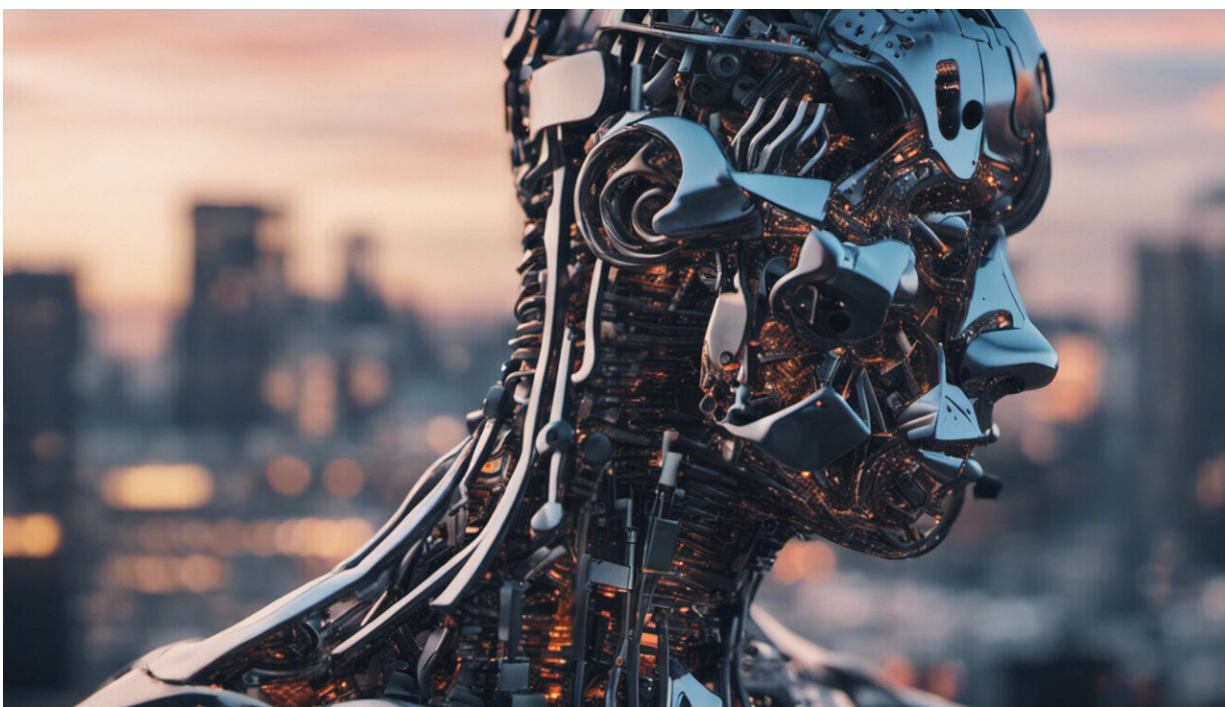


AI can process more information than humans—so will it stop us repeating our mistakes?

April 27 2023, by Anders Sandberg



Credit: AI-generated image

It is a cliché that not knowing history makes one repeat it. As many people have also pointed out, the only thing we learn from history is that we rarely learn anything from history. People engage in [land wars in Asia](#) over and over. They repeat the same dating mistakes, again and

again. But why does this happen? And will technology put an end to it?

One issue is forgetfulness and "[myopia](#)": we do not see how past events are relevant to current ones, overlooking the unfolding pattern. Napoleon ought to have noticed the similarities between his march on Moscow and the Swedish king [Charles XII's failed attempt to do likewise](#) roughly a century before him.

We are also bad at learning when things go wrong. Instead of determining why a decision was wrong and how to avoid it ever happening again, we often try to ignore the embarrassing turn of events. That means that the next time a similar situation comes around, we do not see the similarity—and repeat the [mistake](#).

Both reveal problems with information. In the first case, we fail to remember personal or historical information. In the second, we fail to encode information when it is available.

That said, we also make mistakes when we cannot efficiently deduce what is going to happen. Perhaps the situation is too complex or too time-consuming to think about. Or we are biased to misinterpret what is going on.

The annoying power of technology

But surely [technology](#) can help us? We can now store information outside of our brains, and use computers to retrieve it. That ought to make learning and remembering easy, right?

Storing information is useful when it can be retrieved well. But remembering is not the same thing as retrieving a file from a known location or date. Remembering involves spotting similarities and bringing things to mind.

An [artificial intelligence](#) (AI) also needs to be able to spontaneously bring similarities to our mind—often unwelcome similarities. But if it is good at noticing possible similarities (after all, it could search all of the internet and all our [personal data](#)) it will also often notice false ones.

For failed dates, it may note that they all involved dinner. But it was never the dining that was the problem. And it was a sheer coincidence that there were tulips on the table—no reason to avoid them.

That means it will warn us about things we do not care about, possibly in an annoying way. Tuning its sensitivity down means increasing the risk of not getting a warning when it is needed.

This is a [fundamental problem](#) and applies just as much to any advisor: the cautious advisor will cry wolf too often, the optimistic advisor will miss risks.

A good advisor is somebody we trust. They have about the same level of caution as we do, and we know they know what we want. This is difficult to find in a human advisor, and even more so in an AI.

Where does technology stop mistakes? Idiot-proofing works. Cutting machines require you to hold down buttons, keeping your hands away from the blades. A "dead man's switch" stops a machine if the operator becomes incapacitated.

Microwave ovens turn off the radiation when the door is opened. To launch missiles, two people need to turn keys simultaneously across a room. Here, careful design renders mistakes hard to make. But we don't care enough about less important situations, making the design there far less idiot-proof.

When technology works well, we often trust it too much. Airline pilots

have fewer true flying hours today than in the past due to the amazing efficiency of autopilot systems. This is [bad news](#) when the autopilot fails, and the pilot has less experience to go on to rectify the situation.

The first of a [new breed of oil platform \(Sleipnir A\) sank](#) because engineers trusted the software calculation of the forces acting on it. The model was wrong, but it presented the results in such a compelling way that they looked reliable.

Much of our technology is amazingly reliable. For example, we do not notice how lost packets of data on the internet are constantly being found behind the scenes, how error-correcting codes remove noise or how fuses and redundancy make appliances safe.

But when we pile on level after level of complexity, it looks very unreliable. We do notice when the Zoom video lags, the AI program answers wrong or the computer crashes. Yet ask anybody who used a computer or car 50 years ago how they actually worked, and you will note that they were both less capable and less reliable.

We make technology more complex until it becomes too annoying or unsafe to use. As the parts become better and more reliable, we often choose to add new exciting and useful features rather than sticking with what works. This ultimately makes the technology less reliable than it could be.

Mistakes will be made

This is also why AI is a double-edged sword for avoiding mistakes. Automation often makes things safer and more efficient when it works, but when it fails it makes the trouble far bigger. Autonomy means that smart software can complement our thinking and offload us, but when it is not thinking like we want it to, it can misbehave.

The more complex it is, the more fantastic the mistakes can be. Anybody who has dealt with highly intelligent scholars know how well they can mess things up with great ingenuity when their common sense fails them—and AI has very little human common sense.

This is also a profound reason to worry about AI guiding decision-making: [they make new kinds of mistakes](#). We humans know human mistakes, meaning we can watch out for them. But [smart machines can make mistakes we could never imagine](#).

What's more, AI systems are programmed and trained by humans. And there are lots of examples of such systems [becoming biased and even bigoted](#). They mimic the biases and repeat the mistakes from the human world, even when the people involved explicitly try to avoid them.

In the end, mistakes will keep on happening. There are fundamental reasons why we are wrong about the world, why we do not remember everything we ought to, and why our technology cannot perfectly help us avoid trouble.

But we can work to reduce the consequences of mistakes. The undo button and autosave have saved countless documents on our computers. [The Monument in London](#), [tsunami stones](#) in Japan and other monuments act to remind us about certain risks. Good design practices make our lives safer.

Ultimately, it is possible to learn something from history. Our aim should be to survive and learn from our mistakes, not prevent them from ever happening. Technology can help us with this, but we need to think carefully about what we actually want from it—and design accordingly.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: AI can process more information than humans—so will it stop us repeating our mistakes? (2023, April 27) retrieved 6 December 2023 from <https://techxplore.com/news/2023-04-ai-humansso.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.