that capability on its own. At completion, the original requestor would receive an array of designed charts without having had to provide any instruction beyond the original request for a visualization of the data.

This, White said, is like using a shovel that could stop you from digging once it sensed that you were about to encounter a rock, then jump out of your hands, build a pickaxe and use it to break up the rock, then jump back into your hands so you could keep digging.

“Until ChatGPT,” he said, “we had never been able to think of computers as being able to work on the fly, building new tools on their own to accomplish a goal.” In his view, the Advanced Data Analysis tool and the many more tools like it that are already available or coming soon “are truly transformative.”

The need for new computing infrastructure

AI capabilities require an immense amount of “compute” – that is, the processing power, memory, networking, storage and

other resources required for the computational success of any program.

“Since the advent of cloud computing, much of the processing power that computers demand comes from data centers with thousands of servers managed by cloud providers,” said Srin Pajjuri, a Managing Director and Senior Semiconductor Analyst at Raymond James.

Meanwhile, AI has evolved quickly from its original incarnations when it enabled computers to perform simpler tasks like deciphering an image – determining a photo was a cat, for example. Now, with the increasingly sophisticated uses of AI, like being able to develop visualizations from columns of data, the compute required to perform these tasks has grown exponentially.

As a result, AI is nearly overloading the circuits. “The semiconductor industry, even with the pace of change predicted by Moore’s Law, is struggling to keep up with the demands of processing this vast amount of data and the training models that are used to interpret that data,” said Pajjuri. The oft-cited Moore’s Law, based on the 1965 prediction by Intel co-founder Gordon E. Moore, is that the number of transistors in an integrated circuit, and therefore the amount of data it can process, doubles every two years.

The circuits that power many of the technologies that have become part of everyday life, like smartphones, are central processing units (CPUs). The servers in cloud computing data centers, though, generally rely on more powerful graphics processing units (GPUs).

A CPU, Pajjuri said, is a big processor that can perform a single task very efficiently. A GPU is thousands of small processors, working alongside each other – literally parallel processing – to crunch data. The way AI models are written to analyze data, GPUs are much better suited to the workload required.

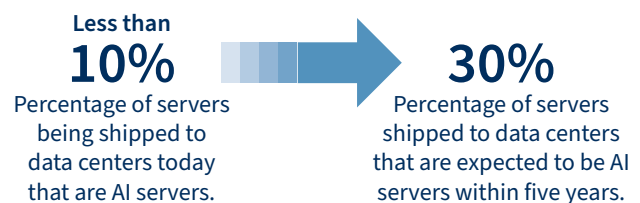
“We’ve seen an explosion in the number of use cases for GPUs with the advent of generative AI,” Pajjuri said. That exponential increase in demand from AI has propelled the soaring stock prices for companies like NVIDIA that manufacture these sophisticated processors.

“The way things are now, the demands from AI don’t meet the hardware that is available,” said Pajjuri. For example, some of the models being developed to deploy ChatGPT require about 10,000 GPUs, running concurrently 24 hours a day, seven days a week, for about three months. “When you consider all the CPUs, GPUs and memory needed, it has a cost of many millions of dollars,” he said.

Fewer than 10% of the servers shipped to data centers today are AI servers, Pajjuri noted.

“We expect that to increase to 30% within the next four to five years,” he said. “But that is a conservative estimate. It wouldn’t be surprising for that number to go much higher, as the amount of compute needed rises significantly.”

Expected growth in AI servers



ChatGPT is a large language model (LLM), which is a type of language model that can perform tasks such as recognizing, translating, predicting or generating text or other content. LLMs use transformer models that are trained using massive data sets. That enables them to learn a vast number of parameters and then understand – and generate – general purpose language. The programs or algorithms in the models can use the data to make predictions or decisions.

The AI models being developed also require vast – and high-quality – data sets to be trained properly, Pajjuri said. The third generation of ChatGPT was trained on 175 billion parameters.³ “The fourth and latest generation of ChatGPT trained on the entire internet, which constituted more than 1 trillion parameters⁴ that were analyzed over a span of three months,” he said.

“The complexity of these AI models is only increasing, and the hardware is struggling to keep up,” Pajjuri added. Everyone is trying to gauge what happens next with AI, but no one can be entirely certain of what the future for AI will look like, he said.

Sectors on the leading edge of change

“It’s clear that generative AI will have a significant impact across sectors and industries ranging from banking and technology to life sciences and biotechnology,” Orton said. Most investors, he acknowledged, are trying to gauge what sectors are likely to feel the biggest impact from AI first.

Look everywhere, White suggested. “Anybody involved in an industry where you use a computer, which is probably just about everybody these days, is going to be impacted,” he said. Areas that he expects to face immediate or far-reaching change include:

Educational institutions. While the initial reaction to AI at educational institutions was to find ways to prevent students from cheating, the response now must go far beyond that, White said: “Educational institutions have to re-architect themselves. There’s going to be a massive demand for reeducating the existing workforce and training the upcoming workforce on how to take advantage of all this. Before we can get started on innovating, we have to build the human capacity” to work with AI.

Law firms and any field immersed in text. “A lot of the tasks that consume massive amounts of time and effort today from humans will be able to get much faster when these tools come out,” White said. “That will be very impactful on many disciplines. People will have to rethink how their time is spent and what services they charge for.”

Software development. “Writing software in the past was a way to control computing” and something “only software developers could do,” he said. Initially, the fact that generative AI enables people to initiate tasks with simple prompts, communicated in everyday language, made some observers fear that AI would eliminate the need for people with degrees in computer science. But White said, “I believe this will actually increase the need for people with skills.” AI will bring “an explosion in the quantity of software that’s out there.” He added, “Anytime you start increasing the volume of software, you massively increase its complexity. ... People trained in computer science will be needed to manage that complexity and manage and maintain all that software over long periods.”

Marketing and communications. “AI is such a capable tool in terms of not only generating first drafts of communications, but also enabling you to do things that would be hard to do any other way, like personalizing messages for every single recipient,” he said. “That scale of personalization, done in a much more sophisticated way, is now possible with these types of tools.”

Ultimately, White said, very few industries or disciplines will go untouched by AI.

“Inevitably, AI will filter into every domain” given “its ability to help humans think outside the box and see information from new perspectives,” he said. “These capabilities are going to be hugely impactful across many disciplines simultaneously.”

He also noted, “There has been a lot of talk about AI being biased, but we also know humans bring biases. For example, once a person develops one way to solve a problem, their future decisions are influenced by confirmation bias that sees any new evidence as proof of their already held views.”

White said AI can help people break free of these cognitive traps, brainstorm and see different perspectives by addressing questions such as:

- “What are 10 other ways to solve this problem?”
- “What are 10 assumptions we’re making that are potentially wrong?” or
- “What are 10 middle-ground solutions that could help us negotiate workable compromises between two polar sets of needs?”

Understanding the investment implications

Ever since ChatGPT-4 was introduced and everyone got a clearer picture of the potential of AI, the stock prices of companies with AI-related business have soared, even if there have been a few minor dips along the way.

Perhaps the biggest beneficiary of the AI enthusiasm to date has been NVIDIA. It “is the primary supplier of the GPUs that are essential to train these models, and it has seen very strong growth as result of generative AI,” said Pajjuri. “As of now, the demand is still outpacing supply, and that is probably going

to continue for the next few quarters.” He noted that few competitors in this space can match the advantages NVIDIA has.

But meeting the demands created by AI is not only a matter of delivering raw computing power.

“Graphics processors are not as easy to program as a CPU,” Pajjuri said. “Historically, there hasn’t been as much infrastructure to train people how to program a GPU. NVIDIA created its own ecosystem and its own software that makes it easy to program GPUs.”

He noted that the firm has also built many tools and libraries that serve as foundation models that can enable its clients to create the equivalent of their own customized ChatGPT for their industry or company.

“NVIDIA is not just a chip supplier,” he said. “When it comes to AI, they’re a full stack solutions provider.” Serving the “full stack” means the firm can help clients meet all their front- and back-end technology needs as they deploy AI capabilities.

A GPU, however, is not the only way AI applications can be processed. Another popular type of chip that can be used is known as “custom silicon.” As Pajjuri explained, “Some of the companies that have the scale to design their own chips are already doing so,” though even mega-cap tech companies need to outsource various aspects of the work. Broadcom has become a key supplier of these kinds of chips.

“In terms of the degree to which AI is contributing to a firm’s revenue, Broadcom is probably second to NVIDIA today,” Pajjuri said. Other beneficiaries of the race to build AI capabilities include Advanced Micro Devices (AMD), which he said already competes with NVIDIA as a supplier of GPUs and data centers and also plans to introduce a memory accelerator for large generative AI models by the end of 2023.

Meanwhile, Marvel Technology delivers networks to meet the bandwidth demands of AI. Pajjuri said firms like AMD and Marvel, as well as Intel and Micron, are poised not only to reap the benefits of AI, but also to benefit from government spending being channeled to the semiconductor industry through the CHIPS (Creating Helpful Incentives to Produce Semiconductors) and Science Act.

A number of companies, Pajjuri said, have already seen their stock prices appreciate because of their AI-related business

even if they haven’t yet experienced a significant contribution from AI to their revenues.

Key players in the AI value chain

Pajjuri provided an example to further illustrate all the key players needed to support any AI application. “If a large firm were looking to implement an AI application, it might build it in-house or go to a public cloud computing provider” with the compute to train AI models.

These mega-cap tech firms “have built the necessary infrastructure by having thousands and thousands of GPUs, known as clusters, connected to work together in parallel to train the AI models,” Pajjuri said.

The servers are sometimes built by large U.S. technology firms, but some are manufactured in Asia by a third-party original design manufacturer (ODM). Many of the key components for these servers also come from U.S. suppliers.

The infrastructure of AI requires GPUs, CPUs, which are still needed, as well as a lot of memory, plus networking chips, which themselves require the most advanced semiconductor manufacturing. Today, Taiwan is home to the dominant provider in that space, although major tech firms in South Korea and the United States are trying to enter this market as well. Meanwhile, the Taiwanese manufacturer relies on a single Dutch company for advanced lithography, a technology that is essential to the manufacturing of these chips, while other suppliers provide equipment and applied materials needed to make advanced chips.

“There are more layers in the supply chain,” Pajjuri said. For semiconductor manufacturing, these include immense global networks of raw material production, refining, product development, manufacturing, packaging, delivery, and corporate interdependency. During fabrication – a process that can take up to 26 weeks – a semiconductor can cross international borders up to 70 times before reaching its final destination, according to the Semiconductor Industry Association.⁵

Much more than a bubble

The dot.com frenzy of the late 1990s painfully taught investors the risks of getting too caught up in technological advances that could ultimately prove to be a bubble. It also showed “that

the initial pioneers of certain technologies don't always come to be the long-term winners," Orton said.

Now, whether companies realize a significant return on their large and growing investments in AI will determine whether this is a bubble, Pajjuri said.

"There are a lot of use cases for AI being developed, and several are already gaining traction," he said. These include AI-driven search engines and work-related software for composing documents, spreadsheets and presentations. Using simple prompts, products in development are expected to help generate or edit text, work with data to create graphs or analyze trends, create presentations, draft e-mails or summarize email threads, create meeting materials, transcribe sessions or provide debriefs that can be reviewed by anyone who missed a meeting or arrived late.

White noted that some may believe AI is still in the early "inflated expectations" part of the hype curve, which tracks the life cycles of new technologies through five distinct phases. He suggested, however, that "it's not even registering yet in terms of what it can do." The scope of AI's potential will become much clearer once more people use it on a daily basis, he predicted. People will soon begin interacting with generative AI all the time, using it, for example, with their e-mail through applications now in development by large, well-known email providers, White said.

The ability to interact with computers with simple conversational language stands to be as profound a change as the one ushered in by graphical user interfaces (GUIs). With

GUIs, people could use more simple and intuitive icons and menus without having to enter disk operating system (DOS) commands to use a computer. Similarly, AI takes ease-of-use many steps forward because people can deliver directions to a computer in straightforward everyday language.

"With a graphical user interface, you were stuck pushing buttons and interacting with it however it had been designed by the programmer and you couldn't go beyond that," said White. "Generative AI is going to be the new interface to computing because it can allow the user to do things that the designers didn't envision originally."

An 'exoskeleton' for the mind

Fears that AI could eventually replace people in the workplace are unwarranted, in White's view.

AI is "all about augmented intelligence," he said. It will amplify people's creativity and their problem-solving and critical-thinking skills" and allow them "to do things bigger, better and faster."

"I think of AI as like an exoskeleton for the mind that gives you superpowers in terms of how quickly you can analyze data and build things" with the use of the technology, White said. He expects generative AI to become an accelerator for all industries as more people start using it.

"What you're going to see is that people will experiment with it," he said. "They will use AI as a thought partner to come up with all kinds of ideas and start turning those ideas into businesses."

Notes:

1. "A Short History of ChatGPT: How We Got to Where We Are Today," Forbes, 5/19/23
2. "ChatGPT sets record for fastest-growing user base – analyst note," Reuters, 2/2/23
3. "OpenAI Presents GPT-3, a 175 Billion Parameters Language Model," NVIDIA Technical Blog, 7/7/20
4. "How ChatGPT and Other LLMs Work—and Where They Could Go Next," Wired, 4/30/23
5. "Building Resilient Supply Chains, Revitalizing American Manufacturing, and Fostering Broad-Based Growth," U.S. White House report, June 2021, citing data from the Semiconductor Industry Association.

Disclosures

Investing involves risk, including risk of loss.

Diversification does not ensure a profit or guarantee against loss.

This material is a general communication being provided for information purposes only. It is educational in nature and not designed to be taken as advice or a recommendation for any specific investment product, strategy, plan feature, or other purpose in any jurisdiction. Nor is it a commitment from Raymond James Investment Management or any of its affiliates to participate in any of the transactions mentioned herein.

Any examples used are generic and for illustration purposes only. This material does not contain sufficient information to support an investment decision, and you should not rely on it in evaluating the merits of investing in any securities or products. In addition, users should make an independent assessment of the legal, regulatory, tax, credit, and accounting implications and make their own determinations together with their own professionals in those fields.

This document is a general communication being provided for informational purposes only. It is educational in nature and not designed to be taken as advice or a recommendation for any specific investment product, strategy, plan feature, or other purpose in any jurisdiction, nor is it a commitment from Raymond James Investment Management or any of its affiliates to participate in any of the transactions mentioned herein. Any examples used are generic and for illustration purposes only. This material does not contain sufficient information to support an investment decision, and you should not rely on it in evaluating the merits of investing in any securities or products. In addition, users should make an independent assessment of the legal, regulatory, tax, credit, and accounting implications and make their own determinations together with their own professionals in those fields.

Any forecasts, figures, opinions or investment techniques and strategies set out are for information purposes only based on certain assumptions and current market conditions and are subject to change without prior notice. All information presented herein is considered to be accurate at the time of production, but no warranty of accuracy is given and no liability in respect of any error or omission is accepted.

It should be noted that investment involves risks. The value of investments and the income from them may fluctuate in accordance with market conditions and taxation agreements, and investors may not get back the full amount invested. Both past performance and yields are not reliable indicators of current and future results. Past performance does not guarantee or indicate future results.

There is no guarantee that any investment strategy will work under all market conditions and each investor should evaluate their ability to invest for the long term, especially during periods of downturn in the market. Investing involves risk and may incur a profit or loss. Investment returns and principal value will fluctuate so that an investor's portfolio when redeemed may be worth more or less than their original cost.

Growth companies are expected to increase their earnings at a certain rate. When these expectations are not met, investors may punish the stocks excessively, even if earnings showed an absolute increase. Growth company stocks also typically lack the dividend yield that can cushion stock prices in market downturns.

The views and opinions expressed are not necessarily those of the broker/dealer or any affiliates. Nothing discussed or suggested should be construed as permission to supersede or circumvent any broker/dealer policies, procedures, rules, and guidelines.

Definitions

The CHIPS (Creating Helpful Incentives to Produce Semiconductors) and Science Act is federal legislation passed in 2022. It aims to increase investments in U.S. semiconductor manufacturing capacity, while also aiming to support the development leading-edge technologies, such as quantum computing, AI, clean energy, and nanotechnology, while also looking to create high-tech hubs that can foster a larger and more inclusive science, technology, engineering, and math (STEM) workforce.

Cloud computing is the practice of using a network of remote servers hosted on the internet to store, manage, and process data, rather than on a local server or a personal computer.

Compute is a generic term used to reference processing power, memory, networking, storage, and other resources required for the computational success of any program. For example, applications that run machine learning algorithms or 3D graphics rendering functions require many gigs of RAM and multiple CPUs to run successfully. In this case, the CPUs, RAM, and Graphic Processing Units required will be called compute resources, and the applications would be compute-intensive applications.

The CPU (central processing unit), or simply processor, is the main chip in a computer responsible for carrying out all tasks. It is responsible for telling all the other components in a computer what to do, according to the instructions it is given by the programs (software) running on that computer.

Confirmation bias is the tendency for people to process information and make decisions by looking at, relying upon, or interpreting information that is

consistent with their own experiences or existing beliefs.

Generative artificial intelligence (AI) is a form of artificial intelligence that can create new content that includes text, audio, code, video, and images.

The graphical user interface (GUI) is a form of user interface (UI) that allows users to interact with electronic devices through graphical icons and audio indicators such as primary notation, instead of using text-based user interfaces, typed command labels, or text navigation.

A graphics processing unit (GPU) is a specialized electronic circuit initially designed to accelerate computer graphics and image processing (either on a video card or embedded on motherboards, mobile phones, personal computers, workstations, and game consoles).

The Gartner hype cycle tracks the life cycle of a new technology through five key phases that can be mapped on a curve: 1. A trigger for new innovation; 2. A peak of inflated expectations; 3. A trough of disillusionment; 4. A rising slope of enlightenment; and 5. A plateau of productivity as mainstream adoption starts to take off. Gartner is a corporate research and consulting firm.

Large language models (LLMs) are artificial intelligence algorithms that can recognize, summarize, translate, predict and generate text, as well as respond to questions in a conversational manner, by massively large sets of data. Large language models learn context and meaning by tracking relationships in sequential data, such as words in a sentence.

A semiconductor lithography system undertakes a process whereby highly complex circuit patterns drawn on a photomask made of a large glass plate are reduced using ultra-high-performance lenses and exposed onto a silicon substrate known as a wafer.

An original design manufacturer (ODM) is a company that designs and manufactures a product that is eventually rebranded by another firm for sale.

About Raymond James Investment Management

Raymond James Investment Management is a global asset management company that combines the exceptional insight and agility of individual investment teams with the strength and stability of a full-service firm. Together with our boutique investment managers – Chartwell Investment Partners, ClariVest Asset Management, Cougar Global Investments, Eagle Asset Management, Reams Asset Management (a division of Scout Investments) and Scout Investments – we offer a range of investment strategies and asset classes, each with a focus on risk-adjusted returns and alpha generation. We believe providing a lineup of seasoned, committed portfolio managers – spanning a wide range of disciplines and investing vehicles – is the best way to help investors seek their long-term financial goals.