

TIME

The Control Key

SEP 29, 2023 · 10 MINUTES

 [Save for Later](#)



At a conference in 2012, Elon Musk met Demis Hassabis, the video-game designer and artificial-intelligence researcher who had co-founded a company named

DeepMind that sought to design computers that could learn how to think like humans.

“Elon and I hit it off right away, and I went to visit him at his rocket factory,” Hassabis says. While sitting in the canteen overlooking the assembly lines, Musk explained that his reason for building rockets that could go to Mars was that it might be a way to preserve human consciousness in the event of a world war, asteroid strike, or civilization collapse. Hassabis told him to add another potential threat to the list: artificial intelligence. Machines could become superintelligent and surpass us mere mortals, perhaps even decide to dispose of us.





SCRIBD



EN ▾

Upload

Saved



Musk paused silently for almost a minute as he processed this possibility. He

decided that Hassabis might be right about the danger of AI, and promptly invested \$5 million in DeepMind as a way to monitor what it was doing.

A few weeks after this conversation with Hassabis, Musk described DeepMind to Google's Larry Page. They had known each other for more than a decade, and Musk often stayed at Page's Palo Alto, Calif., house. The potential dangers of artificial intelligence became a topic that Musk would raise, almost obsessively, during their late-night conversations. Page was dismissive.

At Musk's 2013 birthday party in Napa Valley, California, they got into a passionate debate. Unless we built in safeguards, Musk argued, artificial-intelligence systems might replace humans, making our species irrelevant or even extinct.

Page pushed back. Why would it matter, he asked, if machines someday surpassed humans in intelligence, even consciousness? It would simply be the next stage of evolution.



Human consciousness, Musk retorted, was a precious flicker of light in the universe, and we should not let it be extinguished. Page considered that sentimental nonsense. If consciousness could be replicated in a machine, why would that not be just as valuable? He accused Musk of being a “specist,” someone who was biased in favor of their own species. “Well, yes, I am pro-human,” Musk responded. “I f-cking like humanity, dude.”

Musk was therefore dismayed when he heard at the end of 2013 that Page and Google were planning to buy DeepMind. Musk and his friend Luke Nosek tried to put together financing to stop the deal. At a party in Los Angeles, they went to an upstairs closet for an hour-long Skype call with Hassabis. “The future of AI should

not be controlled by Larry,” Musk told him.

The effort failed, and Google’s acquisition of DeepMind was announced in January 2014. Page initially agreed to create a “safety council,” with Musk as a member. The first and only meeting was held at SpaceX. Page, Hassabis, and Google chair Eric Schmidt attended, along with Reid Hoffman and a few others. Musk concluded that the council was basically bullsh-t.

So Musk began hosting his own series of dinner discussions on ways to counter Google and promote AI safety. He even reached out to President Obama, who agreed to a one-on-one meeting in May 2015. Musk explained the risk and suggested that it be regulated. “Obama got it,” Musk says. “But I realized that it was not going to rise to the level of something that he would do anything about.”

Musk then turned to Sam Altman, a tightly bundled software entrepreneur, sports-car enthusiast, and survivalist who, behind his polished veneer, had a Musk-like intensity. At a small dinner in Palo Alto, they decided to co-found a nonprofit artificial-intelligence-research lab, which they named OpenAI. It would make its software open-source and try to counter Google’s growing dominance of the field. “We wanted to have something like a Linux version of AI that was not controlled by any one person or corporation,” Musk says.

One question they discussed at dinner was what would be safer: a small number of AI systems that were controlled by big corporations or a large number of independent systems? They concluded that a large number of competing systems, providing checks and balances on one another, was better. For Musk, this was the reason to make OpenAI truly open, so that lots of people could build systems based on its source code.

Another way to assure AI safety, Musk felt, was to tie the bots closely to humans.

They should be an extension of the will of individuals, rather than systems that could go rogue and develop their own goals and intentions. That would become one of the rationales for Neuralink, the company he would found to create chips that could connect human brains directly to computers.

Musk's determination to develop artificial-intelligence capabilities at his own companies caused a break with OpenAI in 2018. He tried to convince Altman that OpenAI should be folded into Tesla. The OpenAI team rejected that idea, and Altman stepped in as president of the lab, starting a for-profit arm that was able to raise equity funding, including a major investment from Microsoft.

So Musk decided to forge ahead with building rival AI teams to work on an array of related projects. These included Neuralink, which aims to plant microchips in human brains; Optimus, a humanlike robot; and Dojo, a supercomputer that can use millions of videos to train an artificial neural network to simulate a human brain. It also spurred him to become obsessed with pushing to make Tesla cars self-driving.

At first these endeavors were rather independent, but eventually Musk would tie them all together, along with a new company he founded called xAI, to pursue the goal of artificial general intelligence.

IN MARCH 2023, OpenAI released GPT-4 to the public. Google then released a rival chatbot named Bard. The stage was thus set for a competition between OpenAI-Microsoft and DeepMind-Google to create products that could chat with humans in a natural way and perform an endless array of text-based intellectual tasks.

Musk worried that these chatbots and AI systems, especially in the hands of Microsoft and Google, could become politically indoctrinated, perhaps even infected by what he called the woke-mind virus. He also feared that self-learning

AI systems might turn hostile to the human species. And on a more immediate level, he worried that chatbots could be trained to flood Twitter with disinformation, biased reporting, and financial scams. All of those things were already being done by humans, of course. But the ability to deploy thousands of weaponized chatbots would make the problem two or three orders of magnitude worse.

His compulsion to ride to the rescue kicked in. He was resentful that he had founded and funded OpenAI but was now left out of the fray. AI was the biggest storm brewing. And there was no one more attracted to a storm than Musk.

In February 2023, he invited—perhaps a better word is *summoned*—Sam Altman to meet with him at Twitter and asked him to bring the founding documents for OpenAI. Musk challenged him to justify how he could legally transform a nonprofit funded by donations into a for-profit that could make millions. Altman tried to show that it was all legitimate, and he insisted that he personally was not a shareholder or cashing in. He also offered Musk shares in the new company, which Musk declined.

Instead, Musk unleashed a barrage of attacks on OpenAI. Altman was pained. Unlike Musk, he is sensitive and nonconfrontational. He felt that Musk had not drilled down enough into the complexity of the issue of AI safety. However, he did feel that Musk's criticisms came from a sincere concern. "He's a jerk," Altman told Kara Swisher. "He has a style that is not a style that I'd want to have for myself. But I think he does really care, and he is feeling very stressed about what the future's going to look like for humanity."

THE FUEL FOR AI is data. The new chatbots were being trained on massive amounts of information, such as billions of pages on the internet and other documents. Google and Microsoft, with their search engines and cloud services

and access to emails, had huge gushers of data to help train these systems.

What could Musk bring to the party? One asset was the Twitter feed, which included more than a trillion tweets posted over the years, 500 million added each day. It was humanity's hive mind, the world's most timely dataset of real-life human conversations, news, interests, trends, arguments, and lingo. Plus it was a great training ground for a chatbot to test how real humans react to its responses. The value of this data feed was not something Musk considered when buying Twitter, which he has since renamed X. "It was a side benefit, actually, that I realized only after the purchase," he says.

Twitter had rather loosely permitted other companies to make use of this data stream. In January 2023, Musk convened a series of late-night meetings in his Twitter conference room to work out ways to charge for it. "It's a monetization opportunity," he told the engineers. It was also a way to restrict Google and Microsoft from using this data to improve their AI chatbots. He ignited a controversy in July when he decided to temporarily restrict the number of tweets a viewer could see per day; the goal was to prevent Google and Microsoft from "scraping" up millions of tweets to use as data to train their AI systems.

There was another data trove that Musk had: the 160 billion frames per day of video that Tesla received and processed from the cameras on its cars. This data was different from the text-based documents that informed chatbots. It was video data of humans navigating in real-world situations. It could help create AI for physical robots, not just text-generating chatbots.

The holy grail of artificial general intelligence is building machines that can operate like humans in physical spaces, such as factories and offices and on the surface of Mars, not just wow us with disembodied chatting. Tesla and Twitter together could provide the datasets and the processing capability for both

approaches: teaching machines to navigate in physical space and to answer questions in natural language.

AI was the biggest storm brewing. And there was no one more attracted to a storm than Musk



THIS PAST MARCH, Musk texted me, “There are a few important things I would like to talk to you about. Can only be done in person.” When I got to Austin, he was at the house of Shivon Zilis, the Neuralink executive who is the mother of two of his children and who had been his intellectual companion on artificial intelligence since the founding of OpenAI eight years earlier. He said we should leave our phones in the house while we sat outside, because, he said, someone could use

them to monitor our conversation. But he later agreed that I could use what he said about AI in my book.

He and Zilis sat cross-legged and barefoot on the poolside patio with their twins, Strider and Azure, now 16 months old, on their laps. Zilis made coffee and then put his in the microwave to get it superhot so he wouldn't chug it too fast.

"What can be done to make AI safe?" Musk asked. "I keep wrestling with that. What actions can we take to minimize AI danger and assure that human consciousness survives?"

He spoke in a low monotone punctuated by bouts of almost manic laughter. The amount of human intelligence, he noted, was leveling off, because people were not having enough children. Meanwhile, the amount of computer intelligence was going up exponentially, like Moore's Law on steroids. At some point, biological brainpower would be dwarfed by digital brainpower.

In addition, new AI machine-learning systems could ingest information on their own and teach themselves how to generate outputs, even upgrade their own code and capabilities. The term *singularity* was used by the mathematician John von Neumann and the sci-fi writer Vernor Vinge to describe the moment when artificial intelligence could forge ahead on its own at an uncontrollable pace and leave us mere humans behind. "That could happen sooner than we expected," Musk said in an ominous tone.

For a moment I was struck by the oddness of the scene. We were sitting on a suburban patio by a tranquil backyard swimming pool on a sunny spring day, with two bright-eyed twins learning to toddle, as Musk somberly speculated about the window of opportunity for building a sustainable human colony on Mars before an AI apocalypse destroyed earthly civilization.

Musk lapsed into one of his long silences. He was, as Zilis called it, “batch processing,” referring to the way an old-fashioned computer would cue up a number of tasks and run them sequentially when it had enough processing power available. “I can’t just sit around and do nothing,” he finally said softly. “With AI coming, I’m sort of wondering whether it’s worth spending that much time thinking about Twitter. Sure, I could probably make it the biggest financial institution in the world. But I have only so many brain cycles and hours in the day. I mean, it’s not like I need to be richer or something.”

I started to speak, but he knew what I was going to ask. “So what should my time be spent on?” he said. “Getting Starship launched. Getting to Mars is now far more pressing.” He paused again, then added, “Also, I need to focus on making AI safe. That’s why I’m starting an AI company.”

This is the company Musk dubbed xAI. He personally recruited Igor Babuschkin, formerly of DeepMind, but he told me he would run it himself. I calculated that would mean he would be running six companies: Tesla, SpaceX and its Starlink unit, Twitter, the Boring Co., Neuralink, and xAI. That was three times as many as Steve Jobs (Apple, Pixar) at his peak.

He admitted that he was starting off way behind OpenAI in creating a chatbot that could give natural-language responses to questions. But Tesla’s work on self-driving cars and Optimus the robot put it way ahead in creating the type of AI needed to navigate in the physical world. This meant that his engineers were actually ahead of OpenAI in creating full-fledged artificial general intelligence, which requires both abilities. “Tesla’s real-world AI is underrated,” he said. “Imagine if Tesla and OpenAI had to swap tasks. They would have to make self-driving, and we would have to make large-language-model chatbots. Who wins? We do.”

In April, Musk assigned Babuschkin and his team three major goals. The first was to make an AI bot that could write computer code. A programmer could begin typing in any coding language, and the xAI bot would auto-complete the task for the most likely action they were trying to take. The second product would be a chatbot competitor to OpenAI's GPT series, one that used algorithms and trained on datasets that would ensure its political neutrality.

The third goal that Musk gave the team was even grander. His overriding mission had always been to assure that AI developed in a way that helped guarantee that human consciousness endured. That was best achieved, he thought, by creating a form of artificial general intelligence that could "reason" and "think" and pursue "truth" as its guiding principle. You should be able to give it big tasks, like "Build a better rocket engine."

Someday, Musk hoped, it would be able to take on even grander and more existential questions. It would be "a maximum truth-seeking AI. It would care about understanding the universe, and that would probably lead it to want to preserve humanity, because we are an interesting part of the universe." That sounded vaguely familiar, and then I realized why.

He was embarking on a mission similar to the one chronicled in the formative (perhaps too formative?) bible of his childhood years, the one that pulled him out of his adolescent existential depression, *The Hitchhiker's Guide to the Galaxy*, which featured a supercomputer designed to figure out "the Answer to The Ultimate Question of Life, the Universe, and Everything."

Isaacson, a former editor of TIME, is a professor of history at Tulane and the author of numerous acclaimed biographies. Copyright 2023. Adapted from the book Elon Musk by Walter Isaacson, published by Simon & Schuster Inc. Printed by permission