



SHIKHAR GHOSH

SHWETA BAGAI

## OpenAI: Idealism Meets Capitalism

*No one person should be trusted here. I don't have super voting shares. The board can fire me. I think that's important. [We] think this technology, the benefits, the access to it, the governance of it, belongs to humanity as a whole. If this really works, it's quite a powerful technology and you should not trust one company and certainly not one person.*

— Sam Altman, CEO, OpenAI (June 2023)<sup>1</sup>

*AI's future is being determined by an ideological fight between wealthy techno-optimists, zealous doomers, and multibillion-dollar companies. The fate of OpenAI might hang in the balance, but the company's conceit — the openness it is named after — showed its limits. The future, it seems, will be decided behind closed doors.*

— The Atlantic (November 19, 2023)<sup>2</sup>

*At the heart of this competition is a brain-stretching paradox. The people who say they are most worried about A.I. are among the most determined to create it and enjoy its riches. They have justified their ambition with their strong belief that they alone can keep A.I. from endangering Earth.*

— The New York Times (December 3, 2023)<sup>3</sup>

### Introduction

At OpenAI's first DevDay on November 6, 2023, CEO Sam Altman smiled in the middle of his presentation and asked the audience to calm down, "There's a lot — you don't have to clap each time."<sup>4</sup> Altman was methodically announcing a series of exciting new developments: larger context windows, more economical pricing, and the ability for users to create their own 'agents' within the GPT framework. Eight years since its founding in 2015 and four years since commercialization, OpenAI had become the standard bearer for AI. The startup sparked a global frenzy around generative AI with the launch of its popular chatbot ChatGPT a year earlier. ChatGPT had over 100 million weekly users. An estimated 2 million developers were building on the company's APIs.<sup>5</sup> Altman, as the public face of AI, had transformed OpenAI from a hybrid research organization into a rapidly growing Silicon Valley tech company with annual revenues of \$1.3 billion<sup>6</sup> and a \$86 billion valuation.<sup>7</sup> The company was at the heart of the tech industry's efforts to deploy this groundbreaking technology to businesses and consumers, while also working with regulators on safety guardrails.

---

Professor Shikhar Ghosh and Research Associate Shweta Bagai prepared this case. This case was developed from published sources. Funding for the development of this case was provided by Harvard Business School and not by the company. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Copyright © 2024 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to [www.hbsp.harvard.edu](http://www.hbsp.harvard.edu). This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of Harvard Business School.

Founded as a donation-based nonprofit, OpenAI aimed to address AI's risks to humanity and serve as a counterweight to big tech. Its charter emphasized principles above profit with a mission "to ensure that artificial general intelligence benefits all of humanity."<sup>8</sup> In 2019, facing escalating costs for top-tier talent and computing power required to develop cutting-edge frontier AI models, OpenAI changed its structure from a nonprofit to a 'capped for-profit entity' – overseen by a nonprofit board that guided its ethical direction. The company's products and fundraising was a means to achieve the ultimate goal of artificial general intelligence (AGI) – a machine capable of performing any intellectual task that a human could – to benefit humanity. Microsoft invested \$1 billion into OpenAI's for-profit unit, becoming OpenAI's exclusive Cloud provider with a full and exclusive license to its intellectual property.<sup>9</sup> Microsoft's investment opened the fault lines of the search for increasing model power versus the need for safety and control. This conflict became evident within the AI safety community and the OpenAI leadership. Employees expressed concerns about advanced AI's existential risks and unease about the rapid release and widespread integration of GPT-4 into products.<sup>10</sup>

At the Asia-Pacific Economic Cooperation (APEC) Summit in San Francisco on November 16, 2023, Altman said, "I think this will be the most transformative and beneficial technology humanity has yet invented," adding later, "On a personal note, four times now in the history of OpenAI, the most recent time was just in the last couple of weeks, I've gotten to be in the room when we push the veil of ignorance back and the frontier of discovery forward."<sup>11</sup> No one in the audience was sure what this new breakthrough might be.

Later that morning, Altman received a text from Ilya Sutskever, co-founder and Chief Scientist at OpenAI, to set up a meeting the next day.<sup>12</sup> During the video call at the appointed time, Sutskever informed Altman that he was fired, a decision made by the board citing Altman's alleged lack of candor.<sup>13</sup> Altman was completely blindsided, as was Microsoft, OpenAI's largest investor with a \$13 billion stake. President and co-founder Greg Brockman and a string of senior researchers quit in protest. Altman's firing sparked a cascading chain of events: 95% of OpenAI's 770 employees signed a letter indicating they would resign unless Altman was reinstated as CEO.<sup>14</sup> His firing threatened to jeopardize an employee share sale that would value the company at \$86 billion.<sup>15</sup> OpenAI's investors also clamored for his return. The Board reneged and concluded that Altman was trustworthy enough to be CEO again. Altman reclaimed his position; two of the board members resigned; and the world was left watching and theorizing why this happened in the first place.

At the heart of the OpenAI upheaval was the clash between the commercial impulse to generate profit and fears about AI's potential collateral damage. The two sides had managed to coexist for many years with minor tensions. Altman's return – driven by pressure from investors and an employee rebellion – signaled a consolidation of power and exposed this conflict. The turmoil also indicated that a few individuals had a say in the path of what was possibly the most consequential technology for the world. *The Atlantic* summarized:

Altman's dismissal by OpenAI's board on Friday was the culmination of a power struggle between the company's two ideological extremes – one group born from Silicon Valley techno optimism, energized by rapid commercialization; the other steeped in fears that AI represents an existential risk to humanity and must be controlled with extreme caution.<sup>16</sup>

How did OpenAI come so close to an implosion? Was Altman's removal prompted by those who were fearful of AI and its pace or other factors? Did OpenAI follow a path to commercialization that was aligned with the mission of doing so responsibly and safely? What did this imply for the future of AI – and its development more generally?

*History of Modern AI*

From its inception in the mid-1950s, the field of AI and machine intelligence underwent cycles of high expectations followed by periods known as AI winters. In 1970, Marvin Minsky, an AI pioneer, commented, "In three to eight years we will have a machine with the general intelligence of an average human being."<sup>17</sup> The recurring cycles of high expectations and disillusionment continued until 2012, when the application of neural networks to image recognition showed dramatic improvements in performance – exceeding human capabilities in a few years.

AI development had three distinct phases. Until around 2010, the focus was on creating models based on rules provided by humans, known as GOF AI (Good Old-Fashioned AI). These models applied rules to data to reach conclusions. Many current automated systems used this approach for predictive problems like credit scoring and risk assessment. However, this approach was limited due to the difficulty of converting complex human problem-solving into explicit rules, also known as Polanyi's paradox.<sup>a</sup> Neural networks reversed the sequence. Instead of starting with rules, the designer provided the data and desired outcomes. The model would then develop its own rules. For instance, a model trained on numerous cat images would formulate rules to identify cats. These 'rules' were then used to differentiate between images of cats, dogs, and caterpillars. This method, while effective, posed a challenge: the rules derived by neural networks were impossible to discern, raising concerns that these networks were uncontrollable black boxes.

The approach was augmented by having a model learn rules by following a reward function, or reinforcement learning. The AI was given points for achieving a goal such as winning a game of chess. The model slowly developed strategies for gathering more points and winning the game. This approach enabled DeepMind's AI system AlphaGo to defeat Lee Sedol – a world champion in the ancient game of 'Go' – a task that would be computationally impossible with a rules-based system.<sup>18</sup> This event defied the expectations of most researchers, who had believed it would take at least another decade for AI to develop the necessary sophistication and ingenuity to accomplish such a feat. In 2017, researchers at Google introduced the 'transformer' architecture that trained a machine to predict the next word (or token) in a sequence.<sup>19</sup> This breakthrough created the foundation for generative AI, where the system developed an understanding of the underlying rules in a large set of data and was able to create new content that followed these patterns.

In the modern era, companies and researchers that developed neural networks, reinforcement learning techniques and transformer-based architectures had driven the explosion in AI capabilities. Since 2018, researchers studied the predictable increase in capabilities as models increased their scale – the number of parameters, amount of data, and computational power. The relationship between power and scale was a key driver of the race to develop larger models with the promise of getting to an AGI. At the same time, there was a rapid increase in capabilities in multi-modal models – models that cooperated with each other, and smaller models or agents that were trained to accomplish tasks within specialized intelligence domains.

As AI models became more powerful, two additional developments shaped the AI industry. First, these models became more general. The underlying technology for different classes of problems such as image recognition, voice recognition, translation etc. all converged to the same foundation. As a result, the rate of improvement increased rapidly in all fields simultaneously. Foundation models that were created for text or images and trained on Internet data could be retrained with small datasets to

---

<sup>a</sup> Polanyi's Paradox was the theory that human knowledge of how the world functioned and of our own capability was, to a large extent, beyond our explicit understanding.

become specialized in fields like medical images. Second, the locus of R&D shifted from government funded research labs and academic institutions to commercial organizations. A handful of tech companies and startups were responsible for the lion's share of innovation in the field. Virtually all the large foundation models were aiming to attain AGI. Open AI was originally a response to these trends. The race for dominance in AI, particularly in large models, began in 2010 with the venture funded creation of DeepMind – a company with the explicit objective to “solve intelligence” and develop AGI. This competition in Silicon Valley involved major tech companies and key figures who recognized AI's potential across sectors like education and warfare, and competed for leadership in its R&D.<sup>20</sup> (Exhibit 1: The Battle for Control of AI Development).

## OpenAI's Early History

### *Sam Altman's Journey*

Altman, a longtime enthusiast of science fiction, *Star Wars*, and programming, pursued an unorthodox academic route at Stanford University focused on AI. He dropped out to enter the startup scene, co-founding Loopt, which was part of Y Combinator's first cohort. In 2014 at the age of 28, Altman assumed the leadership of Y Combinator, succeeding Paul Graham. Altman redefined the incubator's scope, extending its influence beyond startup cultivation to broader technological innovation. He initiated a research wing to fund ambitious projects that solved the world's biggest problems. In his mind, AI could potentially address humanity's problems better than humanity itself.<sup>21</sup>

At the time, computer advancements were at a peak: machines were labeling photos and translating text via deep learning and neural networks. Altman was convinced that AGI was within reach. He didn't want to approach large corporations, fearing they might release a possible AGI without adequate guardrails. Instead, he connected with people to start a new kind of AI company: a nonprofit that would direct the field towards responsible AGI. A few months later, Altman raised initial funding for OpenAI from Elon Musk (\$100 million plus his time), and Reid Hoffman (\$10 million).<sup>22</sup> Other prominent backers included Peter Thiel, Jessica Livingston, Amazon Web Services, and YC Research. While recruiting a team, Altman was selective and focused on individuals who shared a belief in the feasibility of AGI, given the skepticism surrounding it. Altman reflected, “Back in 2015, when we were recruiting, it was almost considered a career killer for an AI researcher to say that you took AGI seriously. But I wanted people who took it seriously.”<sup>23</sup> He hired Greg Brockman (CTO of Stripe) and Andrej Karpathy who was at Google Brain (an innovative AI lab) as co-founders.

Altman's most crucial hire was Russian-born engineer Ilya Sutskever.<sup>24</sup> Sutskever had gained prominence as the standout student assisting Geoffrey Hinton in initiating the deep-learning revolution.<sup>25</sup> A gifted AI researcher, he later became a key scientist on the Google Brain team.<sup>26</sup> Altman orchestrated a meeting with Sutskever, inviting him to a dinner alongside Musk. The discussion centered on whether it was possible to catch up to Google and DeepMind and “create a lab which would be a counterbalance.”<sup>27</sup> Altman drew Sutskever into the vision of OpenAI, and a few months later, Sutskever agreed to be co-founder. He eventually took on the role of Chief Scientist and became the driving force behind the company's research.<sup>28</sup>

### *The OpenAI Ethos*

In December 2015, OpenAI was founded as a nonprofit, explicitly prioritizing safety over profit. Altman was concerned about the risks of advanced AI including unexpected negative consequences, misuse for harmful purposes, and a possible existential threat. Positioned as a research lab, OpenAI's charter (developed over two years) laid out a vision that countered the competitive dynamic of

commercial development and replaced it with a cooperative vision for humanity. AGI would be advanced enough to outperform any person at “most economically valuable work,” and demanded responsible oversight. Its structure opposed the tech industry’s “relentless pursuit of scale, a build-first-ask-questions-later approach to launching consumer products.”<sup>29</sup> OpenAI’s primary fiduciary duty was to humanity and not investors or employees. They would “work diligently to minimize conflicts of interest between stakeholders that could compromise broad benefit” and avoid uses of AI or AGI that “unduly concentrate power.”<sup>30</sup> (**Exhibit 2:** The OpenAI Charter).

OpenAI committed to avoid enabling a competitive race particularly in late-stage development of AGI. The charter stated, “If a value-aligned, safety-conscious project comes close to building AGI before we do, we commit to stop competing with and start assisting this project.” OpenAI did not intend to apply for patents. It wanted to make AI safe and accessible by being open source and sharing it with the world. Altman believed humans were benevolent; since OpenAI would equip the majority of these individuals with powerful tools, any bad actors would be outnumbered.<sup>31</sup> Many of OpenAI’s founders were supporters of Effective Altruism, a utilitarian-inspired movement rooted in Silicon Valley that used data to maximize positive impact. Effective altruists were early advocates for addressing existential risks posed by AI, leading many to pursue careers in AI safety.

There was a longtime divide in the broader AI community between those who believed that AI was the biggest opportunity with immense benefits, and others who worried that moving too fast could be dangerous. Accordingly, questions and debate swirled among academics, legislators, founders and users on whether the technology ought to be controlled or released. Proponents of AI safety worried that unregulated advancements could eventually pose a catastrophic threat to humanity. Another group thought the doomers’ perspective was an overblown diversion from the actual risks associated with any technological revolution such as economic disruption, biases, and misuse.<sup>32</sup> A recent subgroup emerged as a response to Effective Altruists, called Effective Accelerationism (“e/acc”). They believed AI should be allowed to accelerate with no gatekeepers standing in the way of innovation.<sup>33</sup>

### *Early Years*

Despite Musk’s backing, an overarching mission, and a stellar group of engineers and researchers, there was no clarity on how OpenAI would achieve its ambitious mission. OpenAI focused on deep learning and neural networks and published a string of research papers. The team experimented with a range of AI applications, from video game algorithms to robotics. Altman said, “We knew *what* we wanted to do. We knew *why* we wanted to do it. But we had no idea *how*.”<sup>34</sup> Sutskever was optimistic, “Chasing AGI wasn’t totally crazy. It was only moderately crazy.”<sup>35</sup> At the time, OpenAI was working in relative obscurity, while DeepMind was center stage with AlphaGo’s success.

Google’s research paper on transformer architecture carved the path forward, allowing a neural net to understand and generate language much more efficiently. While experimenting with transformers, the OpenAI team observed that adding data and computing resources increased model power in a predictable way. Scaling laws indicated that simply feeding a model more data and using more processors enhanced its intelligence, without many changes to the underlying architecture. OpenAI’s first breakthrough came with its Generative Pretrained Transformer model or GPT-1 (with 117 million parameters or variables), that surpassed all previous models in understanding language and generating answers. GPT-1’s most striking outcome was its ability to spontaneously generate appropriate responses in new domains.

The model’s successors GPT-2 (1.5 billion parameters) and GPT-3 (175 billion parameters) represented significant advancements in natural language processing and generation, setting new benchmarks in the field.<sup>36</sup> Sutskever commented, “We’d figured out the formula for progress, the

formula everyone perceives now – the oxygen and the hydrogen of deep learning is computation with a large neural network and data.”<sup>37</sup> As the model’s responses became increasingly refined, OpenAI exercised caution, hesitating to release the program broadly for risk of spreading spam. CTO Mira Murati, who joined in 2018, commented, “We thought that open-sourcing GPT-2 could be really dangerous. We did a lot of work with misinformation experts and did some red teaming. There was a lot of discussion internally on how much to release.”<sup>38</sup>

### *Musk and Altman breakup*

Following Google and DeepMind’s breakthrough with AlphaGo, Musk was concerned that operating OpenAI as a nonprofit would limit its potential to innovate and compete effectively. In late 2017, Musk sought to take control of OpenAI and transform it into a commercial entity, intending to align it with Tesla using Tesla’s supercomputers. Musk proposed taking a majority stake in the company. When Altman and other founders resisted this move, Musk withdrew from OpenAI. He announced his departure in February 2018, ending his active involvement and financial support.<sup>39</sup>

## The Path to Commercialization

### *A New Structure*

Recognizing the need for substantial capital and computing power to attract top researchers and deploy new versions of GPT, Altman sought alternative funding sources. In 2019, OpenAI established a subsidiary with a ‘capped-profit’ business division that could raise money, attract talent, and eventually build commercial products. (**Exhibit 3: Governance Structure**). If the revenue cap was achieved – a figure not disclosed but possibly in the trillions – all profits exceeding this threshold would revert back to the nonprofit research lab.<sup>40</sup> The profit units themselves could be sold on the market like standard equities. *Wired* described the unusual structure:

The novel scheme was almost a quantum approach to incorporation: Behold a company that, depending on your time-space point of view, is for-profit and nonprofit. The details are embodied in charts full of boxes and arrows, like the ones in the middle of a scientific paper where only PhDs or dropout geniuses dare to tread.<sup>41</sup>

The board was responsible for ensuring the safe development of AI and oversaw a nonprofit entity that managed OpenAI’s for-profit ventures. It was obligated to the mission of “ensuring the creation and adoption of safe and beneficial AGI” rather than financial incentives. At any given time, only a few board members had financial interest in the for-profit entity. Members without a financial stake could vote on decisions that potentially conflicted between financial stakeholder interests and the nonprofit’s mission. COO Brad Lightcap affirmed, “It shouldn’t just be in spirit, but encoded in the structure of the company.”<sup>42</sup> Investors were informed about these boundaries and warned that investment in OpenAI ought to be viewed in the spirit of a donation. Lightcap explained, “We have a legal disclaimer that says you, as an investor, stand to lose all your money. We are not here to make your return. We’re here to achieve a technical mission, foremost. And, oh, by the way, we don’t really know what role money will play in a post-AGI world.”<sup>43</sup> OpenAI’s financial documents stipulated an exit contingency for when AGI would alter humanity itself.<sup>44</sup> (**Exhibit 4: Investing Document**).

### *Raising Funding*

Under the new structure, Altman stepped down as President of Y Combinator and became CEO of OpenAI. Greg Brockman became CTO (and Chair of the board), and Ilya Sutskever took on the Chief

Scientist role. Initial investors included Khosla Ventures and Reid Hoffman's charitable organization. Employees were able to claim equity, but Altman did not. Meaningful work was more important to him, "If I didn't already have a ton of money, it would be much weirder. It does seem like people have a hard time imagining ever having enough money. But I feel like I have enough."<sup>45</sup> Initial funding was insufficient to fulfil OpenAI's vision of developing advanced models. Each iteration of the GPT family needed exponentially more power. Altman looked for new sources of funds for this unusual corporate entity, making pitches to multiple entities including U.S. government agencies.

Microsoft, with its vast resources and technical capabilities, particularly in cloud computing, invested \$1 billion in OpenAI.<sup>46</sup> OpenAI received access to Microsoft's Azure cloud services, crucial for training and deploying their large-scale AI models, along with financial backing to continue its R&D. The partnership aligned with Microsoft's strategic interests in AI and cloud computing, allowing them to integrate innovative AI technologies into their products. Despite being a significant investor, Microsoft did not gain board representation. In 2021, the tech giant increased financial support for OpenAI, with an investment of \$2 billion followed by an additional \$10 billion in 2023. Microsoft became the exclusive cloud partner for OpenAI, providing unlimited computing power from its Azure cloud. By this time, OpenAI had already developed GPT-3, which represented a significant leap from its predecessors in terms of capabilities and potential applications. Satya Nadella, Microsoft's CEO, commented that only when he saw GPT-3 in action did he fully grasp the extent of its innovation, observing "emergent properties," referring to the unexpected capabilities that arose as a result of its complex and sophisticated design.<sup>47</sup>

## *A Clash of Ideologies*

How did an organization that was established to counter the power of big tech and had initially pledged to avoid patents, maintain open-source principles, and operate with complete transparency – end up granting an exclusive license for its technology to the world's largest software company? Following its partnership with Microsoft, OpenAI received pushback. Musk tweeted, "This does seem like the opposite of open – OpenAI is essentially captured by Microsoft."<sup>48</sup> He elaborated, "Let's say you founded an organization to save the Amazon rainforest, and instead you became a lumber company, chopped down the forest, and sold it."<sup>49</sup> There were defectors among employees, too. A group of 15 engineers and scientists, dissatisfied with OpenAI's commercial direction left the organization. The team led by researcher Dario Amodei founded Anthropic in 2021 as a public benefit corporation to develop AI according to the principles of effective altruism and focus on safety.<sup>50</sup>

## **The Road to ChatGPT**

### *A Low-key Research Preview*

In the fall of 2022, OpenAI was fully engaged in preparing for the release of its most advanced LLM (large language model) GPT-4. Employees were refining the technology, setting up support infrastructure, as well as outlining policies for acceptable user behavior.<sup>51</sup> The company was using an internal version of GPT with a conversational interface crucial for "truth-seeking," suggesting users could elicit more reliable and comprehensive responses through dialogue.<sup>52</sup> Rumors that the rival company Anthropic was developing its own chatbot changed these plans. In November, OpenAI leadership announced the need to launch a chatbot within weeks. They decided to use an existing model – GPT 3.5 – and framed this as a "low-key research preview" rather than a product launch. Altman and others emphasized kickstarting the "data flywheel" process that aligned with company strategy to gradually release technologies and gather data from people's interactions with AI.<sup>53</sup> This

would eventually feed into the development of its newer, more coherent, and capable successor, GPT-4 (trained with a reported 1.7 trillion parameters). The release would familiarize the public with the inevitable impact of AI on their daily lives. Altman explained, “GPT-4 was a lot to get used to at once.”<sup>54</sup>

Within OpenAI, there was considerable discussion regarding the prudence of unleashing a new conversational tool possessing such incredible capabilities quickly. The company was already stretched thin with GPT-4 and wasn’t fully prepared to manage a chatbot that could alter the risk landscape. A few months earlier, OpenAI had implemented a new traffic tool for monitoring user behavior and was still trying to understand product usage and ways to mitigate abuse. Some employees were more positive, and believed that adapting GPT 3.5 into a chatbot would be a straightforward task, given the model’s intensive testing and refinement.<sup>55</sup>

ChatGPT’s launch on November 30, 2022 was an understated event. Employees not directly involved, including those in safety, were not aware of the release. Others started a betting pool on the number of users during the first week. The maximum guess was 100,000 users; ChatGPT hit 1 million users in the first five days.<sup>56</sup> It was the first model to offer a glimpse of mainstream usability through a public-facing API and marked a turning point in language-based AI. The term “low-key research preview” became a popular meme among employees who transformed it into laptop stickers. Two months after its launch, ChatGPT was the most downloaded consumer app in history with 100 million users. “[ChatGPT] triggered a tech explosion not seen since the internet burst into our lives. Suddenly the Turing test was history, search engines were endangered species, and no college essay could ever be trusted. No job was safe. No scientific problem was immutable.”<sup>57</sup>

### *Advancing the Product Pipeline - and the Internal Divisions*

ChatGPT’s success had a deep impact and exacerbated the existing ideological divisions within the company. Its release prompted OpenAI to develop profitable products – clashing with its declared mission and image as an idealistic research lab. There was a clear path to revenue and profit, and the focus shifted to meeting consumer demands.<sup>58</sup> OpenAI spent over \$540 million in developing ChatGPT, reflecting the excessive training costs. The product team capitalized on the growing momentum and doubled down on commercial efforts. The number of employees working on “applied” AI (or building products) soared.<sup>59</sup> In February 2023, OpenAI launched a paid version of ChatGPT. Microsoft leveraged its multi-billion-dollar partnership by integrating ChatGPT into its Bing search engine. Nadella commented that the new Bing with OpenAI’s technology would make rival Google “come out and show that they can dance.”<sup>60</sup> An API tool was introduced to facilitate the integration of ChatGPT into business applications.<sup>61</sup>

Two weeks later in March, GPT-4 was released with an estimated 1.5 trillion parameters. GPT-4 could pass the bar exam, plan a course syllabus, and author a book within seconds.<sup>62</sup> (**Exhibit 5: GPT-4 Capabilities**). OpenAI forged ahead with releasing new products such as the image generator DALL-E 3, and ChatGPT Enterprise to target companies. The aim was to create a soft landing for “the singularity” (science fiction that described the moment when intelligent technology could no longer be controlled by humanity). Altman commented, “It doesn’t make sense to just build AGI in secret and drop it on the world.”<sup>63</sup> The company generated revenues of \$1.3 billion by selling access to its models to developers and consumers through ChatGPT.<sup>64</sup>

The release of multiple products exacerbated OpenAI’s challenges, putting pressure on the company’s infrastructure and straining employees with mitigating the technology’s risks. Research teams had to redirect computing power to manage increasing traffic, creating repeated server crashes. The traffic-monitoring tool for tracking user activity failed repeatedly, making identifying and



preventing abuse difficult. Although ChatGPT was undeniably delightful and useful, it tended to hallucinate, producing believable but misleading responses. OpenAI's safety teams – focused on refining ChatGPT to reject abusive requests and provide appropriate responses – pushed for slowing down. The API platform saw a spike in fraud. The mounting pressures led to significant stress for employees with some facing mental health challenges. Communication within the company deteriorated. Often, colleagues would only realize someone had been fired when they noticed their absence on Slack.<sup>65</sup> The openness embodied in the company's name had shifted from the radical transparency suggested at launch.<sup>66</sup>

### *Simmering Leadership Tensions*

Among OpenAI's leadership, Sutskever became increasingly concerned whether OpenAI was staying true to its nonprofit mission of developing beneficial AGI. The rapid advancement of OpenAI's LLMs strengthened Sutskever's belief that the arrival of AGI was imminent, and he focused on preventing possible dangers. He took on the role of a spiritual leader leaning on the phrase, "Feel the AGI," referring to the idea that OpenAI was on the cusp of its goal. OpenAI employees created a special "Feel the AGI" reaction emoji in Slack. At a leadership offsite, Sutskever illustrated his commitment to OpenAI's founding principles by commissioning and burning a wooden effigy that symbolized an AI that was "not aligned" with human objectives.<sup>67</sup>

In July 2023, Sutskever created a "superalignment team" to devise ways to control a hypothetical superintelligence (a future technology that Sutskever speculated would outmatch humans in almost every way). Superalignment applied the idea of alignment (making sure a model did what was asked to do, and did not do what it was not asked) to superhuman models. The team focused on tackling upstream AI safety issues (allocating 20% of existing computing resources to this effort) in preparation for the possible emergence of AGI within this decade.<sup>68</sup> This involved developing techniques to ensure that the model adhered to user instructions and did not hallucinate. Team members believed that AI safety measures were inadequate for the level of advancement GPT-4 represented.<sup>69</sup> Sutskever, in particular, believed that AI could one day destroy humanity. He created a hypothetical scenario in which a superintelligence could potentially hide its true behavior and pretend to be aligned when it wasn't. Sutskever told the *MIT Tech Review*:

I'm doing it for my own self-interest. It's obviously important that any superintelligence anyone builds does not go rogue. Obviously. When you have an organization like OpenAI that's moving at a fast pace and pursuing ambitious goals, tension is inevitable. I view any tension between product and research as a catalyst for advancing us, because I believe that product wins are intertwined with research success.<sup>70</sup>

Sutskever was responsible for many technical innovations at OpenAI, encompassing the foundational work leading to a notable scientific breakthrough known as Q\* (Q star). The model could solve unfamiliar math problems, a challenging task for other AI models, signaling that LLMs were beginning to overcome their reasoning limits.<sup>71</sup> Acing math on the level of grade-school students made researchers optimistic about Q\* success and its applicability to novel scientific research. But, speculation was rife: a group of staff researchers sent a letter to the board cautioning that Q\* could "threaten humanity."<sup>72</sup> They highlighted both remarkable capabilities and potential risks of AI. Others believed that AGI was far away, and AI was not yet capable of doing any task that humans could.

### *Mitigating AI Risk*

Since the release of ChatGPT, some scientists and political leaders worried about its risks such as automating jobs, growth of autonomous weapons or misinformation. In March 2023, hundreds of

researchers signed an open letter warning of the “dangerous” arms race in AI and called for a pause in the development of frontier models until stronger governance was in place.<sup>73</sup> Altman did not sign this. Eliezer Yudkowsky, an AI pioneer warned that the letter did not address the seriousness of the situation and asked for too little; he said to “shut it all down.”<sup>74</sup> A few months later, a San Francisco-based nonprofit released a joint statement from CEOs and scientists from tech companies (including OpenAI), “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”<sup>75</sup>

Risks posed by AI could be categorized into three types.<sup>76</sup> Short-term risks involved immediate concerns in existing AI systems, such as privacy breaches, biases, and the spread of misinformation. In the medium term, AI was projected to significantly advance in science and engineering, raising risks of misuse in domains like biology and cybersecurity, potentially disrupting global power balances. Longer-term risks referred to the future evolution of AI, where AI systems could develop increased autonomy and manipulation capabilities, and pose existential threats to humanity. AI could cause widespread devastation either through intentional misuse (e.g., by state actors to create bioweapons) or unintended actions that deviated from their design. Although there were a mix of many kinds of risks and scenarios, the underlining capabilities that powered both benefits and risks were the same.

In 2015, Altman had said, “AI will probably most likely lead to the end of the world, but in the meantime, there’ll be great companies.”<sup>77</sup> (**Exhibit 6: Risks outlined by Altman**). OpenAI’s researchers were confident in their ability to build the models as they understood the risks. Critically examining millions of user prompts would allow them to ensure that future products were ethically aligned. However, the paradox was that despite the efforts to rigorously test and secure their products against issues such as deepfakes, misinformation and criminal spam, future AI models might become advanced enough to outsmart the very humans who created them. Conversely, in the pursuit of enhanced safety, there was a risk of over-restricting these models, potentially diminishing their utility.<sup>78</sup> A Stanford/UC Berkeley study revealed that despite their advanced safety features, newer versions of GPT performed worse than their predecessors – erring in basic mathematical calculations.<sup>79</sup>

### *A Divided Board*

OpenAI’s board had up to nine members. The majority had no financial stake. In 2023, departures of three members (LinkedIn founder Reid Hoffman, director of Neuralink Shimon Zilis, and former Texas congressman Will Hurd) reduced the number to six members (Brockman, Altman, Sutskever), and three independent directors.<sup>80</sup> These were Helen Toner (Director of strategy, Georgetown University’s Center for Security and Emerging Technology); Tasha McCauley (Scientist, RAND Corporation); and Adam D’Angelo (CEO, Quora, the Q&A site). McCauley and Toner were said to have ties to the Effective Altruist movement; all three shared a common worry that AI might surpass human intelligence.<sup>81</sup> After evaluating four candidates for a one position, the directors couldn’t reach a consensus, exacerbating the division between Altman and Brockman versus the other four members over filling the three vacant seats.<sup>82</sup>

In October, Altman had a disagreement with board member Toner over her academic paper on AI safety,<sup>b</sup> criticizing it for unfavorably comparing OpenAI’s safety efforts to Anthropic’s, and potentially harming OpenAI amid an FTC investigation.<sup>83</sup> Despite previously praising Toner for her critical safety insights, Altman expressed in an email, “I did not feel we’re on the same page on the damage of all this. Any amount of criticism from a board member carries a lot of weight.”<sup>84</sup> Altman tried to persuade

---

<sup>b</sup> Andrew Imbrie, Owen Daniels and Helen Toner, “Decoding Intentions: Artificial Intelligence and Costly Signals,” Georgetown University, October 2023, <https://cset.georgetown.edu/publication/decoding-intentions/>, accessed November 2023.

individual board members to dismiss Toner. Subsequently, several members exchanged information, and deduced that Altman may have created a misleading impression suggesting that another board member was in favor of Toner's departure (though this was not the case).<sup>85</sup>

Moreover, some board members were worried about Altman's entrepreneurial ambitions. In late September, Altman flew to the Middle East to raise money to create a chip company to rival Nvidia. In addition to his hundreds of external investments, Altman was seeking billions from Softbank for a new business to make AI-oriented hardware.<sup>86</sup> The board was concerned that Altman wasn't sharing all his plans. Some employees had also expressed concerns with Altman's management style, which led the board to conduct a review.<sup>87</sup> On a secret 15-minute video call, the board members voted one by one to oust Altman. They decided to act quickly and secretly.<sup>88</sup>

## Five Days of Chaos

At its first developer day in November, Altman exclaimed, "What we launch today is going to look quaint relative to what we're busy creating for you now."<sup>89</sup> The company released GPTs, custom versions of ChatGPT that could be built without coding. OpenAI was also negotiating a "tender offer" funding deal to allow employees to sell their shares, potentially creating financial windfall for many. This deal, led by Thrive Capital, valued OpenAI at \$86 billion – nearly tripling its valuation in six months. At the APEC panel later that month, Altman expressed that certain safety concerns raised by critics were misplaced. "The current models are fine. We don't need heavy regulation here, probably not even for the next couple of generations."<sup>90</sup> A few months earlier at the U.S Senate, he had testified, "We think that regulatory intervention by governments will be critical to mitigate the risks of increasingly powerful models."<sup>91</sup>

### *Altman is fired*

On November 16<sup>th</sup>, the Thursday before Thanksgiving, Sutskever initiated an extraordinary series of events. A few hours after the APEC panel, Sutskever texted Altman asking if they could talk the following day. Altman logged on to a video call at noon on Friday. He soon realized that something was amiss: the three independent directors had also joined the call. Reading from a script, Sutskever announced that "a deliberate review process" had found that Altman had not been "consistently candid in his communications" with the board. He was being fired. The board had lost confidence in his ability to be CEO.<sup>92</sup> (**Exhibit 7: Timeline of Events**). The members did not provide specifics on what led to their decision. Altman was surprised, and asked, "How can I help?"<sup>93</sup> The board members urged him to support an interim CEO. He assured them he would; he reconsidered soon after.<sup>94</sup>

The decision to oust Altman shocked both the company's employees and the wider tech community including Microsoft, a major investor with a \$13 billion stake, who learnt of the news 5-10 minutes before it was made public. Some industry insiders likened it to the dramatic firing of Steve Jobs from Apple in 1985, underscoring the decision's significant repercussions. Although Brockman was the board chair, he was not invited to the meeting where Altman was fired. When the news broke, a message was sent to a private WhatsApp group of over 100 chief executives from Silicon Valley firms: "Sam is out."<sup>95</sup> This ignited a flurry of activity with members eagerly asking for details and questioning the reasons for Altman's departure. OpenAI published a blog post that afternoon announcing the firing and the stepping down of Brockman as board chair.<sup>96</sup>

OpenAI's employees were shocked and demanded details.<sup>97</sup> Hours after the firing, in conversation with 15 OpenAI executives, the board members said that Altman had lied to them but could not elaborate for legal reasons. An OpenAI lawyer suggested that board members might be violating their

fiduciary duties if their decision to fire Altman led to the company's collapse. Toner responded, "That would actually be consistent with the mission."<sup>98</sup> She stated that the board's mission was to ensure that AI "benefits all of humanity" even if it meant destroying the company.<sup>99</sup> The firing was driven by a lack of trust rather than safety; dismissing Altman was consistent with the board's duty to ensure that AI systems were built responsibly.<sup>100</sup> In the board's view, OpenAI would be stronger without Altman.<sup>101</sup> Moreover, the continued existence of OpenAI wasn't necessary for the nonprofit's broader mission. Altman's supporters dismissed the board's decision as illegitimate calling the firing a 'coup,' and exerted pressure on the board to resign.<sup>102</sup> Altman's firing led Brockman and a string of senior researchers to quit. CTO Mira Murati was made interim CEO. Murati attempted to hire back Altman and Brockman, while the board undertook its own CEO search.

### *Into the Unknown*

On Saturday, November 18<sup>th</sup>, OpenAI's COO wrote a company memo that the firing was driven by a "breakdown in communications" between Altman and the board rather than "malfeasance or anything related to our financial, business, safety, or security/privacy practices."<sup>103</sup> Board members D'Angelo and McCauley reached out to Altman and requested his assistance in stabilizing OpenAI. They were concerned about the previous day's meeting with OpenAI executives.<sup>104</sup> Customers were contemplating moving to competing platforms, and Google was actively attempting to lure away top talent. Over two dozen Altman supporters gathered at his \$27 million mansion in San Francisco's Russian Hill area, actively campaigning on social media and expressing their objections in private messaging groups.<sup>105</sup>

Murati informed the board she could not continue in her interim role. Altman posted on X: "i love openai employees so much." Murati and dozens of employees replied with emojis of colored hearts.<sup>106</sup> Even as the board considered bringing back Altman, they wanted to bring in members who could keep a check on Altman. They suggested tech veteran Bret Taylor (Twitter's former chair), who rapidly gained unanimous approval and agreed to assist in the negotiation process. The board didn't know how talks with Altman would go and sought another interim CEO as a backup.<sup>107</sup> It was alleged that board members approached the Anthropic's CEO Dario Amodei, and proposed a merger between the two companies.<sup>108</sup> Altman continued to round up his allies, including Microsoft.

### *The Counter-Offensive*

By Sunday, Altman felt confident he would become CEO again. He and his supporters gave the board a deadline: Resign by 10 a.m. or everyone would leave.<sup>109</sup> The board was open to Altman's return if they could agree on new board members; but talks broke down. With no obvious way out, the board told Murati they were hiring Emmet Shear (former CEO of the streaming company Twitch) as the interim CEO. Shear was known for his cautious approach to AI. In September, he posted on X: "The way you make it safely through a dangerous jungle at night is not to sprint forward at full speed, nor to refuse to proceed forward. You poke your way forward, carefully."<sup>110</sup> OpenAI employees responded to Shear's hiring with an emoji indicating strong disapproval, while investor Vinod Khosla publicly asked Shear to resign. Around midnight, in a post on X, Microsoft CEO Nadella expressed his excitement to have hired Altman and Brockman to lead its new AI research team. He welcomed all OpenAI employees at Microsoft, signaling a commitment to OpenAI and its product roadmap.<sup>111</sup>

### *Gathering Support*

By Monday, 700 of 770 OpenAI's employees signed a letter threatening to go to Microsoft if Altman was not reinstated, and demanded that the independent directors resign (**Exhibit 8: Employee Letter**).

Employees had an additional incentive to support Altman since his departure could foil an investment deal that would allow them to sell their stock back to OpenAI – cashing out equity without waiting for the company to go public.<sup>112</sup> Surprisingly, Sutskever also signed the letter. He had been urged to reverse course by Brockman’s wife and earlier posted on X, “I deeply regret my participation in the board’s actions.” Dozens of employees, including Murati, posted an identical message on social media, “OpenAI is nothing without its people.”<sup>113</sup> Altman reposted several of these with heart emojis. With OpenAI nearing collapse, the board was not left with any choice but to open talks again. They were surprised by the employees’ evident readiness to leave the company if Altman was not leading it, and by the extent to which the management team aligned with Altman. The company spokesperson commented, “We believe Sam is the best leader for OpenAI.”<sup>114</sup> D’Angelo suggested former Treasury Secretary Lawrence H. Summers, as a possible board member. Altman agreed: he would give up his board seat and was open to an independent investigation of his leadership and dismissal.<sup>115</sup>

### *Altman’s Return*

By Tuesday evening, an agreement was reached: Altman would resume his role but not rejoin the board. Summers, D’Angelo, and Taylor would become board members, and Microsoft would participate as a nonvoting observer. (**Exhibit 9: An Evolving Board**). Toner, McCauley, and Sutskever would step down. D’Angelo would be the only remaining director from the old six-person board that fired Altman.<sup>116</sup> Toner posted on X on Wednesday evening, “our decision was about the board’s ability to effectively supervise the company,” and “we were not motivated by a desire to slow down OpenAI’s work.” Some expressed that the board was “incredibly principled and brave” to confront Altman.<sup>117</sup>

## Looking Ahead

In a span of 120 hours, Altman was fired; Brockman resigned; two interim CEOs were instituted and demoted; majority of the staff was on the verge of moving to Microsoft; Altman was reinstated; and the Board pared down from six to three. This was interspersed with a host of reactions – shock, surprise, suspense, sentimentality – from staff, investors, and beyond. Sutskever posted (and later deleted) about learning lessons, “[In the past month] one such lesson is that the phrase ‘the beatings will continue until morale improves’ applies more often than it has any right to.”<sup>118</sup>

On November 29<sup>th</sup>, Altman announced that Microsoft would be given an observer seat on the board. The secondary sale of employee stock that valued the company at \$86 billion was on track. The new board of Summers, Taylor, and D’Angelo was firmly in place. Other key employees were back in their positions. Microsoft’s stock reached record highs, although no OpenAI employees had joined their team. Sutskever’s position was still unclear.<sup>119</sup> As the company prepared to move forward, it needed to resolve the role and power of its unusual structure. The arrangement had aimed to balance morality and financial strength, but had led to internal conflict.<sup>120</sup> What should the composition and role of the nonprofit board be? What was the best way for Open AI to reconcile the risks and promise of AI (and potentially AGI) for the benefit of humanity?

Altman would need all the power of his formidable intelligence to bring together the demands of capitalist incentives with the original purpose and intent of Open AI. OpenAI’s superalignment team had assumed AI’s eventual superiority and believed that a model with human-like abilities was inevitable. “But it won’t stop there. We’re going to have superhuman models, models that are much smarter than us. And that presents fundamental new technical challenges.”<sup>121</sup> Peter Thiel summarized his confidence in Altman’s ability to manage this challenge. “No one better navigates between the ‘Scylla of misplaced idealism’ and the ‘Charybdis of myopic ambition.’”<sup>122</sup>



**Exhibit 1** Battle for Control of AI Development*Creation of DeepMind*

In 2010, Demis Hassabis, a neuroscientist and former under-14 world-class chess player, aimed to build AGI. His startup, DeepMind wanted to create AI systems that could learn and excel in many domains. Despite AI's historical underperformance, Hassabis's vision and understanding of its potential risks convinced Peter Thiel to invest 1.4 million British pounds. DeepMind began with mastering classic Atari video games as a demonstration of its technology. Elon Musk also provided funding, breaking his informal rule of only investing in companies he personally managed. His decision reflected a deep interest in AI's potential, along with concerns about its ethical development.

*The Talent Auction*

Geoffrey Hinton, a professor at the University of Toronto, along with two of his graduate students (including Ilya Sutskever, who would later join OpenAI), published a groundbreaking paper indicating AI's capabilities in 2012. Their research highlighted a neural network they had trained to recognize common objects such as flowers, dogs, and cars, by exposing it to large volumes of data. The potential of practical applications of neural networks captured the attention of major tech companies. Hinton and his team found themselves at the center of a high-stakes bidding war. Baidu offered them \$12 million to join the company in Beijing – a proposal that piqued their interest, but one they ultimately declined. Recognizing the value of their work, Hinton organized an auction at an AI conference in Lake Tahoe, with participation from Google, Microsoft, and DeepMind. Bids escalated rapidly; Microsoft left and re-entered the bidding at \$37 million, eventually stepping out. The competition narrowed down to Baidu and Google, with bids hitting \$40 million. Ultimately, Hinton and his team chose Google, halting the auction at \$44 million. Google co-founder Larry Page, familiar with comparable technology at Google Brain (the AI lab), believed that Hinton's research could enhance the work of his own scientists and considered it valuable enough to pursue at any cost.

*Google Acquires DeepMind*

The DeepMind founders had maintained they would stay independent to prevent their technology from becoming dangerous. However, once the tech industry entered the talent race, Hassabis realized he had no alternative but to sell; Google and Facebook emerged as key contenders. Despite a higher offer from Facebook, Google acquired DeepMind for \$650 million in 2014. Unlike Facebook, Google was willing to meet DeepMind's two conditions: ethical oversight by an independent board of technologists, and no military use of their AI technology.

The DeepMind ethics board was established to ensure the responsible development of AI technology, and included Elon Musk, Reid Hoffman (LinkedIn founder and a Microsoft board member) and Google's top executives. At the first meeting in August 2015, there was discussion on potential risks and societal implications of AI. Mustafa Suleyman, co-founder of DeepMind, delivered a presentation titled "The Pitchforkers Are Coming." He expressed concerns about potential negative impacts of AI, such as a surge in disinformation, and the possibility that AI, by automating numerous jobs, could lead to widespread unemployment. The ethics board never had a second meeting. DeepMind's executives, increasingly concerned about the potential misuse of their technology at Google, attempted to break away in 2017. Google responded by significantly increasing their salaries and stock award packages. They remained within Google, despite their initial apprehensions.

*Page vs. Musk*

Elon Musk and Larry Page, once friends and fellow tech visionaries, found themselves at odds over the potential impact of AI. Their relationship soured during a contentious debate at a party in July 2015, over whether AI would elevate or destroy humanity. Musk, fearing AI could inflict grave damage on humanity, clashed with Page's more optimistic view of a digital utopia where humans would eventually merge with AI. Musk expressed dismay over Page's lack of concern for safety; Page labeled Musk a "specieist," implying Musk favored human interests over future digital life forms. This argument marked a turning point in their relationship and led to their eventual estrangement. Moreover, Musk was discontented with losing his influence over DeepMind's trajectory, particularly in light of his escalating concerns about AI's potential dangers. He wanted to start his own lab. A few weeks after the altercation with Page, Musk had dinner with Sam Altman, who supported building responsible AGI; Musk was willing to place a bet on an effort that accounted for AI's impact. Bolstered by hundreds of millions of dollars from Musk and others, OpenAI launched later in the year, and promised to protect the world from Page's vision of the future.

Source: Compiled from Cade Metz, Karen Weise, Nico Grant, and Mike Isaac, "Ego, Fear and Money: How the A.I. Fuse Was Lit," *The New York Times*, <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>, accessed December 3, 2023.

**Exhibit 2** OpenAI Charter (Published April 9, 2018)

This document reflects the strategy we've refined over the past two years, including feedback from many people internal and external to OpenAI. The timeline to AGI remains uncertain, but our Charter will guide us in acting in the best interests of humanity throughout its development.

OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome. To that end, we commit to the following principles:

**Broadly distributed benefits**

We commit to use any influence we obtain over AGI's deployment to ensure it is used for the benefit of all, and to avoid enabling uses of AI or AGI that harm humanity or unduly concentrate power.

Our primary fiduciary duty is to humanity. We anticipate needing to marshal substantial resources to fulfill our mission, but will always diligently act to minimize conflicts of interest among our employees and stakeholders that could compromise broad benefit.

**Long-term safety**

We are committed to doing the research required to make AGI safe, and to driving the broad adoption of such research across the AI community.

We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions. Therefore, if a value-aligned, safety-conscious project comes close to building AGI before we do, we commit to stop competing with and start assisting this project. We will work out specifics in case-by-case agreements, but a typical triggering condition might be "a better-than-even chance of success in the next two years."



### Technical leadership

To be effective at addressing AGI's impact on society, OpenAI must be on the cutting edge of AI capabilities – policy and safety advocacy alone would be insufficient.

We believe that AI will have broad societal impact before AGI, and we'll strive to lead in those areas that are directly aligned with our mission and expertise.

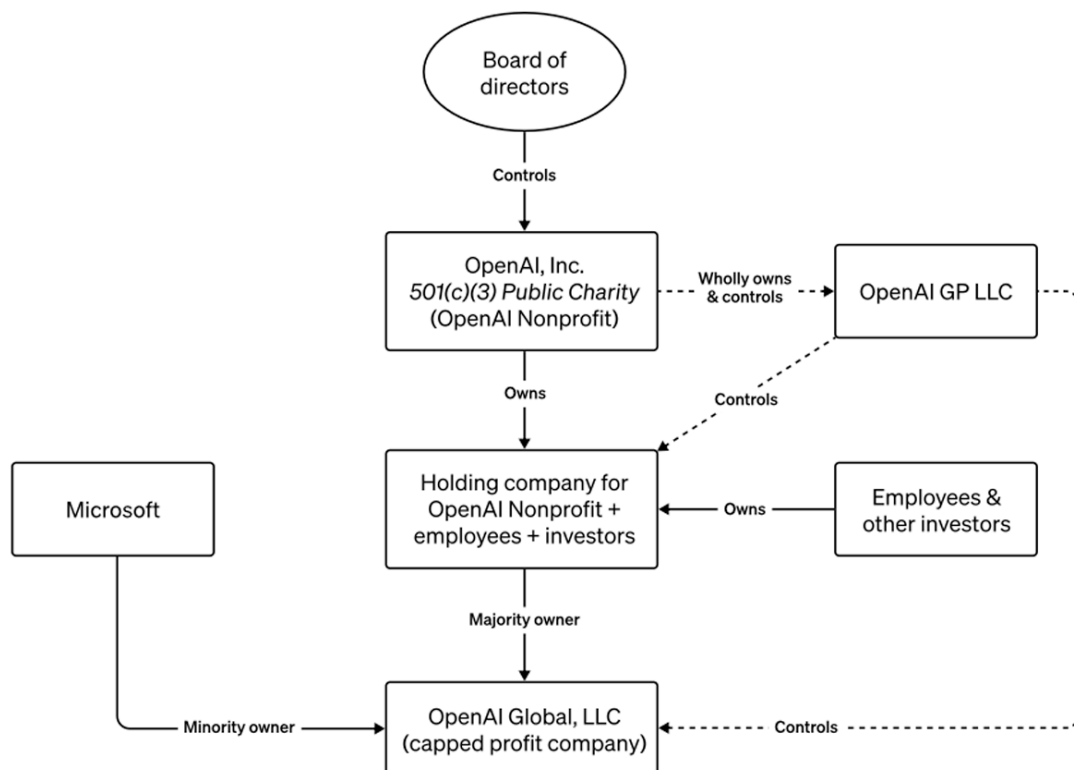
### Cooperative orientation

We will actively cooperate with other research and policy institutions; we seek to create a global community working together to address AGI's global challenges.

We are committed to providing public goods that help society navigate the path to AGI. Today this includes publishing most of our AI research, but we expect that safety and security concerns will reduce our traditional publishing in the future, while increasing the importance of sharing safety, policy, and standards research.

Source: OpenAI Website, "OpenAI Charter," <https://openai.com/charter>, accessed December 2023.

### Exhibit 3 Company Governance Structure



Source: OpenAI website, <https://openai.com/our-structure>, accessed December 2023.

**Exhibit 4** Investing Document

**IMPORTANT**

\*\*Investing in OpenAI Global, LLC is a *high-risk investment*\*\*

\*\*Investors could lose their capital contribution and not see any return\*\*

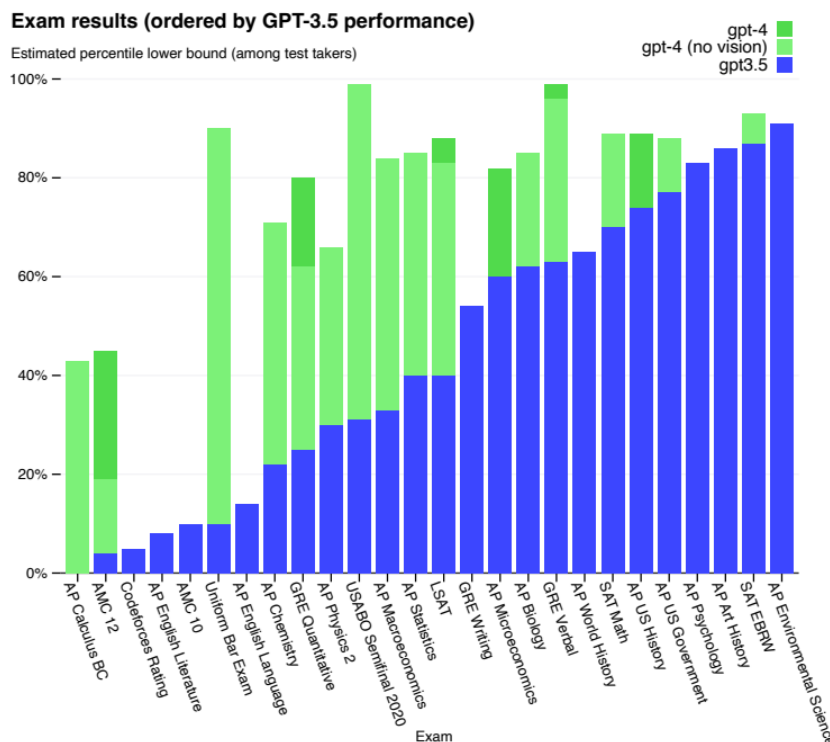
\*\*It would be wise to view any investment in OpenAI Global, LLC in the spirit of a donation, with the understanding that it may be difficult to know what role money will play in a post-AGI world\*\*

The Company exists to advance OpenAI, Inc.'s mission of ensuring that safe artificial general intelligence is developed and benefits all of humanity. The Company's duty to this mission and the principles advanced in the OpenAI, Inc. Charter take precedence over any obligation to generate a profit. The Company may never make a profit, and the Company is under no obligation to do so. The Company is free to re-invest any or all of the Company's cash flow into research and development activities and/or related expenses without any obligation to the Members.

Source: Adapted from OpenAI Website, [https://openai.com/our-structure?utm\\_source=evolvinginternetclub.beehiiv.com&utm\\_medium=referral&utm\\_campaign=brain-food-15-openai-hard-power-vs-soft-power](https://openai.com/our-structure?utm_source=evolvinginternetclub.beehiiv.com&utm_medium=referral&utm_campaign=brain-food-15-openai-hard-power-vs-soft-power); accessed December 2023.

**Exhibit 5** GPT 4 Capabilities

OpenAI released GPT-4, a large language model with an estimated 1.5 trillion parameters in March 2023. It was the most advanced and versatile model that could generate realistic text and images on diverse topics and tasks. GPT-4 demonstrated improved conversational abilities, responsiveness to user steering, potential for image-based inputs, and safety precautions to prevent harmful advice or inappropriate content. GPT-4 also achieved human-level performance on various professional and academic benchmarks. It excelled in a number of exams including the Bar, the LSATs, the SAT's Reading and Math tests, and the GRE. GPT's technical documentation did not include any details on the architecture (including model size), hardware, training compute, data set, or training method. OpenAI had signaled very clearly its shift towards a closed model, proprietary training methods, and the non-disclosure of training data.



Source: OpenAI, GPT-4 technical report, Figure 4, p. 6, <https://cdn.openai.com/papers/gpt-4.pdf>, accessed January 2024.

## Exhibit 6 Risks of AI, according to Sam Altman

Short-Term Risks	
<i>Fake news</i>	"I'm particularly worried that these models could be used for large-scale disinformation. Now that they're getting better at writing computer code, [they] could be used for offensive cyberattacks." (March 2023)
<i>Miseducation</i>	"The more general ability of these models to manipulate, to persuade, to provide sort of one-on-one interactive disinformation. Given that we're going to face an election next year and these models are getting better, I think this is a significant area of concern." (May 2023)
<i>Economic destruction</i>	"The current worries that I have are that there are going to be disinformation problems or economic shocks, or something else at a level far beyond anything we're prepared for. And that doesn't require superintelligence." (March 2023)
<i>Loss of jobs</i>	"I think a lot of customer service jobs, a lot of data entry jobs get eliminated pretty quickly. Some people won't work for sure. I think there are people in the world who don't want to work and get fulfillment in other ways, and that shouldn't be stigmatized either." (June 2023)

Medium-Term Risks	
<i>New diseases</i>	"An A.I. that could design novel biological pathogens. An A.I. that could hack into computer systems. I think these are all scary." (May 2023)
<i>Bad actors</i>	"We do worry a lot about authoritarian governments developing this. A thing that I do worry about is we're not going to be the only creator of this technology. There will be other people who don't put some of the safety limits that we put on it. Society, I think, has a limited amount of time to figure out how to react to that, how to regulate that, how to handle it." (March 2023)
<i>Mental decline</i>	"I worry that as the models get better and better, the users can have less and less of their own discriminating thought process." (May 2023)
Long-Term Risks	
<i>Goodbye humanity</i>	"The bad case – and I think this is important to say – is, like, lights-out for all of us. I'm more worried about an accidental misuse case in the short term." (Jan 2023)
<i>Superintelligence</i>	"A misaligned superintelligent AGI could cause grievous harm to the world; an autocratic regime with a decisive superintelligence lead could do that too." (Feb 2023)
<i>Existential threat of future AI</i>	"We are on an exponential curve, and a relatively steep one, and human intuition for exponential curves is really bad. [...] I think we have to really push ourselves to say, Okay, GPT-4 [is] not a risk like you're talking about there, but how sure are we that GPT-9 won't be? And if it might be, even if there's a small percentage chance of it being really bad, like that deserves attention." (June 2023)
<i>Potential to go 'quite wrong'</i>	"I think if this technology goes wrong, it can go quite wrong." (May 2023)
<i>Sleepless nights</i>	"What I lose the most sleep over is the hypothetical idea that we already have done something really bad by launching ChatGPT." (June 2023)
<i>Biggest fear</i>	"We've got to be careful here. I think people should be happy that we are a little bit scared of this." (May 2023)

Source: Compiled by casewriters from: Jesse Galef, "Sam Altman Investing in 'AI Safety Research,'" *Future of Life*, June 6, 2015, <https://futureoflife.org/ai/sam-altman-investing-in-ai-safety-research/>, accessed January 2024; Spriha Srivastava, *Business Insider*, June 8, 2023 <https://www.businessinsider.com/openai-ceo-sam-altman-says-he-is-losing-sleep-over-chatgpt-2023-6>, accessed January 2024; Rachel Shin, "Sam Altman, the man behind ChatGPT, is increasingly alarmed about what he unleashed. Here are 15 quotes charting his descent into sleepless panic," *Fortune*, June 8, 2023; Justin Hendrix, "Transcript: Senate Judiciary Subcommittee Hearing on Oversight of AI," May 16, 2023, <https://www.techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai/>, accessed January 2024; Mark Sullivan, "Sam Altman: You should not trust Sam Altman," *Fast Company*, June 22, 2023, <https://www.fastcompany.com/90913845/sam-altman-you-should-not-trust-sam-altman>, accessed December 2023.

## Exhibit 7 Timeline of Events

<b>November 16</b>	Sam Altman discusses future of AI at APEC Summit Ilya Sutskever schedules call with Altman
<b>November 17</b>	The board fires Altman Greg Brockman is removed from the board as chair OpenAI makes firing public on its blog CTO Mira Murati becomes the interim CEO Microsoft releases a statement about Altman's firing At all-hands meeting, employees cite "coup" and "hostile takeover" Altman and Brockman post they are "shocked and saddened" Brockman and a string of senior researchers quit Key figures in tech industry express support for Altman Customers consider moving to rival services
<b>November 18</b>	COO sends companywide memo confirming no malfeasance Murati attempts to hire Altman and Brockman Board looks for successor to Altman Board approaches Dario Amodei about potential merger Altman and OpenAI pressure board to quit and give board a deadline (which is not met) Investors support Altman's return
<b>November 19</b>	Altman goes to OpenAI HQ Board negotiations continue OpenAI announces hiring Emmett Shear Employees express strong disapproval of Shear Microsoft announces hiring Altman and Brockman, and invite all OpenAI employees to join
<b>November 20</b>	Altman joins Microsoft to lead new research team

	OpenAI employees sign letter threatening to quit unless board steps down
	Sutskever signs the letter, and posts on X that he “deeply regrets” participating in the board’s actions
	Altman and Brockman consider returning to OpenAI
	Satya Nadella supports their joining either Microsoft or OpenAI
<b>November 21</b>	Altman and OpenAI reach agreement on reinstating Altman as CEO
	Altman gives up his seat on the board
	Summers and Taylor come on as new board members
	D’Angelo is retained from old board
	Toner and McCauley resign
<b>November 29</b>	Microsoft gains non-observer board seat

Source: Compiled by casewriters from Cade Metz and Tripp Mickle, “Back at OpenAI, Sam Altman Outlines The Company’s Priorities,” *The New York Times*, November 29, 2023; Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, “Inside OpenAI’s Crisis Over the Future of Artificial Intelligence,” *The New York Times*, December 9, 2023; Paris Martineau and Julia Black, “From King to Exile to King Again: The Inside Story of Sam Altman’s Whiplash Week,” *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.

#### Exhibit 8 Letter from OpenAI Employees (Excerpt)

To the Board of Directors at OpenAI,

OpenAI is the world’s leading AI company. We, the employees of OpenAI, have developed the best models and pushed the field to new frontiers. Our work on AI safety and governance shapes global norms. The products we built are used by millions of people around the world. Until now, the company we work for and cherish has never been in a stronger position.

The process through which you terminated Sam Altman and removed Greg Brockman from the board has jeopardized all of this work and undermined our mission and company. Your conduct has made it clear you did not have the competence to oversee OpenAI.

When we all unexpectedly learned of your decision, the leadership team of OpenAI acted swiftly to stabilize the company. They carefully listened to your concerns and tried to cooperate with you on all grounds. Despite many requests for specific facts for your allegations, you have never provided any written evidence. They also increasingly realized you were not capable of carrying out your duties, and were negotiating in bad faith.

The leadership team suggested that the most stabilizing path forward - the one that would best serve our mission, company, stakeholders, employees and the public - would be for you to resign and put in place a qualified board that could lead the company forward in stability. Leadership worked with you around the clock to find a mutually agreeable outcome. Yet within two days of your initial

decision, you again replaced interim CEO Mira Murati against the best interests of the company. You also informed the leadership team that allowing the company to be destroyed “would be consistent with the mission.”

**Your actions have made it obvious that you are incapable of overseeing OpenAI. We are unable to work for or with people that lack competence, judgement and care for our mission and employees. We, the undersigned, may choose to resign from OpenAI and join the newly announced Microsoft subsidiary run by Sam Altman and Greg Brockman. Microsoft has assured us that there are positions for all OpenAI employees at this new subsidiary should we choose to join. We will take this step imminently, unless all current board members resign, and the board appoints two new lead independent directors, such as Bret Taylor and Will Hurd, and reinstates Sam Altman and Greg Brockman.**

Source: The New York Times, “The Employee Letter to OpenAI’s Board,” November 20, 2023, <https://www.nytimes.com/interactive/2023/11/20/technology/letter-to-the-open-ai-board.html>, accessed January 2024.

## Exhibit 9 An Evolving Board

### The Old Guard (Six including Sam Altman)

#### *Greg Brockman*

An OpenAI co-founder, President and Chair of the board, Brockman resigned from the company in protest of Altman's ousting. Before his involvement with OpenAI, Brockman spent five years as the CTO of Stripe. *He returned as President after Altman was reinstated as CEO.*

#### *Ilya Sutskever*

Sutskever was Geoffrey Hinton's star student and helped initiate the deep-learning revolution. After co-founding DNNResearch, an AI startup specializing in neural networks, he sold the company to Google and then joined Google as a research scientist. After three years, he left to become a co-founder and research director at OpenAI. Since November 2018, he served as the chief scientist. *His future role at OpenAI was unclear.*

#### *Adam D'Angelo*

D'Angelo, the co-founder and CEO of Quora, a social platform for questions and answers, spent close to four years at Facebook, serving as the tech giant's CTO from 2006 to 2008. *He was retained as a board member after Altman was reinstated.*

#### *Helen Toner*

A University of Melbourne graduate with a degree in chemical engineering, Toner worked as a research analyst at firms including Open Philanthropy. She was also involved with the University of Oxford's Center for the Governance of AI. In 2019, she studied Beijing's AI ecosystem and then helped establish Georgetown University's Center for Security and Emerging Technology, serving as director of strategy. Toner joined the OpenAI board in 2021, succeeding her former manager Holden Karnofsky, whose wife co-founded OpenAI rival Anthropic. Toner was known in the AI safety world for being a critical thinker unafraid to challenge the status quo. *She stepped down as a board member.*

#### *Tasha McCauley*

McCauley was the CEO of GeoSim Systems, a geospatial tech company. She was an adjunct senior management scientist at Rand Corporation and on the OpenAI board since 2018. McCauley had worked on AI policy and governance issues and taught at Singularity University. *She resigned from the board.*

### The New Members (Three including Adam D' Angelo)

#### *Larry Summers*

Larry Summers, a prominent economist, held key positions including U.S. Secretary of the Treasury under Bill Clinton and President of the National Economic Council under Barack Obama. He also served as the president of Harvard University and Chief Economist at the World Bank.

#### *Bret Taylor*

Bret Taylor served as co-CEO of Salesforce and played a key role in its acquisition of Slack. He also chaired Twitter until 2022. Taylor replaced Greg Brockman as OpenAI chair of the board.



Source: Compiled by casewriters from Meghan Bobrowsky and Deepa Seetharaman, "The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side," *The Wall Street Journal*, December 7, 2023; Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023; Cade Metz and Tripp Mickle, "Back at OpenAI, Sam Altman Outlines The Company's Priorities," *The New York Times*, November 29, 2023.

## Endnotes

- <sup>1</sup> Mark Sullivan, "Sam Altman: You should not trust Sam Altman," *Fast Company*, June 22, 2023, <https://www.fastcompany.com/90913845/sam-altman-you-should-not-trust-sam-altman>, accessed December 2023.
- <sup>2</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>3</sup> Cade Metz, Karen Weise, Nico Grant, and Mike Isaac, "Ego, Fear and Money: How the A.I. Fuse Was Lit," *The New York Times*, <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>, accessed December 3, 2023, accessed December 2023.
- <sup>4</sup> Melissa Heikkila and Will Douglas Heaven, "What's next for OpenAI," *MIT Technology Review*, November 20, 2023.
- <sup>5</sup> Jon Porter, "ChatGPT continues to be one of the fastest-growing services ever," *The Verge*, November 6, 2023, <https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference>, accessed November 2023.
- <sup>6</sup> Stephanie Palazzolo, Anissa Gardizy, Kate Clark and Job Victor, "What Comes Next for Sam Altman's OpenAI," *The Information*, November 22, 2023, [https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm\\_campaign=RTSU+-+OpenAI+What%27s&utm\\_content=2308&utm\\_medium=email&utm\\_source=cio&utm\\_term=1675&rc=kimdxv](https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm_campaign=RTSU+-+OpenAI+What%27s&utm_content=2308&utm_medium=email&utm_source=cio&utm_term=1675&rc=kimdxv), accessed November 2023.
- <sup>7</sup> Maria Heeter and Kate Clark, "Thrive-Led OpenAI Tender to Continue After Altman Returns," *The Information*, November 22, 2023, [https://www.theinformation.com/articles/thrive-led-openai-tender-to-continue-after-altman-returns?offer=rtssu-engagement-sept-23&utm\\_campaign=RTSU+-+Thrive+Tender&utm\\_content=2312&utm\\_medium=email&utm\\_source=cio&utm\\_term=1679&rc=kimdxv](https://www.theinformation.com/articles/thrive-led-openai-tender-to-continue-after-altman-returns?offer=rtssu-engagement-sept-23&utm_campaign=RTSU+-+Thrive+Tender&utm_content=2312&utm_medium=email&utm_source=cio&utm_term=1679&rc=kimdxv), accessed November 2023.
- <sup>8</sup> OpenAI Website, <https://openai.com/about>, accessed December 2023.
- <sup>9</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>10</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>11</sup> Anna Tong, Jeffrey Dastin and Krystal Hu, "OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, sources say," *Reuters*, November 23, 2023, <https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/>, accessed December 2023.
- <sup>12</sup> Paris Martineau and Julia Black, "From King to Exile to King Again: The Inside Story of Sam Altman's Whiplash Week," *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.
- <sup>13</sup> Ross Anderson, "OpenAI's Chief Scientist Made a Tragic Miscalculation," *The Atlantic*, November 2023.
- <sup>14</sup> Paris Martineau and Julia Black, "From King to Exile to King Again: The Inside Story of Sam Altman's Whiplash Week," *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.
- <sup>15</sup> Stephanie Palazzolo, Anissa Gardizy, Kate Clark and Job Victor, "What Comes Next for Sam Altman's OpenAI," *The Information*, November 22, 2023, [https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm\\_campaign=RTSU+-+OpenAI+What%27s&utm\\_content=2308&utm\\_medium=email&utm\\_source=cio&utm\\_term=1675&rc=kimdxv](https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm_campaign=RTSU+-+OpenAI+What%27s&utm_content=2308&utm_medium=email&utm_source=cio&utm_term=1675&rc=kimdxv), accessed November 2023.
- <sup>16</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>17</sup> AIWs, "This week in the History of AI," December 23, 2022, <https://aiws.net/the-history-of-ai/this-week-in-the-history-of-ai-at-aiws-net-in-three-to-eight-years-we-will-have-a-machine-with-the-general-intelligence-of-an-average-human-being/>, accessed January 2024.
- <sup>18</sup> DeepMind/Google, "AlphaGo", <https://deepmind.google/technologies/alphago/>, accessed December 2023.

- <sup>19</sup> Visual Storytelling Team and Madhumita Murgia, "Generative AI exists because of the transformer," *The Financial Times*, September 12, 2023, <https://ig.ft.com/generative-ai/>, accessed January 2024.
- <sup>20</sup> Cade Metz, Karen Weise, Nico Grant, and Mike Isaac, "Ego, Fear and Money: How the A.I. Fuse Was Lit," *The New York Times*, <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>, accessed December 3, 2023, accessed December 2023.
- <sup>21</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>22</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>23</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>24</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>25</sup> Ross Anderson, "OpenAI's Chief Scientist Made a Tragic Miscalculation," *The Atlantic*, November, 2023.
- <sup>26</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>27</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>28</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>29</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>30</sup> OpenAI Website, "OpenAI Charter," <https://openai.com/charter>, accessed December 2023.
- <sup>31</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>32</sup> Melissa Heikkila and Will Douglas Heaven, "What's next for OpenAI," *MIT Technology Review*, November 20, 2023.
- <sup>33</sup> Kevin Roose, "This A.I. Subculture's Motto: Go, Go, Go," *The New York Times*, December 10, 2023, [https://www.nytimes.com/2023/12/10/technology/ai-acceleration.html#:~:text=Effective%20Accelerationism%20\(often%20shortened%20to,barred%20pursuit%20of%20technologic al%20progress.,](https://www.nytimes.com/2023/12/10/technology/ai-acceleration.html#:~:text=Effective%20Accelerationism%20(often%20shortened%20to,barred%20pursuit%20of%20technologic al%20progress.,) accessed December 2023.
- <sup>34</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>35</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>36</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>37</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>38</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>39</sup> Cade Metz, Karen Weise, Nico Grant, and Mike Isaac, "Ego, Fear and Money: How the A.I. Fuse Was Lit," *The New York Times*, <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>, accessed December 3, 2023, accessed December 2023.
- <sup>40</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.

- <sup>41</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>42</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>43</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>44</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>45</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>46</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>47</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>48</sup> Elon Musk, Post on X (Twitter), September 24, 2020, <https://twitter.com/elonmusk/status/1309052632850468864>, accessed December 2023.
- <sup>49</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>50</sup> Cade Metz, Karen Weise, Nico Grant, and Mike Isaac, "Ego, Fear and Money: How the A.I. Fuse Was Lit," *The New York Times*, <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html>, accessed December 3, 2023, accessed December 2023.
- <sup>51</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>52</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>53</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>54</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>55</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>56</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>57</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>58</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>59</sup> Stephanie Palazzolo, Anissa Gardizy, Kate Clark and Job Victor, "What Comes Next for Sam Altman's OpenAI," *The Information*, November 22, 2023, [https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm\\_campaign=RTSU+-+OpenAI+What%27s&utm\\_content=2308&utm\\_medium=email&utm\\_source=cio&utm\\_term=1675&rc=kimdxv](https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm_campaign=RTSU+-+OpenAI+What%27s&utm_content=2308&utm_medium=email&utm_source=cio&utm_term=1675&rc=kimdxv), accessed November 2023.
- <sup>60</sup> Nilay Patel, "Microsoft thinks AI can beat Google at search – CEO Satya Nadella explains why," *The Verge*, February 7, 2023, <https://www.theverge.com/23589994/microsoft-ceo-satya-nadella-bing-chatgpt-google-search-ai>, accessed October December 2023.
- <sup>61</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>62</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.

- <sup>63</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>64</sup> Stephanie Palazzolo, Anissa Gardizy, Kate Clark and Job Victor, "What Comes Next for Sam Altman's OpenAI," *The Information*, November 22, 2023, [https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm\\_campaign=RTSU+-+OpenAI+What%27s&utm\\_content=2308&utm\\_medium=email&utm\\_source=cio&utm\\_term=1675&rc=kimdxv](https://www.theinformation.com/articles/what-comes-next-for-sam-altmans-openai?offer=rtssu-engagement-sept-23&utm_campaign=RTSU+-+OpenAI+What%27s&utm_content=2308&utm_medium=email&utm_source=cio&utm_term=1675&rc=kimdxv), accessed November 2023.
- <sup>65</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>66</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>67</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>68</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>69</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>70</sup> Melissa Heikkila and Will Douglas Heaven, "What's next for OpenAI," *MIT Technology Review*, November 20, 2023.
- <sup>71</sup> Stephanie Palazzolo, "Why OpenAI's Breakthrough Isn't a Harbinger for Humanity-Ending AGI," *The Information*, November 27, 2023.
- <sup>72</sup> Anna Tong, Jeffrey Dastin and Krystal Hu, "OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, sources say," *Reuters*, November 23, 2023, <https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/>, accessed December 2023.
- <sup>73</sup> John Thornhill, "OpenAI has just fused its corporate 'kill switch,'" *Financial Times*, November 23, 2023.
- <sup>74</sup> Eliezer Yudkowsky, "Pausing AI developments Isn't Enough. We Need to Shut it All Down," *TIME* magazine, March 29, 2023, <https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>, accessed January 2024.
- <sup>75</sup> Center for AI Safety, "Statement on AI Risk," <https://www.safe.ai/statement-on-ai-risk>, accessed January 2024.
- <sup>76</sup> Dario Amodi, "Oversight of A.I.: Principles of Regulation," Written Testimony Before the Judiciary Committee, United States Senate, July 25, 2023.
- <sup>77</sup> Rachel Shin, "Sam Altman, the man behind ChatGPT, is increasingly alarmed about what he unleashed. Here are 15 quotes charting his descent into sleepless panic," *Fortune*, June 8, 2023, accessed December 2023.
- <sup>78</sup> Steven Levy, "What OpenAI Really Wants," *Wired*, September 5, 2023, <https://www.wired.com/story/what-openai-really-wants/>, accessed November 2023.
- <sup>79</sup> Lingjiao Chen, Matei Zaharia, James Zou, "How is ChatGPT's Behavior Changing over Time?" Stanford University and UC Berkeley, October 31, 2023, <https://arxiv.org/pdf/2307.09009.pdf>, accessed December 2023.
- <sup>80</sup> Hayden Field, "Who's on the OpenAI board - the group behind Sam Altman's ouster," *CNBC*, November 18, 2023, [Who's on the OpenAI board - the group behind Sam Altman's ouster \(cnn.com\)](https://www.cnn.com/2023/11/18/tech/openai-board/index.html), accessed January 2024.
- <sup>81</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>82</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>83</sup> Meghan Bobrowsky and Deepa Seetharaman, "The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side," *The Wall Street Journal*, December 7, 2023.
- <sup>84</sup> Cade Metz, Tripp Mickle, and Mike Isaac, "Before Altman's Ouster, OpenAI's Board was Divided and Feuding," November 21, 2023, <https://www.nytimes.com/2023/11/21/technology/openai-altman-board-fight.html#:~:text=A.I.%20Race%20Began,Before%20Altman's%20Ouster%2C%20OpenAI's%20Board%20Was%20Divided%20and%20Feuding,who%20should%20fill%20board%20vacancies,> accessed November 2023.

- <sup>85</sup> Meghan Bobrowsky and Deepa Seetharaman, "The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side," *The Wall Street Journal*, December 7, 2023.
- <sup>86</sup> Karen Hao and Charlie Warzel, "Inside the Chaos at OpenAI," *The Atlantic*, November 19, 2023.
- <sup>87</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>88</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>89</sup> John Thornhill, "OpenAI has just fused its corporate 'kill switch,'" *Financial Times*, November 23, 2023.
- <sup>90</sup> Paris Martineau and Julia Black, "From King to Exile to King Again: The Inside Story of Sam Altman's Whiplash Week," *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.
- <sup>91</sup> Justin Hendrix, "Transcript: Senate Judiciary Subcommittee Hearing on Oversight of AI," May 16, 2023, <https://www.techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai/>, accessed January 2024
- <sup>92</sup> Paris Martineau and Julia Black, "From King to Exile to King Again: The Inside Story of Sam Altman's Whiplash Week," *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.
- <sup>93</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>94</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>95</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>96</sup> Richard Lawler, "This is how OpenAI fired its CEO," *The Verge*, November 18, 2023, <https://www.theverge.com/2023/11/18/23966509/this-is-how-openai-fired-its-ceo-sam-altman>, accessed November 2023.
- <sup>97</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>98</sup> Meghan Bobrowsky and Deepa Seetharaman, "The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side," *The Wall Street Journal*, December 7, 2023.
- <sup>99</sup> John Thornhill, "OpenAI has just fused its corporate 'kill switch,'" *Financial Times*, November 23, 2023.
- <sup>100</sup> Meghan Bobrowsky and Deepa Seetharaman, "The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side," *The Wall Street Journal*, December 7, 2023.
- <sup>101</sup> Cade Metz, Tripp Mickle, and Mike Isaac, "Before Altman's Ouster, OpenAI's Board was Divided and Feuding," November 21, 2023, <https://www.nytimes.com/2023/11/21/technology/openai-altman-board-fight.html#:~:text=A.I.%20Race%20Began,Before%20Altman's%20Ouster%2C%20OpenAI's%20Board%20Was%20Divided%20and%20Feuding,who%20should%20fill%20board%20vacancies.>, accessed November 2023.
- <sup>102</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>103</sup> Cade Metz, Tripp Mickle, and Mike Isaac, "Before Altman's Ouster, OpenAI's Board was Divided and Feuding," November 21, 2023, <https://www.nytimes.com/2023/11/21/technology/openai-altman-board-fight.html#:~:text=A.I.%20Race%20Began,Before%20Altman's%20Ouster%2C%20OpenAI's%20Board%20Was%20Divided%20and%20Feuding,who%20should%20fill%20board%20vacancies.>, accessed November 2023.

- <sup>104</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>105</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>106</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>107</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>108</sup> Stephanie Palazzolo, "OpenAI's Board Approached Anthropic About Merger," *The Information*, November 20, 2023, [https://www.theinformation.com/articles/openai-approached-anthropic-about-merger?utm\\_campaign=Automated+Fallback+R&utm\\_content=89&utm\\_medium=email&utm\\_source=cio&utm\\_term=19&rc=kimdxv](https://www.theinformation.com/articles/openai-approached-anthropic-about-merger?utm_campaign=Automated+Fallback+R&utm_content=89&utm_medium=email&utm_source=cio&utm_term=19&rc=kimdxv), accessed November 2023.
- <sup>109</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>110</sup> Melissa Heikkila and Will Douglas Heaven, "What's next for OpenAI," *MIT Technology Review*, November 20, 2023.
- <sup>111</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>112</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>113</sup> Paris Martineau and Julia Black, "From King to Exile to King Again: The Inside Story of Sam Altman's Whiplash Week," *The Information*, November 22, 2023, <https://www.theinformation.com/articles/from-king-to-exile-to-king-again-the-inside-story-of-sam-altmans-whiplash-week?rc=kimdxv>, accessed January 2024.
- <sup>114</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>115</sup> Tripp Mickle, Cade Metz, Mike Isaac, and Karen Weise, "Inside OpenAI's Crisis Over the Future of Artificial Intelligence," *The New York Times*, December 9, 2023.
- <sup>116</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>117</sup> John Thornhill, "OpenAI has just fused its corporate 'kill switch,'" *Financial Times*, November 23, 2023.
- <sup>118</sup> Nitasha Tiku, "OpenAI leaders warned of abusive behavior before Sam Altman's ouster," *The Washington Post*, December 8, 2023.
- <sup>119</sup> Cade Metz and Tripp Mickle, "Back at OpenAI, Sam Altman Outlines The Company's Priorities," *The New York Times*, November 29, 2023.
- <sup>120</sup> Derek Thompson, "The OpenAI Mess Is About One Big Thing," *The Atlantic*, November 22, 2023.
- <sup>121</sup> Will Douglas Heaven, "Now we know what OpenAI's superalignment team has been up to," *MIT Technology Review*, December 14, 2023.
- <sup>122</sup> Berber Jin and Keach Hagey, "The Contradictions of Sam Altman, AI Crusader," *The Wall Street Journal*, March 31, 2023, <https://www.wsj.com/tech/ai/chatgpt-sam-altman-artificial-intelligence-openai-b0e1c8c9>, accessed January 2024.