

Is AI our friend or foe?

Thoughts on AI and humanity

By Nicole Fung · Sunday, March 26, 2023 · Updated Oct 2026 · 7 min read

Thanks for listening to the beat.

The topic of artificial intelligence (AI) is a hot topic especially after the release of Midjourney, ChatGPT, and a cohort of other applications and plug-ins this past year ('22-'23). Suddenly, the idea of AI being plausible and capable of creating a big ripple in humanity seems both exhilarating and ominously daunting. Will humanity fall victim to our own device?

Updated Oct 2026. This piece first ran in March 2023. The original argument stands. New sections and numbers show where things have moved since, with sources at the end.

Three years on: where we stand

When this post was written, ChatGPT was four months old. Today AI is part of ordinary office life. Stanford's 2026 AI Index reports that organizational adoption of AI reached 88%.¹ McKinsey's 2026 global survey finds that eight in ten respondents say AI has improved their own productivity.²

Public feeling has not caught up with the technology, though. In a March 2026 Pew Research Center survey, half of U.S. adults said the growing use of AI in daily life makes them feel more concerned than excited. Just 10% said they were more excited than concerned, and 38% said they felt both equally. Back in 2021, 37% said they were more concerned than excited.³

The gap between experts and everyone else is also wide. Pew found that 56% of AI experts think AI will have a very or somewhat positive impact on the U.S. over the next 20 years, compared with 17% of the general public.³ So the question in the headline is still the right one to ask. Most people simply haven't made up their minds.

The AGI worry

There is some consensus that once AI is developed past its current form into the realm of AGI (*Artificial general intelligence*), it would inherently have the capability of self-determination and could possibly, as a scenario, decide that humanity is subservient to its own agenda or other holistic sustainability

consideration that we have helped instill into its training. A scenario that is often used is—it may consider humans as the biggest obstacle to climate change and therefore, decide that eliminating humanity is the simplest, quickest, and easiest solution to solving the problem.

Two months after this post first ran, hundreds of researchers and executives signed a one-sentence statement from the Center for AI Safety: “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”⁴ Signatories included Geoffrey Hinton and leaders from OpenAI and Google DeepMind. The fear is not just science fiction. Some of the people building these systems share it.

The most careful recent summary is calmer, though. The International AI Safety Report 2026, chaired by Yoshua Bengio, describes “loss of control” as scenarios where AI systems operate outside anyone’s control. It then adds: “Current systems lack the capabilities to pose such risks, but they are improving in relevant areas such as autonomous operation.”⁵

The same report notes that today’s AI is uneven: “capabilities remain ‘jagged’: leading systems may excel at some difficult tasks while failing at other, simpler ones.”⁵ That is a useful picture to keep in mind. The systems we use every day are brilliant in places and clumsy in others. They are nowhere near the all-knowing machine of the climate scenario.

A tool, with trade-offs

However, we should remind ourselves that although AI is a tool that we should have healthy trepidation over, as with any tool, it is always a trade-off of benefits, costs, and probability.

There is no denying that as I am creating this website on technology, humanity, and using businesses to do good, I am enhanced by AI to fill my many shortcomings, including graphics and contents that are my own voice and thoughts but could not have come out as beautifully and quickly as they could have otherwise. It endows me with powers that I would not have without greater effort and gives me much more control over my own artistic license.

The research since then backs up that personal experience. In a 2023 MIT experiment with 453 college-educated professionals, access to ChatGPT cut the time needed for writing tasks by 40 percent, and output quality rose by 18 percent.⁶ Benefits like that are real and measurable.

The costs are real too, and they tend to be ordinary rather than dramatic. Stanford’s AI Index counted 362 documented AI incidents in 2025, up from 233 in 2024.¹ Most of them involve errors, misuse, and poor oversight, not machines with an agenda.

The self-driving dilemma

Jumping back to the autonomous decision-making by AI, I am reminded of the thought experiment relating to self-driving where the computer would have to make the split decision that in order to avoid the crash with a group of schoolchildren. The best-case scenario would be for my smart car to kill me

and avoid the kids. Yet, no driver wants that even though it is morally right. As a follower of Tesla and the full self-driving journey, I am inclined to believe that the best way to avoid AI having to make that decision is to eliminate that scenario by beefing up all other safety features and warning systems.

Researchers have tested exactly this tension. In a 2016 study in *Science*, participants approved of self-driving cars that would sacrifice their passengers for the greater good, and wanted others to buy them. But they “would themselves prefer to ride in AVs that protect their passengers at all costs.”⁷ People agree with the moral rule in theory and reject it for their own family.

The industry has mostly taken the route described above: make the impossible choice rarer. Waymo, which runs driverless taxis in several U.S. cities, reports 271.3 million rider-only miles through June 2026. Compared with human drivers over the same distance in the same areas, it reports 82% fewer injury-causing crashes and 95% fewer crashes with serious injuries or worse.⁸ Those are the company’s own figures, but it publishes its methods alongside them.

Supervision still matters. In October 2025, the U.S. National Highway Traffic Safety Administration opened a preliminary evaluation covering an estimated 2,882,566 Tesla vehicles with Full Self-Driving. The investigation looks at traffic-law violations, including driving through red lights.⁹ Better safety systems reduce the dilemma. They don’t remove the need for checks.

The everyday risks come first

For most businesses, the near-term risks look less like AGI and more like a confident mistake in a customer conversation. In a 2024 Canadian tribunal case, Air Canada’s website chatbot told a grieving customer he could claim a bereavement discount after booking. That was wrong. The airline argued the chatbot was responsible for its own words. The tribunal disagreed and ordered the airline to pay damages of \$812.02.¹⁰

A second risk is how people react to AI, not what the AI does. The International AI Safety Report notes that “reliance on AI tools can weaken critical thinking skills” and can encourage “automation bias,” which it describes as “the tendency to trust AI system outputs without sufficient scrutiny.”⁵

Both risks are manageable with the same habit: a named person checks what the AI says before it reaches a customer, and the business owns the result. That is the practical version of “safeguarding” AI.

So, friend or foe?

Yes, AGI could be scary, but we can't and shouldn't stop AI's progress. We should use AI for the good of humanity and safeguard it, so it wouldn't have to make any drastic decision that is undesirable in the future to people and society.

Three years later, that still feels like the right balance. The long-term questions deserve serious people working on them, and they now have them. The rest of us can make AI a friend in the places we control: our teams, our products, and our customer promises.

WHAT IT MEANS FOR YOU

- Treat AI as a powerful tool with trade-offs, not as a friend or a foe by default.
- Put a person in charge of checking AI output before it reaches customers. You own what your chatbot says.
- Design out the hard choices where you can, the way safer cars reduce split-second dilemmas.
- Keep an eye on the long-term debate, but spend your energy on the everyday risks you can manage now.

SOURCES

1. The 2026 AI Index Report (<https://hai.stanford.edu/ai-index/2026-ai-index-report>), Stanford HAI, 2026
2. The state of AI in 2026: On the road to ROI (<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>), McKinsey & Company, Aug 25, 2026
3. Key findings about how Americans view artificial intelligence (<https://www.pewresearch.org/short-reads/2026/03/12/key-findings-about-how-americans-view-artificial-intelligence/>), Pew Research Center, Mar 12, 2026
4. Statement on AI risk (<https://aistatement.com/>), Center for AI Safety, May 2023
5. International AI Safety Report 2026: Executive summary (<https://internationalaisafetyreport.org/publication/2026-report-executive-summary>), International AI Safety Report, 2026
6. Study finds ChatGPT boosts worker productivity for some writing tasks (<https://news.mit.edu/2023/study-finds-chat-gpt-boosts-worker-productivity-writing-0714>), MIT News, Jul 14, 2023
7. The social dilemma of autonomous vehicles (<https://www.science.org/doi/10.1126/science.aaf2654>), Science (Bonnefon, Shariff and Rahwan), Jun 24, 2016
8. Safety impact (<https://waymo.com/safety/impact/>), Waymo, data through June 2026
9. ODI resume, preliminary evaluation PE25012 (<https://static.nhtsa.gov/odi/inv/2025/INOA-PE25012-19171.pdf>), National Highway Traffic Safety Administration, Oct 7, 2025
10. A word of caution: Company liable for misrepresentations made by chatbot (<https://mcmillan.ca/insights/a-word-of-caution-company-liable-for-misrepresentations-made-by-chatbot/>), McMillan LLP, Feb 2024

Frequently asked questions

Is AI our friend or foe?

It can be a friend, if we safeguard it. AGI could be scary, but we can't and shouldn't stop AI's progress. We should use AI for the good of humanity and make it a friend in the places we control: our teams, our products, and our customer promises.

Can today's AI slip out of human control?

Not yet. The International AI Safety Report 2026 says: "Current systems lack the capabilities to pose such risks, but they are improving in relevant areas such as autonomous operation."

What is the biggest near-term AI risk for a business?

A confident mistake in a customer conversation. Air Canada had to pay \$812.02 in damages after its website chatbot gave a grieving customer wrong advice about a bereavement discount. The fix: a named person checks what the AI says before it reaches a customer.

What is automation bias?

The tendency to trust AI system outputs without sufficient scrutiny, as the International AI Safety Report describes it. Reliance on AI tools can also weaken critical thinking skills.

First published on nicolefung.pro on March 26, 2023. Updated Oct 2026 with new data and sources; the original text is kept, lightly edited for US spelling and obvious typos.
