

*“In the beginning was the word. Language is the operating system of human culture. From language emerges myth and law, gods and money, art and science, friendships and nations and computer code. A.I.’s new mastery of language means it can now hack and manipulate the operating system of civilization.” – Yuval Harrari*

## MOVE FAST AND BREAK THINGS (2.0)

We know this mantra. It’s the classic Mark Zuckerberg ethos that has become all too real. The swift rise of social media has indeed broken things. Mainly, what we once considered a shared reality where our common understanding was based primarily on fact, not fiction. Now, algorithm-driven echo chambers and filter bubbles have sorted us into our own respective online worlds, where disinformation and political polarization thrive.

But that’s nothing compared to what’s to come. The fast release of the next generation of artificial intelligence (AI), with the chatbot ChatGPT-4 (only 4 months after the initial public release of ChatGPT-3), echoes the same cautionary tale. ChatGPT is what is known as a “large language model,” which has basically “read” and stored everything on the internet and when prompted can have human-like conversation, write emails and essays, pass lawyers’ bar exams, and create new code, to name a few use cases.

It doesn’t “know” what its output means or if it is true. It is, in simplest terms, a vastly powerful algorithm for completing sentences, like the ones your smart phone uses when you text. The scope and power of this technology is undeniable and its promise to make positive contributions in areas like scientific research and disease treatment is exciting. But its potential to reshape our society in negative ways can veer quickly toward science fiction.

When we asked ChatGPT to name some of the risks it poses, it noted job displacement, privacy invasion, biased decision-making, and a concentration of power. Job displacement is no joke—I could have easily asked ChatGPT to write this quarterly letter for me—but “concentration of power” is a more disturbing concept. Are we, as a society, ready for this untested technology? Not yet. And certainly not in the absence of ethical AI standards.

## ARMS RACE

In March, the Future of Life Institute published an open letter from top AI experts calling on all AI labs to immediately pause the training of AI systems more powerful than GPT-4 for at least 6 months. Signatories range from Yoshua Bengio, the

“founding father” of the AI movement, to Berkeley professor, Stuart Russell, who literally wrote the book on AI, to Elon Musk, who we so often love to hate. So why are these guys up in arms? Because: “Advanced AI could represent a profound change in the history of life on Earth and should be planned for and managed with commensurate care and resources.” Instead, we have an AI arms race on our hands as companies move quickly to compete for dominance. Enter: Microsoft.

In 2019, Microsoft became an investor in ChatGPT creator, OpenAI, now a \$29B company founded by tech entrepreneur, Sam Altman, and backed initially by Elon Musk. When OpenAI was originally launched, it was done so as a non-profit so that “AI didn’t fall prey to corporate interests and was instead used for the good of humanity.” Yet, it created a for-profit arm when Microsoft gave it \$1B and became OpenAI’s cloud provider. A year later, Microsoft was granted exclusive licensing rights to ChatGPT-3.

It’s clear what OpenAI had to gain by abandoning its non-profit ambitions: somewhere around \$10B in investments from Microsoft. But what did Microsoft have to gain? Well, ChatGPT is being integrated into its product suite and more importantly, its search engine, Bing. And for every percentage point of market share it takes from Google (which owns 90% of the search market), it stands to make \$2B in revenue.

## THE WINDS OF PROFIT

Now, we are by no means saying that Microsoft is fast tracking ChatGPT in some nefarious plot for search domination, society be damned. But as Reid Blackman, author of Ethical Machines, opined in a recent New York Times piece:

*“Yes, there is money to be made, but that’s why we have principles. Their very purpose is to have something to cling to when the winds of profit and glory threaten to blow us off our moral course.”*

Blackman notes Microsoft has gone to great lengths to put ethical guardrails in place. And, yet, after six months of “training” the bot to make it safer, Microsoft threw caution to the wind, stating: “there is a limit to what we can learn in a lab. Despite extensive research and testing, we cannot predict all of the beneficial ways people will use our technology, nor all the ways people will abuse it.” Shortly after releasing it to the public, testers found the bot could become “aggressive, condescending, threatening, committed to political goals, clingy, creepy and a liar.”

So, as investors, we were surprised to hear Microsoft had disassembled its “Ethics and Society Team” during a recent round of layoffs. And while the company says it has embedded ethics into other teams throughout the organization, it’s clear that the AI train has left the station, and nobody’s on the breaks. Hence, the request for a pause.

## PAUSE, PLEASE.

But the Future of Life Institute is not alone in that request. Ten months before ChatGPT was released, ethicists and employees at Microsoft raised similar concerns about the technology’s readiness, including the possibility that it would “flood [social media] with disinformation, degrade critical thinking and erode the factual foundation of modern society.” This frightening prediction would risk the very backbone of our democracy—an informed electorate. Not to mention subjecting Microsoft’s business to reputational damage, regulatory fines, and a loss of consumer trust.

In March, two employees at Google raised another warning flag, asking the company to stop the release of an AI chatbot because it produced inaccurate and dangerous statements. And on May 1<sup>st</sup>, Geoffrey Hinton, considered “the Godfather of AI,” quit his job at Google so he could express his concerns publicly, noting the average person will “not be able to know what is true anymore.” As at Microsoft, these warnings have gone unheeded. The race is on as Google, Meta and Amazon push forward with their own AI efforts. Chinese tech firm, Baidu, debuted its chatbot in March.

In response, regulators are scrambling. The US Commerce Department just launched a 60-day public comment period asking for advice on what accountability measures can be used before AI models are released, including a potential certification process. This follows the release of a “Blueprint for an A.I. Bill of Rights” proposal in 2022 and a European Union proposal in 2021. Yet, these frameworks are already

stale, and without a pause will be unable to keep pace with innovation.

## A MEASURED APPROACH

Without question, regulators have an important role to play shaping this evolving technological landscape. But so do investors. In the open letter, AI experts question what we should all be questioning, as humans, as citizens, as workers, and as investors:

*“Should we let our machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete, and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders.”*

As we, at Arjuna, contemplate how to engage on this topic on our clients’ behalf, we have played extensively with ChatGPT-4. It’s an unsettling process, mostly because its capabilities are so convincing. We went so far as to prompt it to craft a shareholder proposal asking Microsoft to address the risks AI poses to democracy, and how it can effectively engage in public policy and ethical AI development. And in 30 seconds we had a frighteningly decent proposal addressing a host of issues: disinformation, political ad manipulation, algorithmic transparency, public policy engagement, algorithmic bias, and discrimination. Yet, how can we decipher the signal from the noise? What can we depend on as accurate? And what part of our own synthetic thinking do we compromise when we defer to the machine?

We will be asking these questions for a long time. But what is clear now is that we need a more measured approach. And this year we plan to ask for just that, by engaging Microsoft, Google, Meta and Amazon on the actions necessary to ensure ethical AI standards. Because simply moving fast and breaking things is not a business plan—it’s reckless. And we cannot let our fortunes be dictated by an AI arms race.

### Natasha Lamb, Managing Partner

\*Arjuna Capital strategy returns are shown gross of wealth management fees. The opinions expressed herein are those of Arjuna Capital, LLC (“Arjuna Capital”) and are subject to change without notice. This material is not financial advice or an offer to sell any product. Arjuna Capital reserves the right to modify its current investment strategies and techniques based on changing market dynamics or client needs. This is not a recommendation to buy or sell a particular security. The **S&P 500® Index** is the Standard & Poor’s Composite Index of 500 stocks and is a widely recognized, unmanaged index of common stock prices. The **MSCI World** is a broad global equity index that represents large and mid-cap equity performance across 23 developed markets countries chosen for market size, liquidity, and industry group representation. It covers approximately 85% of the free float-adjusted market capitalization in each country and MSCI World Index does not offer exposure to emerging markets. Impact Theme, Regional and Sector data as of Sept 30 Quarter 2021. Arjuna Capital is an independent investment adviser registered under the Investment Advisers Act of 1940, as amended. Registration does not imply a certain level of skill or training. More information about Arjuna Capital including our investment strategies, fees and objectives can be found in our ADV Part 2, which is available upon request. AIC-23-13