The New York Times

Mark Zuckerberg, Elon Musk and the Feud Over Killer Robots

As the tech moguls disagree over the risks presented by something that doesn't exist yet, all of Silicon Valley is learning about unintended consequences of A.I.

By Cade Metz

June 9, 2018

SAN FRANCISCO — Mark Zuckerberg thought his fellow Silicon Valley billionaire Elon Musk was behaving like an alarmist.

Mr. Musk, the entrepreneur behind SpaceX and the electric-car maker Tesla, had taken it upon himself to warn the world that artificial intelligence was "potentially more dangerous than nukes" in television interviews and on social media.

So, on Nov. 19, 2014, Mr. Zuckerberg, Facebook's chief executive, invited Mr. Musk to dinner at his home in Palo Alto, Calif. Two top researchers from Facebook's new artificial intelligence lab and two other Facebook executives joined them.

As they ate, the Facebook contingent tried to convince Mr. Musk that he was wrong. But he wasn't budging. "I genuinely believe this is dangerous," Mr. Musk told the table, according to one of the dinner's attendees, Yann LeCun, the researcher who led Facebook's A.I. lab.

Mr. Musk's fears of A.I., distilled to their essence, were simple: If we create machines that are smarter than humans, they could turn against us. (See: "The Terminator," "The Matrix," and "2001: A Space Odyssey.") Let's for once, he was saying to the rest of the tech industry, consider the unintended consequences of what we are creating before we unleash it on the world.

Neither Mr. Musk nor Mr. Zuckerberg would talk in detail about the dinner, which has not been reported before, or their long-running A.I. debate.

The creation of "superintelligence" — the name for the supersmart technological breakthrough that takes A.I. to the next level and creates machines that not only perform narrow tasks that typically require human intelligence (like self-driving cars) but can actually outthink humans — still feels like science fiction. But the fight over the future of A.I. has spread across the tech industry.

More than 4,000 Google employees recently signed a petition protesting a $9 million A.I. contract the company had signed with the Pentagon — a deal worth chicken feed to the internet giant, but deeply troubling to many artificial intelligence researchers at the company. Last week, Google executives, trying to head off a worker rebellion, said they wouldn't renew the contract when it expires next year.

Artificial intelligence research has enormous potential and enormous implications, both as an economic engine and a source of military superiority. The Chinese government has said it is willing to spend billions in the coming years to make the country the world's leader in A.I., while the Pentagon is aggressively courting the tech industry for help. A new breed of autonomous weapons can't be far away.

All sorts of deep thinkers have joined the debate, from a gathering of philosophers and scientists held along the central California coast to an annual conference hosted in Palm Springs, Calif., by Amazon's chief executive, Jeff Bezos.

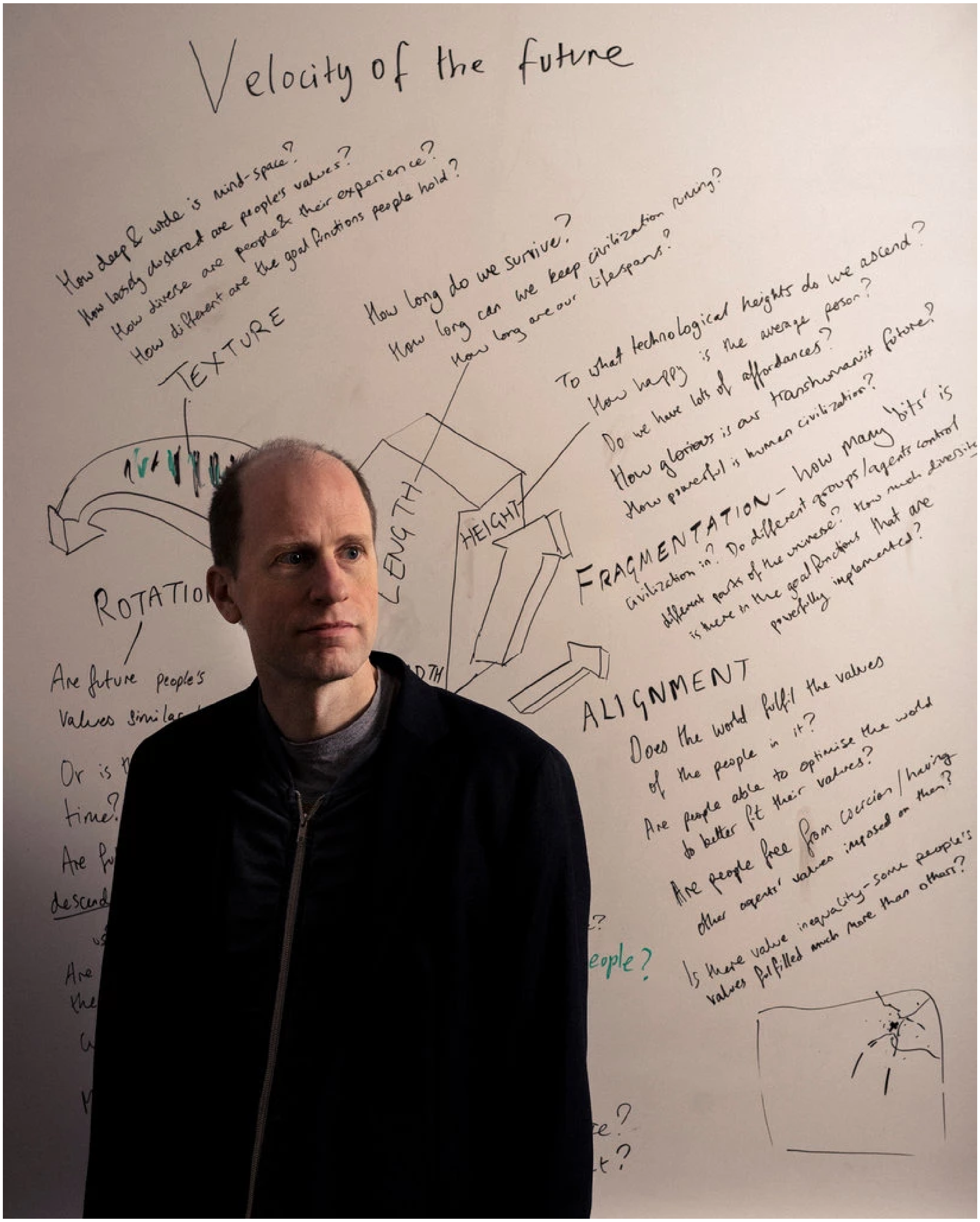
"You can now talk about the risks of A.I. without seeming like you are lost in science fiction," said Allan Dafoe, a director of the governance of A.I. program at the Future of Humanity Institute, a research center at the University of Oxford that explores the risks and opportunities of advanced technology.

And the public roasting of Facebook and other tech companies over the past few months has done plenty to raise the issue of the unintended consequences of the technology created by Silicon Valley.

In April, Mr. Zuckerberg spent two days answering questions from members of Congress about data privacy and Facebook's role in the spread of misinformation before the 2016 election. He faced a similar grilling in Europe last month.

Facebook's recognition that it was slow to understand what was going on has led to a rare moment of self-reflection in an industry that has long believed it is making the world a better place, whether the world likes it or not.

Even such influential figures as the Microsoft founder Bill Gates and the late Stephen Hawking have expressed concern about creating machines that are more intelligent than we are. Even though superintelligence seems decades away, they and others have said, shouldn't we consider the consequences before it's too late?


Velocity of the future
How deep & wide is mind-space?
How broadly dispersed are people's values?
How diverse are people & their experience?
How different are the goal functions people hold?
TEXTURE
How long do we survive?
How long can we keep civilization running?
How long are our lifespans?
To what technological heights do we ascend?
How happy is the average person?
Do we have lots of affordances?
How glorious is our transhumanist future?
How powerful is human civilization?
FRAGMENTATION – how many 'bits' is civilization in? Do different groups/agents control different parts of the universe? How much diversity is there in the goal functions that are powerfully implemented?
LENGTH
HEIGHT
ROTATION
Are future people's values similar
Or is
time?
ALIGNMENT
Does the world fulfil the values of the people in it?
Are people able to optimise the world to better fit their values?
Are people free from coercion / having other agents' values imposed on them?
Is there value inequality – some people's values fulfilled much more than others?
people?

Nick Bostrom, whose 2014 book, "Superintelligence: Paths, Dangers, Strategies" had an outsize — some would argue fear-mongering — effect on the A.I. discussion.
Tom Jamieson for The New York Times

"The kind of systems we are creating are very powerful," said Bart Selman, a Cornell University computer science professor and former Bell Labs researcher. "And we cannot understand their impact."

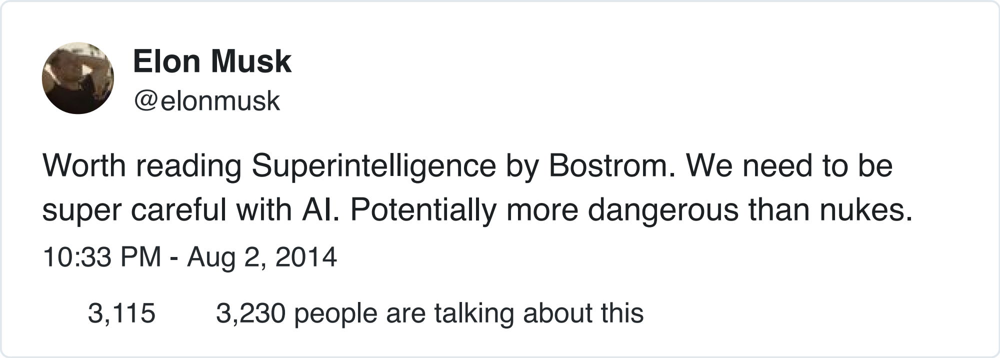
The Imperfect Messenger

Pacific Grove is a tiny town on the central coast of California. A group of geneticists gathered there, in the winter of 1975 to discuss whether their work — gene editing — would end up harming the world. In January 2017, the A.I. community held a similar discussion in the beachside grove.

The private gathering at the Asilomar Hotel was organized by the Future of Life Institute, a think tank built to discuss the existential risks of A.I. and other technologies.

The heavy hitters of A.I. were in the room — among them Mr. LeCun, the Facebook A.I. lab boss who was at the dinner in Palo Alto, and who had helped develop a neural network, one of the most important tools in artificial intelligence today. Also in attendance was Nick Bostrom, whose 2014 book, "Superintelligence: Paths, Dangers, Strategies" had an outsized — some would argue fear-mongering — effect on the A.I. discussion; Oren Etzioni, a former computer science professor at the University of Washington who had taken over the Allen Institute for Artificial Intelligence in Seattle; and Demis Hassabis, who heads DeepMind, an influential Google-owned A.I. research lab in London.

And so was Mr. Musk, who in 2015 had donated $10 million to the Cambridge, Mass., institute. That same year, he also helped create an independent artificial intelligence lab, OpenAI, with an explicit goal: create superintelligence with safeguards meant to ensure it won't get out of control. It was a message that clearly aligned him with Mr. Bostrom.



Elon Musk
@elonmusk

Worth reading Superintelligence by Bostrom. We need to be super careful with AI. Potentially more dangerous than nukes.

10:33 PM - Aug 2, 2014

3,115 3,230 people are talking about this

On the second day of the retreat, Mr. Musk took part in a nine-person panel dedicated to the superintelligence question. Each panelist was asked if superintelligence was possible. As they passed the microphone down the line, each said "Yes," until the microphone reached Mr. Musk. "No," he said. The small auditorium rippled with knowing laughter. Everyone understood that Mr. Musk thought superintelligence was not only possible, but very dangerous.

Mr. Musk later added: "We are headed toward either superintelligence or civilization ending."

At the end of the panel, Mr. Musk was asked how society can best live alongside superintelligence. What we needed, he said, was a direct connection between our brains and our machines. A few months later, he unveiled a start-up, called Neuralink, backed by $100 million that aimed to create that kind of so-called neural interface by merging computers with human brains.

Warnings about the risks of artificial intelligence have been around for years, of course. But few of those Cassandras have the tech cred of Mr. Musk. Few, if any, have spent as much time and money on it. And perhaps none has had as complicated a history with the technology.

Just a few weeks after Mr. Musk talked about his A.I. concerns at the dinner in Mr. Zuckerberg's house, Mr. Musk phoned Mr. LeCun, asking for the names of top A.I. researchers who could work on his self-driving car project at Tesla. (That autonomous technology was in use at the time of two fatal Tesla car crashes, one in Florida in May 2016 and the other in March of this year.)

During a recent Tesla earnings call, Mr. Musk, who has struggled with questions about his company's financial losses and concerns about the quality of its vehicles, chastised the news media for not focusing on the deaths that autonomous technology could prevent — a remarkable stance from someone who has repeated warned the world that A.I. is a danger to humanity.

The tussle in Palm Springs

There is a saying in Silicon Valley: We overestimate what can be done in three years and underestimate what can be done in 10.

On Jan. 27, 2016, Google's DeepMind lab unveiled a machine that could beat a professional player at the ancient board game Go. In a match played a few months earlier, the machine, called AlphaGo, had defeated the European champion Fan Hui — five games to none.

Even top A.I. researchers had assumed it would be another decade before a machine could solve the game. Go is complex — there are more possible board positions than atoms in the universe — and the best players win not with sheer calculation, but through intuition. Two weeks before AlphaGo was revealed, Mr. LeCun said the existence of such a machine was unlikely.

A few months later, AlphaGo beat Lee Sedol, the best Go player of the last decade. The machine made moves that baffled human experts but ultimately led to victory.

Many researchers, including the leaders of DeepMind and OpenAI, believe the kind of self-learning technology that underpins AlphaGo provided a path to "superintelligence." And they believe progress in this area will significantly accelerate in the coming years.

OpenAI recently "trained" a system to play a boat racing video game, encouraging it to win as many game points as it could. It proceeded to win those points but did so while spinning in circles, colliding with stone walls and ramming other boats.

It's the kind of unpredictability that raise grave concerns about the rise of A.I., including superintelligence.

All sorts of deep thinkers have joined the debate over artificial intelligence, including those at an annual conference hosted in Palm Springs, Calif., by Amazon's chief executive, Jeff Bezos.
Jack Nicas/The New York Times

But the deep opposition to these concerns was on display in March at an exclusive conference organized by Amazon and Mr. Bezos in Palm Springs.

One evening, Rodney Brooks, a roboticist at the Massachusetts Institute of Technology, debated the potential dangers of A.I. with the neuroscientist, philosopher and podcaster Sam Harris, a prominent voice of caution on the issue. The debate got personal, according to a recording obtained by The Times.

Mr. Harris warned that because the world was in an arms race toward A.I., researchers may not have the time needed to ensure superintelligence is built in a safe way.

"This is something you have made up," Mr. Brooks responded. He implied that Mr. Harris's argument was based on unscientific reasoning. It couldn't be proven right or wrong — a real insult among scientists.

"I would take this personally, if it actually made sense." Mr. Harris said.

A moderator finally ended the tussle and asked for questions from the audience. Mr. Etzioni, the head of the Allen Institute, took the microphone. "I am not going to grandstand," he said. But urged on by Mr. Brooks, he walked onto the stage and laid into Mr. Harris for three minutes, saying that today's A.I. systems are so limited, spending so much time worrying about superintelligence just doesn't make sense.

The people who take Mr. Musk's side are philosophers, social scientists, writers — not the researchers who are working on A.I., he said. Among A.I. scientists, the notion that we should start worrying about superintelligence is "very much a fringe argument."

Mr. Zuckerberg goes to Washington

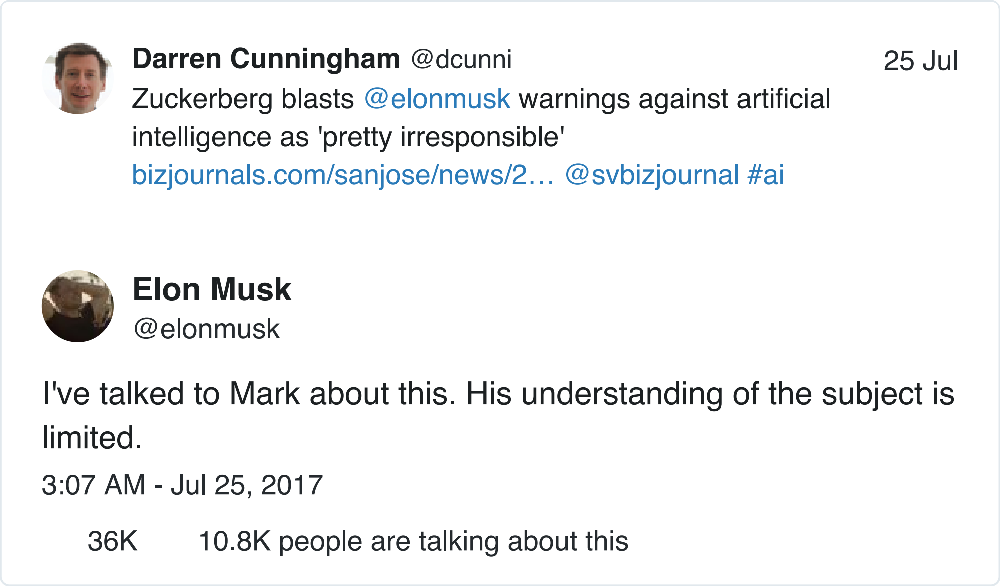
Since their dinner three years ago, the debate between Mr. Zuckerberg and Mr. Musk has turned sour. Last summer, in a live Facebook video streamed from his backyard as he and his wife barbecued, Mr. Zuckerberg called Mr. Musk's views on A.I. "pretty irresponsible."

Panicking about A.I. now, so early in its development, could threaten the many benefits that come from things like self-driving cars and A.I. health care, he said.

"With A.I. especially, I'm really optimistic," Mr. Zuckerberg said. "People who are naysayers and kind of try to drum up these doomsday scenarios — I just, I don't understand it."

In other words: You're getting ahead of reality, Elon. Relax.

Mr. Musk responded with a tweet . "I've talked to Mark about this," Mr. Musk wrote. "His understanding of the subject is limited."



Darren Cunningham @dcunni 25 Jul

Zuckerberg blasts @elonmusk warnings against artificial intelligence as 'pretty irresponsible' bizjournals.com/sanjose/news/2… @svbizjournal #ai

Elon Musk
@elonmusk

I've talked to Mark about this. His understanding of the subject is limited.

3:07 AM - Jul 25, 2017

36K 10.8K people are talking about this

In April, Mr. Zuckerberg testified before Congress, explaining how Facebook was going to fix the problems it had helped create.

One way to do it? By leaning on artificial intelligence. But in his testimony, Mr. Zuckerberg acknowledged that scientists haven't exactly figured out how some types of artificial intelligence are learning.

"This is going to be a very central question for how we think about A.I. systems over the next decade and beyond," he said. "Right now, a lot of our A.I. systems make decisions in ways that people don't really understand."

Tech bigwigs and scientists may mock Mr. Musk for his Chicken Little routine on A.I., but they seem to be moving toward his point of view.

Inside Google, a group is exploring flaws in A.I. methods that can fool computer systems into seeing things that are not there. Researchers are warning that A.I. systems that automatically generate realistic images and video will soon make it even harder to trust what we see online. Both DeepMind and OpenAI now operate research groups dedicated to "A.I safety."

Mr. Hassabis, the founder of DeepMind, still thinks Mr. Musk's views are extreme. But he said the same about the views of Mr. Zuckerberg. The threat is not here, he said. Not yet. But Facebook's problems are a warning.

"We need to use the downtime, when things are calm, to prepare for when things get serious in the decades to come," said Mr. Hassabis. "The time we have now is valuable, and we need to make use of it."

Follow Cade Metz on Twitter: @CadeMetz

Cade Metz reported from San Francisco. Jack Nicas contributed reporting from Palm Springs, Calif.

A version of this article appears in print on June 10, 2018, on Page BU1 of the New York edition with the headline: Moguls and Killer Robots