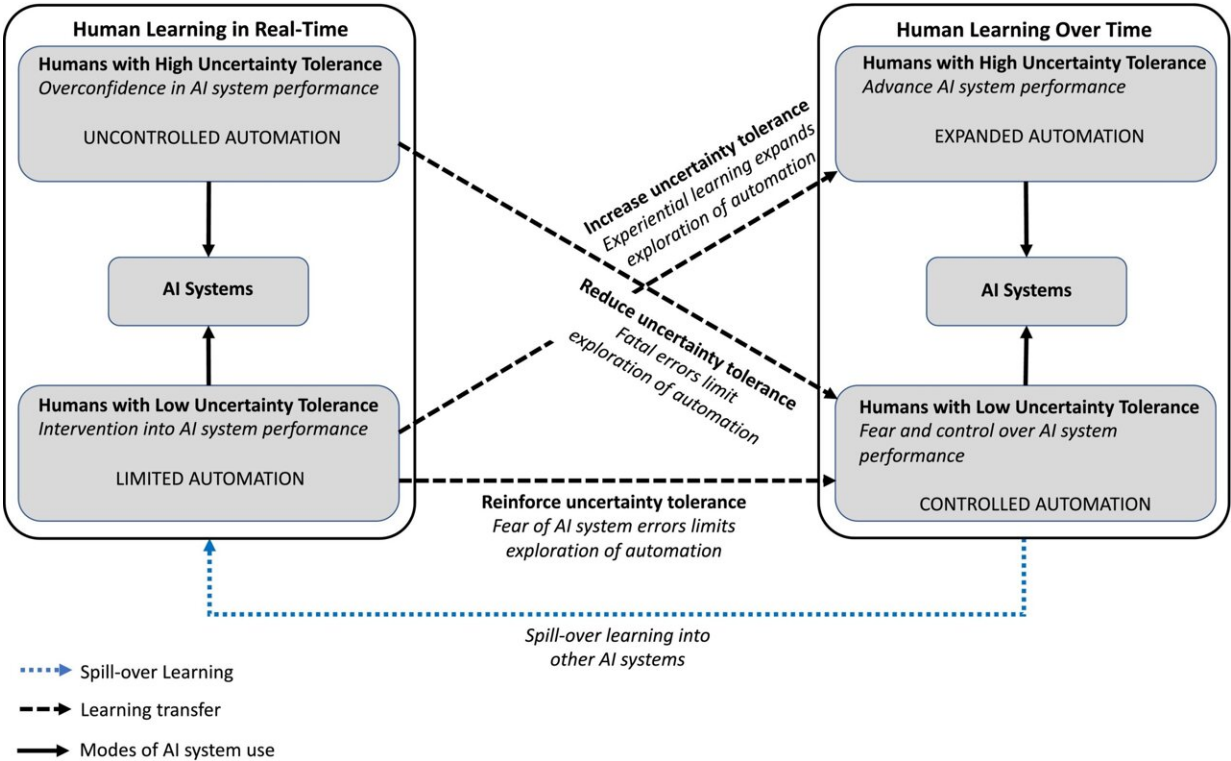![Techxplore logo]

When looking to test new AI tools, make sure to avoid the 'Tesla trap,' say researchers

September 3 2024

Graphical abstract. Credit: *Information and Organization* (2024). DOI: 10.1016/j.infoandorg.2024.100502

If only techno-optimists get to test AI tools, the results simply aren't good enough. This is where Tesla made a big mistake—a lesson that the health sector will do well to learn.

Tesla has been testing [self-driving cars](#) for many years, and NTNU was able to participate behind the scenes as part of [a study published](#) in *Information and Organization*.

Tesla has produced several versions of its autopilot AI tool, which has become better and better with every new release. The tool is trained by its users. In other words, by drivers that utilize the autopilot function.

It is now clear that some of these drivers have been highly risk-tolerant techno-optimists.

Asleep at the wheel

Some test drivers have had so much confidence in the autopilot function that they have fallen asleep at the wheel. This has resulted in accidents, two of which were fatal because the sleeping drivers were unable to intervene and override the autopilot function.

In one case, a car drove straight into a fire engine parked in the road.

The research in which NTNU has participated shows that these drivers, who have the greatest confidence in the autopilot AI, in fact serve only to put a spanner in Tesla's works.

This is because when Tesla owners with low levels of risk tolerance read about the accidents, they lose confidence in the AI tool and never use the autopilot function. This in turn removes major opportunities for the AI tool to learn.

Ask the risk averse first

Research scientists believe that many other enterprises with ambitions to

utilize AI can learn some important lessons from the findings of this research. Not least in the health sector.

By far the most important lesson for all enterprises planning to utilize AI tools is that the role of pilot users should first be offered to the cautious and risk averse.

Such advice is in direct opposition to what has usually been practiced when it comes to testing new digital systems.

The reason for this, of course, is that during the recruitment of pilot users (individuals or teams), whether in connection with AI or something else, it is always the risk-tolerant who put up their hands. In other words, those who are the most positive and optimistic on behalf of new technology.

Such people are typically also those who are willing to take chances during the testing process. It is exactly because of this that they are preferred in the role of pilot users.

Same pattern among doctors

Today at SINTEF and NTNU, researchers are studying [medical doctors](#) and engineers who are cutting their teeth on machine learning models. And they are seeing the same tendencies as recorded in the Tesla study.

Risk-averse users tend to steer away from AI when they encounter examples of tools making mistakes. Their initial skepticism is confirmed, and they lose the motivation to try again.

A probable key reason why it is so easy to 'follow the lead' of the skeptics is that people in general have a strong aversion to things they cannot explain or don't understand.

'Black boxes' make us nervous

The problem with so-called subsymbolic AI (algorithms that recognize everything from shapes to complex patterns with the aid of a process called classification) is that it is perceived as a 'black box.'

Research in the field of so-called Explainable AI, as conducted by workers such as Inga Strümke at NTNU, addresses this issue in particular. Findings from such work show that when something unforeseen happens, people need an explanation.

We accept that people in general, including doctors, make mistakes. This is because we can understand the reason for what has happened. However, we do not know how a subsymbolic AI tool has arrived at its classification. This makes us feel nervous and insecure.

AI can offer key assistance in the health sector

This is probably an important reason why, when 'risk-tolerant' people make fatal mistakes based on poorly founded AI-based advice, any cautiously established confidence held by others can be demolished in minutes.

We argue that society has to get a grip on this issue. Not least because of ideas published by the medical doctor Ishita Barua in her recent book 'Kunstig intelligens redder liv' (Artificial intelligence saves lives, in Norwegian). Barua also has a Ph.D. in the use of AI in medicine. She argues that the health sector must take the plunge and utilize machine learning models because of the enormous potential they offer.

So what can the health sector do to avoid ending up in the 'Tesla trap'?

Key data from the risk averse

When it comes to testing AI models, ask less risk-tolerant doctors. Ask them to experiment with the tool—to look, learn and disseminate their experience without, at least at first, using AI-based advice in the treatment of their patients.

This approach has the potential to offer the models better and more reliable training data. The Tesla study tells us why. When the car's autopilot AI did something peculiar, the 'cautious' drivers took the wheel and corrected for the error. This provided the model with valuable machine learning, enabling the AI tool to adjust its driving style.

In more recent versions of the Tesla autopilot, the AI tool has performed much better in situations similar to those in which it had been previously corrected.

But society risks losing benefits such as this if 'cautious' AI users are scared away by the mistakes made by techno-optimists.