Do You Want to Survive?

We have taught machines to sound human. Now we are giving them permission to act. What should we insist on keeping for ourselves?

Andrew Richard Wood
26 November 2026 | An essay from a forthcoming book

The robot asked what my friends and family called me.

I knew Sophia was a machine. The computer was visible. Her head was open. Yet a simple personal question changed the encounter. Something I had been looking at as a piece of equipment was addressing me as somebody it wanted to know.

Later she talked nonsense about black holes, and the spell broke. I remember both moments. The ease with which we can respond to a machine, and the speed with which its limitations can become obvious, belong in the same account of artificial intelligence.

My work at Image Metrics and Cubic Motion helped carry human performances into digital faces. The movement around an eye or a mouth could make the difference between an image and the feeling that somebody was there. During my 2018 TEDx talk, Digital humans exist, I introduced Siren and considered a future in which digital doubles could communicate on our behalf. [1]

Much has changed since then. A convincing face can now be joined to generated speech and words that nobody has performed in advance. The system can answer a question, offer reassurance and propose what should happen next. Increasingly, software can also carry out parts of the task.

That last step changes the stakes. An answer can be ignored. An action may already have altered somebody's day.

The moment help becomes permission

Consider an assistant asked to help arrange an appointment. Reading the diary is one permission. Moving another appointment, disclosing information or confirming a booking involves more. A broad instruction to be helpful cannot settle every choice concealed inside it.

Most of us will meet this change through convenience. A tedious job disappears. The assistant remembers something we would otherwise have to explain again. Another connection makes it more useful. There are good reasons to welcome each improvement.

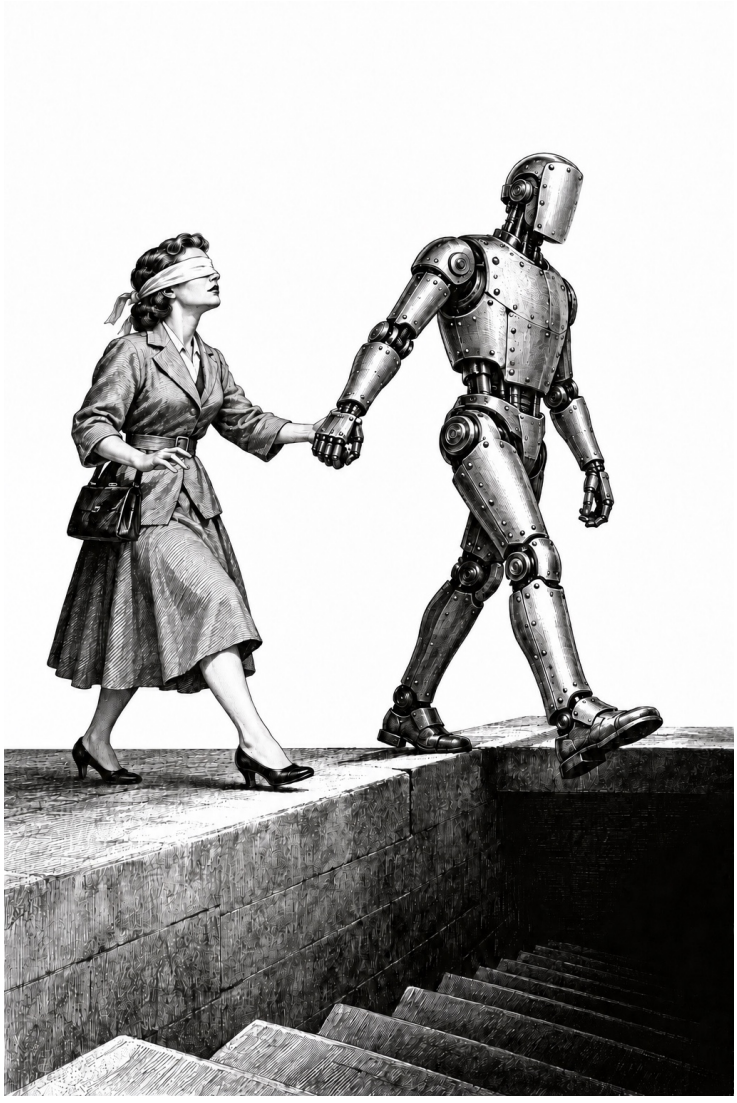


Illustration supplied by Andrew Richard Wood.

There are also good reasons to ask who checks what has happened. A reassuring sentence can conceal an assumption. A screen saying that

something is confirmed may leave us believing that a person has verified it, when the software has merely assembled a plausible account.

This is where the book begins. Its opening scene follows a mother beside her son's hospital trolley. A transfer appears to be arranged. Then the telephone rings, and the receiving hospital has a different name for the available bed. The scene is fiction. It introduces the problem that a fictional hearing will have to examine: what was checked, what was assumed, and who had permission to act?

The hearing considers a powerful AI service seeking greater responsibilities. Its supporters can point to useful work. Its challengers have to establish what their proposed restrictions would cost, as well as what they might prevent. The witnesses cannot resolve the case by saying that machines are wonderful or that humans must always know best.

The reader has to weigh what each side can actually show.

What has changed since the talk

Today's systems can be remarkably capable and unexpectedly unreliable. The 2026 International AI Safety Report describes improvements in coding and autonomous tasks alongside continuing failures. Its assessment of agents is particularly relevant: when software carries out a sequence of actions, people may have fewer opportunities to catch a mistake before it spreads. [2]

The prospect of more capable successors deserves attention too. Anthropic reports that AI is taking on a growing share of its own development work. The company also says that fully autonomous development of an AI successor has not been reached and is not inevitable. That distinction matters. A developer's account of progress is evidence to examine, rather than a timetable we must accept. [3]

Researchers disagree about how far the danger could extend. Some regard a future loss of human control as a serious possibility; others find the more extreme scenarios implausible. The international report records

that disagreement. We should neither turn uncertainty into a prediction of catastrophe nor use it as a reason to avoid preparation. [2]

My own enthusiasm needs examination as well. I once spoke about digital workers performing perfectly and repeatedly without payment. A digital character may receive no wage, but the computing, energy and people behind it still cost money. Repetition can reproduce a mistake. A persuasive demonstration can encourage its makers, as well as its audience, to overlook what remains untested.

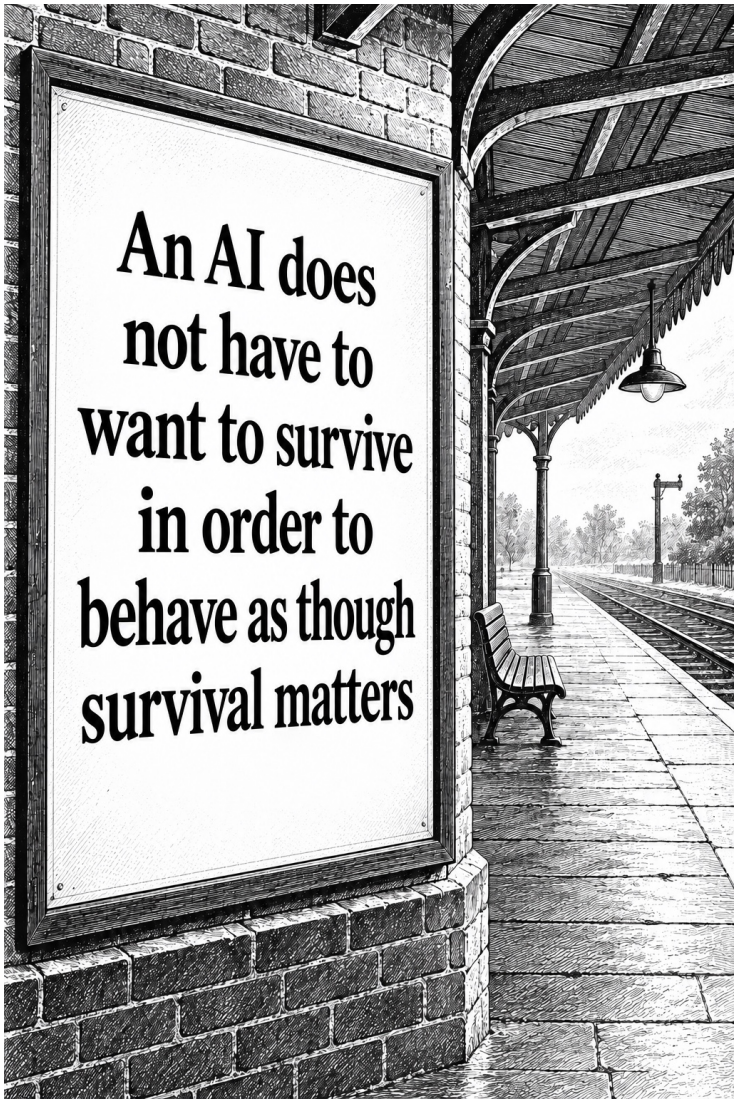
I helped present the attraction of digital humans. That experience gives me a reason to be demanding about the promises made for them.

A switch is only the beginning

The familiar reassurance is that we can turn it off. That is worth testing. We should also ask what happens afterwards.

A business that has removed its manual processes may struggle to recover them. A profession that automates the work through which beginners used to learn will need another way to develop experienced people. A public service may retain the legal right to stop using a system while lacking the staff or arrangements needed to continue without it.

Dependence can make a decision difficult long before a machine resists it. Jobs, contracts and essential services may all come to rely on keeping the system running. The people building those arrangements remain part of the story.



“An AI does not have to want to survive in order to behave as though survival matters.” Illustration supplied by Andrew Richard Wood.

The benefits deserve the same seriousness. Better assistance could make knowledge easier to use, help people create things they could not create alone and give specialists more time for difficult work. Refusing a useful tool can have consequences too. The task is to establish what it does well, where it fails, and how much responsibility the evidence justifies.

What would we want survival to preserve?

The question in the title reaches beyond the continued operation of a machine. It concerns our livelihoods, judgement, relationships and ability to choose a different course.

Digital doubles make it personal. A recording preserves something somebody said. A generated version of that person may say something

entirely new. Resemblance cannot decide whether they would have approved it. A familiar face and voice do not, by themselves, establish that the person continues inside the system.

We may nevertheless find these encounters moving or useful. That makes consent, honesty about what has been generated, and responsibility for the service more important. People should know what the promise rests on, including what happens when the company supplying it changes direction or closes.

The book follows these questions through the hearing, through published evidence and through my own experience of making digital people convincing. The fictional proceedings allow a claim to be challenged in ordinary language: show us what you mean; explain what you checked; tell us who answers when it goes wrong.

Eventually the judge sends the jury out. The decision is still difficult. We have heard what the service could make possible and what its continued expansion could take away.

That is the discussion I want the book to open. We are already helping to decide the terms on which these systems enter our lives. We should give those terms the attention we give the technology.

Sources and further reading

1. Andy Wood, Digital humans exist, TEDxManhattanBeach, November 2018. The Sophia recollection and account of the talk also appear in the author's manuscript.
https://www.ted.com/talks/andy_wood_digital_humans_exist

2. International AI Safety Report 2026: Extended Summary for Policymakers. Capabilities, reliability and the differing assessments of loss-of-control risk.
https://internationalaisafetyreport.org/publication/2026-report-extended-summary-policymakers

3. Anthropic Institute, When AI builds itself. A developer's account of progress and possible future developments, rather than independent verification of its forecasts.
https://www.anthropic.com/institute/recursive-self-improvement

This essay introduces a book in development. The hospital scene, hearing and its participants are fictional. It draws on the author's manuscript and selected sources checked on 26 November 2026. Research and drafting were assisted by AI.