

AI: Why installing 'robot judges' in courtrooms is a really bad idea

July 13 2023, by Rónán Kennedy



Credit: AI-generated image ([disclaimer](#))

Science fiction's visions of the future include many versions of artificial intelligence (AI), but relatively few examples where software replaces human judges. For once, the real world seems to be changing in ways that are not predicted in stories.

In February, a Colombian [judge asked ChatGPT for guidance](#) on how to decide an insurance case. Around the same time, [a Pakistani judge used ChatGPT](#) to confirm his decisions in two separate cases. There are also reports of [judges in India and Bolivia](#) seeking advice from ChatGPT.

These are unofficial experiments, but some systematic efforts at reform do involve AI. In China, judges are [advised and assisted by AI](#), and this development is likely to continue. [In a recent speech](#), the master of the rolls, Sir Geoffrey Vos—the second most senior judge in England and Wales—suggested that, as the legal system in that jurisdiction is digitized, AI might be used to decide some "less intensely personal disputes", such as commercial cases.

AI isn't really that smart

This might initially seem to be a good idea. The law is supposed to be applied impartially and objectively, "without fear or favor". Some say, [what better way to achieve this](#) than to use a computer program? AI doesn't need a lunch break, can't be bribed, and doesn't want a pay rise. AI justice can be applied more quickly and efficiently. Will we, therefore, see "robot judges" in courtrooms in the future?

There are four principal reasons why this might not be a good idea. The first is that, in practice, AI generally acts as an [expert system](#) or as a [machine learning](#) system. Expert systems involve encoding rules into a model of decisions and their consequences—called a decision tree—in software. These had their heyday in law in the 1980s. However, they ultimately [proved unable to deliver good results](#) on a large scale.

Machine learning is a form of AI that improves at what it does over time. It is [often quite powerful](#), but no more so than a very educated guess. One strength is that [it can find correlations and patterns](#) in data that we don't have the capacity to calculate. However, one of its

weaknesses is that it fails in ways that are different to the way people do, reaching conclusions that are obviously incorrect.

In a notable example, an AI was tricked into [recognizing a turtle as a gun](#). Facial recognition often has issues correctly identifying women, children and those with dark skin. So it's possible that AI could also [erroneously place someone at a crime scene](#) who wasn't there. It would be difficult to be confident in a legal system that produced outcomes that were clearly incorrect but also very difficult to review, as the reasoning behind machine learning is not transparent. It has outstripped our ability to understand its inner workings—a phenomenon known as the "black box problem".

When AI is used in legal processes, and it fails, the consequences can be severe. Large language models, the technology underlying AI chatbots such as ChatGPT, are known to write text that is completely untrue. This is known as an AI hallucination, even though it implies that the software is thinking rather than statistically determining what the next word in its output should be.

This year, it emerged that a New York lawyer had used ChatGPT to write submissions to a court, only to discover that it cited [cases that do not exist](#). This indicates that these types of tools are not capable of replacing lawyers yet, and in fact, may never be.

Historical biases

Second, [machine learning](#) systems rely on historical data. In crime and law, these will often [contain bias and prejudice](#). Marginalized communities [will often feature more](#) in records of arrests and convictions, so an AI system might draw the unwarranted conclusion that people from particular backgrounds are more likely to be guilty.

A prominent example of this is the Compas system, an AI algorithm used by US judges to make decisions on granting bail and sentencing. An investigation claimed that [it generated "false positives"](#) for people of color and "false negatives" for white people. In other words, it suggested that people of color would re-offend when they did not in fact do so, and suggested that white people would not re-offend when they did. However, the developer of the system challenges these claims.

Third, it is not clear that legal rules can be reliably converted into software rules. Individuals will interpret the same rule in different ways. When 52 programmers were assigned the task of automating the enforcement of speed limits, the programs that they [wrote issued very different numbers of tickets](#) for the same sample data.

Individual judges may have different interpretations of the law, but they do so in public and are subject to being overturned on appeal. This should reduce the amount of variation in judgments over time—at least in theory. But if a programmer is too strict or too lenient in their implementation of a rule, that may be very difficult to discover and correct.

Automated government systems fail at a scale and speed that's very difficult to recover from. The Dutch government used an [automated system](#) (SyRI) [to detect benefits fraud](#), which wrongly accused many families, destroying lives in the process.

The Australian "Online Compliance Intervention" scheme is used to automatically assess debts from recipients of social welfare payments. It's commonly known as ["Robodebt"](#). The scheme overstepped its bounds, negatively affecting hundreds of thousands of people and was the [subject of a Royal Commission in Australia](#). (Royal Commissions are investigations into matters of public importance in Australia.)

Finally, judging is [not all that judges do](#). They have many other roles in the [legal system](#), such as managing a courtroom, a caseload, and a team of staff, and those would be even more difficult to replace with software programs.

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

Provided by The Conversation

Citation: AI: Why installing 'robot judges' in courtrooms is a really bad idea (2023, July 13) retrieved 28 April 2024 from

<https://techxplore.com/news/2023-07-ai-robot-courtrooms-bad-idea.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.