

Opinion: To understand AI's problems, look at the shortcuts taken to create it

May 12 2023, by Nello Cristianini



Credit: AI-generated image

A machine can only "do whatever we know how to order it to perform," [wrote](#) the 19th-century computing pioneer Ada Lovelace. This reassuring statement was made in relation to Charles Babbage's description of the first mechanical computer.

Lady Lovelace could not have known that in 2016, a program called [AlphaGo](#), designed to play and improve at the board game "Go", would not only be able to defeat all of its creators, but would do it in ways that they could not explain.

In 2023, the AI chatbot [ChatGPT](#) is taking this to another level, holding conversations in multiple languages, solving riddles and even passing legal and medical exams. Our machines are now able to do things that we, their makers, do not know "how to order them to do".

This has provoked both excitement and concern about the potential of this technology. Our anxiety comes from not knowing what to expect from these new machines, both in terms of their immediate behavior and of their future evolution.

We can make some sense of them, and the risks, if we consider that all their successes, and most of their problems, come directly from the particular recipe we are following to create them.

The reason why machines are now able to do things that we, their makers, do not fully understand is because they have become capable of learning from experience. AlphaGo became so good by playing more games of Go than a human could fit into a lifetime. Likewise, no human could read as many books as ChatGPT has absorbed.

Reducing anxiety

It's important to understand that machines have become intelligent without thinking in a human way. This realization alone can greatly reduce confusion, and therefore anxiety.

Intelligence is [not exclusively a human ability](#), as any biologist will tell you, and our specific brand of it is neither its pinnacle nor its

destination. It may be difficult to accept for some, but intelligence has more to do with chickens crossing the road safely than with writing poetry.

In other words, we should not necessarily expect machine intelligence to evolve towards some form of consciousness. Intelligence is the ability to do the right thing in unfamiliar situations, and this can be found in machines, for example those that recommend a new book to a user.

If we want to understand how to handle AI, we can return to a crisis that hit the industry from the late 1980s, when many researchers were still trying to mimic what we thought humans do. For example, they were trying to understand the rules of language or human reasoning, to program them into machines.

[That didn't work](#), so they ended up taking some shortcuts. This move might well turn out to be one of the most consequential decisions in our history.

Fork in the road

The first shortcut was to rely on making decisions based on statistical patterns found in data. This removed the need to actually understand the complex phenomena that we wanted the machines to emulate, such as language. The [auto-complete](#) feature in your messaging app can guess the next word without understanding your goals.

While others had similar ideas before, the first to make this method really work, and stick, was probably [Fredrick Jelinek at IBM](#), who invented "statistical language models", the ancestors of all GPTs, while working on machine translation.

In the early 1990s, [he summed up that first shortcut](#) by quipping:

"Whenever I fire a linguist, our systems performance goes up. Though the comment may have been said jokingly, it reflected a real-world shift in the focus of AI away from attempts to emulate the rules of language.

This approach rapidly spread to other domains, introducing a new problem: sourcing the data necessary to [train statistical algorithms](#).

Creating the data specifically for training tasks would have been expensive. A second shortcut became necessary: data could be harvested from the web instead.

As for knowing the intent of users, such as in [content recommendation systems](#), a third shortcut was found: to constantly observe users' behavior and infer from it what they might click on.

By the end of this process, AI was transformed and a new recipe was born. Today, this method is found in all online translation, recommendations and question-answering tools.

Fuel to operate

For all its success, this recipe [also creates problems](#). How can we be sure that [important decisions](#) are made fairly, when we cannot inspect the machine's inner workings?

How can we stop machines from amassing our personal data, when this is the very fuel that makes them operate? How can a machine be expected to stop harmful content from reaching users, when it is designed to learn what makes people click?

It doesn't help that we have deployed all this in a very influential position at the very center of our digital infrastructure, and have delegated many important decisions to AI.

For instance, algorithms, rather than human decision makers, dictate what we're shown on social media in real time. In 2022, the coroner who ruled on the tragic death of 14-year-old Molly Russell partly blamed an algorithm for [showing harmful material to the child](#) without being asked to.

As these concerns derive from the same shortcuts that made the technology possible, it will be challenging to find good solutions. This is also why the initial decisions of the [Italian privacy authority to block ChatGPT](#) created alarm.

Initially, the authority raised the issues of [personal data](#) being gathered from the web without a legal basis, and of the information provided by the chatbot containing errors. This could have represented a serious challenge to the entire approach, and the fact that it was solved by adding legal disclaimers, or changing the terms and conditions, might be a preview of future regulatory struggles.

We need good laws, not doomsaying. The paradigm of AI shifted long ago, but it was not followed by a corresponding shift in our legislation and culture. That time has now come.

An important conversation has started about what we should want from AI, and this will require the involvement of different types of scholars. Hopefully, it will be based on the technical reality of what we have built, and why, rather than on sci-fi fantasies or doomsday scenarios.

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

Provided by The Conversation

Citation: Opinion: To understand AI's problems, look at the shortcuts taken to create it (2023, May 12) retrieved 10 December 2023 from <https://techxplore.com/news/2023-05-opinion-ai-problems-shortcuts.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.