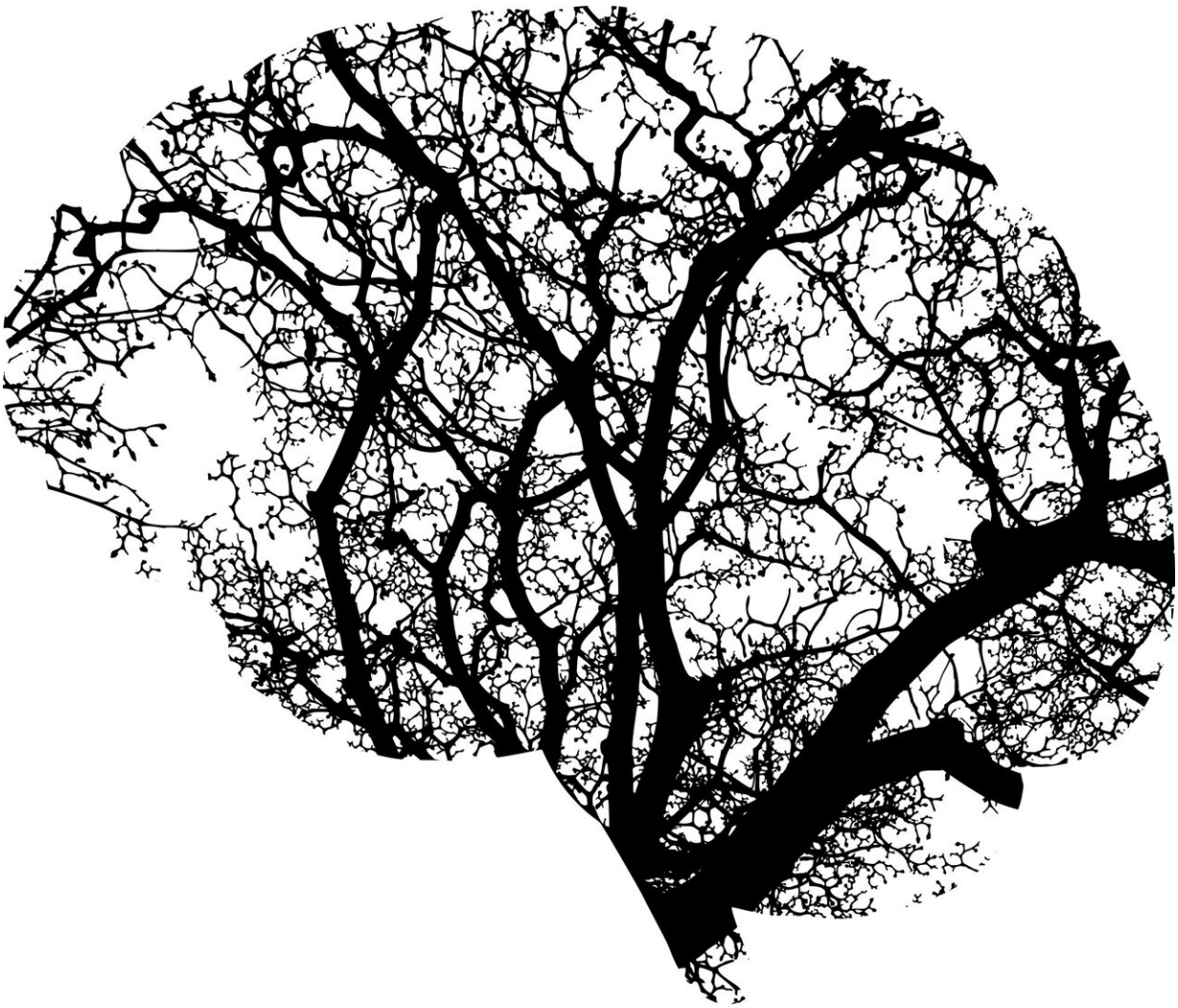
Reining in AI means figuring out which regulation options are feasible, both technically and economically

January 18 2024, by Saurabh Bagchi



Credit: CC0 Public Domain

Concern about generative artificial intelligence technologies seems to be growing almost as fast as the spread of the technologies themselves. These worries are driven by unease about the possible spread of disinformation at a scale never seen before, and fears of loss of employment, loss of control over creative works, and, more futuristically, AI becoming so powerful that it causes extinction of the human species.

The concerns have given rise to calls for regulating AI technologies. Some governments, for example the European Union, have responded to their citizens' push for regulation, while some, such as the U.K. and India, are taking a more laissez-faire approach.

In the U.S., the White House issued an executive order on Oct. 30, 2023, titled Safe, Secure, and Trustworthy Artificial Intelligence. It sets out guidelines to reduce both immediate and long-term risks from AI technologies. For example, it asks AI vendors to share safety test results with the federal government and calls for Congress to enact consumer privacy legislation in the face of AI technologies soaking up as much data as they can get.

In light of the drive to regulate AI, it is important to consider which approaches to regulation are feasible. There are two aspects to this question: what is technologically feasible today and what is economically feasible. It's also important to look at both the training data that goes into an AI model and the model's output.

1. Honor copyright

One approach to regulating AI is to limit the training data to public domain material and copyrighted material that the AI company has secured permission to use. An AI company can decide precisely what data samples it uses for training and can use only permitted material.

This is technologically feasible.

It is partially economically feasible. The quality of the content that AI generates depends on the amount and richness of the training data. So it is economically advantageous for an AI vendor to not have to limit itself to content it's received permission to use. Nevertheless, today some companies in generative AI are proclaiming as a sellable feature that they are only using content they have permission to use. One example is Adobe with its Firefly image generator.

2. Attribute output to a training data creator

Attributing the output of AI technology to a specific creator—artist, singer, writer and so on—or group of creators so they can be compensated is another potential means of regulating generative AI. However, the complexity of the AI algorithms used makes it impossible to say which input samples the output is based on. Even if that were possible, it would be impossible to determine the extent each input sample contributed to the output.

Attribution is an important issue because it's likely to determine whether creators or the license holders of their creations will embrace or fight AI technology. The 148-day Hollywood screenwriters' strike and the resultant concessions they won as protections from AI showcase this issue.

In my view, this type of regulation, which is at the output end of AI, is technologically not feasible.

3. Distinguish human- from AI-generated content

An immediate worry with AI technologies is that they will unleash automatically generated disinformation campaigns. This has already

happened to various extents—for example, disinformation campaigns during the Ukraine-Russia war. This is an important concern for democracy, which relies on a public informed through reliable news sources.

There is a lot of activity in the startup space aimed at developing technology that can tell AI-generated content from human-generated content, but so far, this technology is lagging behind generative AI technology. The current approach focuses on identifying the patterns of generative AI, which is almost by definition fighting a losing battle.

This approach to regulating AI, which is also at the output end, is technologically not currently feasible, though rapid progress on this front is likely.

4. Attribute output to an AI firm

It is possible to attribute AI-generated content as coming from a specific AI vendor's technology. This can be done through the well-understood and mature technology of cryptographic signatures. AI vendors could cryptographically sign all output from their systems, and anyone could verify those signatures.

This technology is already embedded in basic computational infrastructure—for example, when a web browser verifies a website you are connecting to. Therefore, AI companies could easily deploy it. It's a different question whether it's desirable to rely on AI-generated content from only a handful of big, well-established vendors whose signatures can be verified.

So this form of regulation is both technologically and economically feasible. The regulation is geared toward the output end of AI tools.

It will be important for policymakers to understand the possible costs and benefits of each form of regulation. But first they'll need to understand which of these is technologically and economically feasible.

Provided by The Conversation