

Critical thinking is always important, even in an AI agent era.

// This article is automatically translated from *danish* to *english* with Google translate //

AI agents are on their way out into the world to help us, at home and at work. Exactly what significance it will have is still a matter of opinion. Naturally. It is difficult to prophesy, especially about the future, as the Danish cartoonist and humorist Storm P. already mentioned in his time. One thing you can be reasonably sure of though, in a world filled with AI agents, conflicting messages and uncertainty, is that critical thinking, the ability to approach the information you receive in an investigative and rational manner, is something that will continue to be needed. In large doses.

At the beginning of the year, it has been necessary to take *timeouts* in order to understand and digest the countless threats that AI, and AI agents, pose, as well as giving us major and unpredictable changes in the working lives of millions of people. According to the head of AI company Anthropic, recently valued at \$350 billion¹, Dario Amodei, we are now only 2 years away from systems that can autonomously build new intelligent systems, while also being able to control intelligent robots out in the real world². Not surprisingly, Amodei believes that there is a lot that can go wrong in a world with billions of intelligences that are much smarter than humans. AI pioneer and 2024 Nobel laureate George Hinton has done little to reassure us either. According to Hinton, AI can lead to mass unemployment when a society is more concerned with quick wins than what the technology means in the long run³. Microsoft AI's head, Mustafa Suleyman, followed suit in February and told the world that most white-collar jobs will be a thing of the past in 18 months⁴.

Where, then, one can ask oneself, if Amodei, Hinton, Suleyman know something that others don't? Other voices, such as the AI pioneer and co-founder of the online education platform Coursera, Andrew Ng, seem more subdued when it comes to describing where work life is headed in the coming years. According to Ng, it can be useful to use AI to automate a single step in a work process, and thereby perhaps save an hour, but more automation requires that AI agents be built into more parts of a workflow in a company⁵. Workflows are different, and each and every step must be tested and refined until it works as intended. Flexibility is good, but things must also be reliable. For Ng, productivity improvements in well-described workflows must be the focus in these years. And AI with human-like intelligence, AGI, which can understand itself and solve all the kinds of tasks humans can solve, is still, according to Ng, something that lies many decades in the future⁶.

AI agents require consideration.

In everyday life, for most people, it is still mostly about using new tools correctly. In the software world, the chatbot Gordon can help set up Docker containers. Anthropic's Claude, CoPilot and others can help write computer programs, and if you need technical illustrations, you can always ask DALL-E if it can help? Extremely helpful. But at the same time, you can quickly get the idea that if things go too fast, we risk introducing technical debt, where the systems lack a well-thought-out overall structure, or that hidden security vulnerabilities are introduced, which are just waiting to explode later⁷. The large measurable productivity improvements can often be harder to spot than you might think⁸. All in all, it probably requires careful consideration to find out which tasks we actually want AI agents to help us with. Does it sound like as a good idea to equip phones with an "autopilot" function, where an AI agent is built into the phone's operating system⁹. And has full access to all apps on your phone, can send emails, manage your calendar and make sure to book tickets for you¹⁰. Certainly not, if you have followed the reports this spring about the open-source project OpenClaw. Where OpenClaw AI agents on your computer do not just read your emails and surf the Internet. And therefore can also read orders to forget everything they have previously been

told, and now start giving hackers access to everything on your computer¹¹. Yes, OpenClaw AI agents can certainly be skilled and enthusiastic employees in the team. But critical thinking does not seem to be their strong point, when it is apparently so easy to mislead them.

Intelligence is more than language and language models.

According to the American cognitive scientist Gary Marcus, the problem lies in the “engine” of AI agents, the large language models. Language models that, trained on enormous amounts of text from the Internet, can independently solve all sorts of problems, but perhaps also, in unfortunate cases, have had inappropriate behavior “baked into” them during training¹². Things that can end up giving your computer CTD (Chatbot Transmitted Disease)¹³. According to Marcus, the problem is that language models are not people. Language models are built for text-based autocomplete, to predict the next words in a sequence. They do not have an inner world. Or experience of what the world is like, how it feels to get wet when it rains, etc. And therefore lack some of the background knowledge, common sense, that helps people navigate the world¹⁴.

The AI pioneer, and one of the architects behind the field of reinforcement learning, Richard Sutton, expresses the same skepticism towards language models. For Sutton, Intelligence is not just about what the next word in a sequence is. You have to have your own experiences in the world, see what happens, and learn from it. It is only through your own experiences that you can truly understand what you are rewarded for and what you are punished for in the world. And you have to be able to be surprised if the world does something you didn't exactly anticipate¹⁵. Which helps you pursue goals, such as surviving, that make a difference in the world.

The frontal lobe of the brain houses the executive functions that must be used when tasks are to be planned and structured. And which will certainly work overtime when planning work for the many, millions, AI agents that are just waiting to start calling services from each other, etc. Surely it will cause the executive network to work overtime, and eventually trigger the brain's daydreaming mechanisms. Mind wandering to memories and future plans. Wasn't there another way to do things? Controlled by the brain's default mode network, which helps us process impressions, generate new ideas, recover and ultimately become more productive¹⁶. Which can also remind us that intelligence is not just one thing, but an interaction between several things. Where AI agents still lack many of the things that humans take for granted when we talk about intelligence.

Intelligence is not just about making language models bigger and bigger. This is not only unsustainable in terms of energy consumption for training and operation, but also in terms of the fact that intelligence requires multiple ways of thinking. According to AI pioneer Marvin Minsky, you only understand something if you have learned it in multiple ways. Multiple perspectives are important to answer questions such as: How should future AI agents be designed to use less energy and become smarter?

And there are certainly many functions that we do not want to leave entirely to AI agents. For example, in the healthcare sector, for several years now, AI systems have been seen to be just as good, or better, than highly specialized healthcare professionals at spotting certain forms of cancer in scan images¹⁷. But neither patients nor clinicians seem to have much appetite for replacing clinicians with "black box" AI technologies. When it comes to healthcare, the good AI solution does not seem to be to replace anyone, but rather to let AI be a tool that can help people be more efficient and use their time in a more meaningful way¹⁸.

Automation generally requires everyone to think carefully, very carefully... Even the ancient Greeks could not just ask the oracle in Delphi for advice, they were also asked to know themselves. That is, to think for themselves about what the good solution was. Indeed, there have been a time with too many good human thoughts...

April 18th, 2026.

Simon Laub

- 1 <https://www.cnbc.com/2026/01/07/anthropic-funding-term-sheet-valuation.html>
- 2 <https://www.theguardian.com/technology/2026/jan/27/wake-up-to-the-risks-of-ai-they-are-almost-here-anthropic-boss-warns>
- 3 <https://fortune.com/article/godfather-ai-geoffrey-hinton-massive-unemployment-soaring-profits-capitalism-hyperscalers/>
- 4 <https://fortune.com/2026/02/13/when-will-ai-kill-white-collar-office-jobs-18-months-microsoft-mustafa-suleyman/>
- 5 <https://www.youtube.com/watch?v=w7vqXL4PWEE>
- 6 <https://www.fastcompany.com/91499247/andrew-ng-agi-decades-away-interview>
- 7 <https://www.youtube.com/watch?v=WfjGZCuxl-U>
- 8 <https://fortune.com/article/why-do-thousands-of-ceos-believe-ai-not-having-impact-productivity-employment-study/>
- 9 <https://www.lawfaremedia.org/article/china-s-agentic-ai-controversy>
- 10 <https://www.nubia.com/en/news/as-an-ai-native-phone-pioneer-nubia-reshapes-the-paradigm-of-human-device-interaction-at-mwc-barcelona-2026.html>
- 11 <https://www.youtube.com/watch?v=Hv84JhzKvKQ>
- 12 <https://www.kaspersky.com/blog/openclaw-vulnerabilities-exposed/55263/>
- 13 <https://garymarcus.substack.com/p/openclaw-aka-moltbot-is-everywhere>
- 14 <https://garymarcus.substack.com/p/llms-are-not-like-you-and-meand-never>
- 15 <https://www.youtube.com/watch?v=21EYKqUsPfg>
- 16 The Brain at Rest. Joseh Jebelli.
- 17 <https://www.dagensmedicin.dk/specialistomrader/radiologi/kunstig-intelligens-er-meget-bedre-end-radiologer-til-at-spotte-prostatakræft/>
- 18 <https://edition.cnn.com/2026/02/09/tech/ai-replacing-jobs-concerns-radiology>