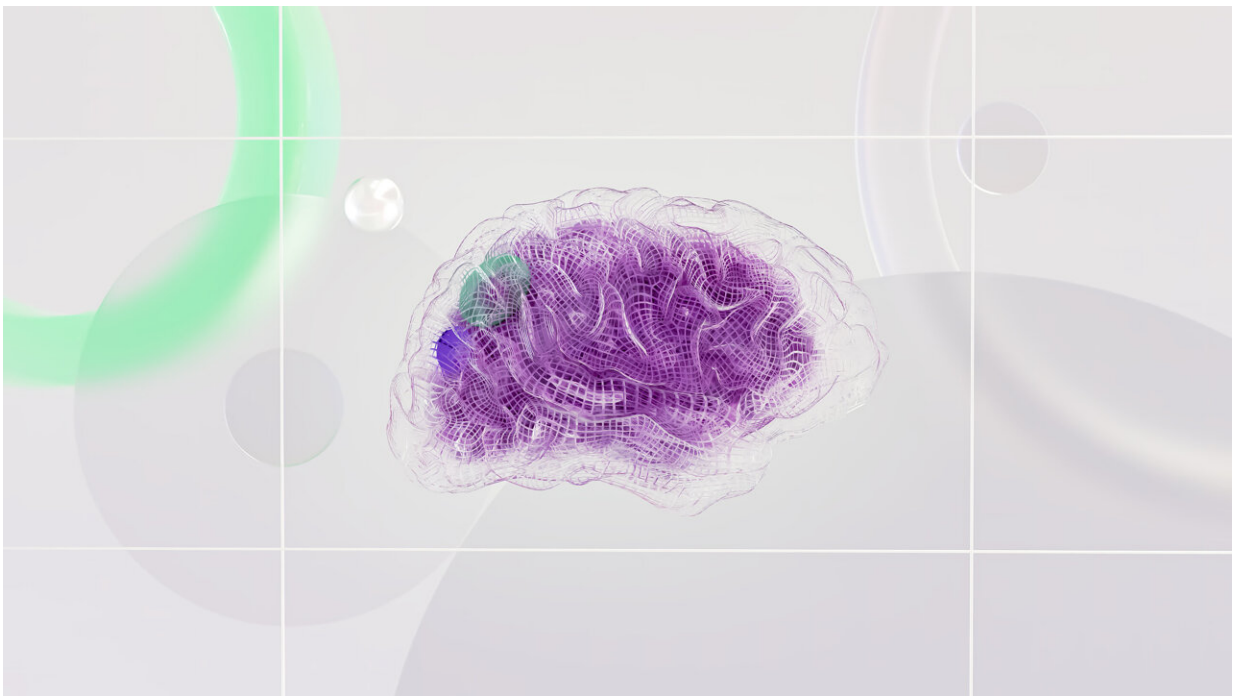


Irony machine: Why are AI researchers teaching computers to recognize irony?

July 25 2022, by Charles Barbour



Credit: Google DeepMind from Pexels

What was your first reaction when you heard about Blake Lemoine, the Google engineer who announced last month the AI program he was working on had developed consciousness?

If, like me, you're instinctively suspicious, it might have been something like: Is this guy serious? Does he honestly believe what he is saying? Or

is this an elaborate hoax?

Put the answers to those questions to one side. Focus instead on the questions themselves. Is it not true that even to ask them is to presuppose something crucial about Blake Lemoine: specifically, he is conscious?

In other words, we can all imagine Blake Lemoine being deceptive.

And we can do so because we assume there is a difference between his inward convictions—what he genuinely believes—and his outward expressions: what he *claims* to believe.

Isn't that difference the mark of consciousness? Would we ever assume the same about a [computer](#)?

Consciousness: 'The hard problem'

It is not for nothing philosophers have taken to calling consciousness "[the hard problem](#)." It is notoriously difficult to define.

But for the moment, let's say a conscious being is one capable of having a thought and not divulging it.

This means consciousness would be the prerequisite for [irony](#), or saying one thing while meaning the opposite. I know you are being ironic when I realize your words *don't* correspond with your thoughts.

That most of us have this capacity—and most of us routinely convey our unspoken meanings in this manner—is something that, I think, should surprise us more often than it does.

It seems almost discretely human.

Animals can certainly be funny—but not deliberately so.

What about machines? Can they deceive? Can they keep secrets? Can they be ironic?

AI and irony

It is a truth universally acknowledged (among academics at least) that any research question you might cook up with the letters "AI" in it is already being studied somewhere by an army of obscenely well-resourced computational scientists—often, if not always, funded by the U.S. military.

This is certainly the case with the question of AI and irony, which has recently attracted a significant amount of [research interest](#).

Of course, given that irony involves saying one thing while meaning the opposite, creating a machine that can detect it, let alone generate it, is no simple task.

But if we could create such a machine, it would have a multitude of practical applications, some more sinister than others.

In the age of online reviews, for example, retailers have become very keen on [so-called](#) "opinion mining" and "sentiment analysis," which uses AI to map not merely the content, but the mood of reviewer's comments.

Knowing whether your product is being praised or becoming the butt of the joke is valuable information.

Or consider content moderation on social media. If we want to limit online abuse while protecting freedom of speech, would it not be helpful to know when someone is serious and when they are joking?

Or what if someone tweets that they have just joined their local terrorist cell or they're packing a bomb in their suitcase and heading for the airport? (Don't ever tweet that, by the way.) Imagine if we could determine instantly whether they are serious, or whether they are just "being ironic."

In fact, given irony's proximity to lying, it's not hard to imagine how the entire shadowy machinery of governmental and corporate surveillance that has grown up around new communications technologies would find the prospect of an irony-detector extremely interesting.

And that goes a long way towards explaining the growing literature on the topic.

AI, from Clippy to facial recognition

To understand the state of current research into AI and irony, it is helpful to know a little about [the history of AI](#) more generally.

That history is typically broken down into two periods.

Until the 1990s, researchers sought to program computers with a set of handcrafted formal rules for how to behave in predefined situations.

If you used Microsoft Word in the 1990s, you might remember the irritating office assistant Clippy, who was endlessly popping up to offer unwanted advice.

Since the turn of the century, that model has been replaced by data-driven machine learning and [neural networks](#).

Here, enormous caches of examples of a given phenomena are translated into numerical values, on which computers can perform complex

mathematical operations to determine patterns no human could ever discover.

Moreover, the computer does not merely apply a rule. Rather, it learns from experience, and develops new operations independent of human intervention.

The difference between the two approaches is the difference between Clippy and, say, facial recognition technology.

Researching sarcasm

To build [a neural network with the ability to detect irony](#), researchers focus initially on what some would consider its simplest form: sarcasm.

The researchers begin with data stripped from social media.

For instance, they might collect all tweets labeled #sarcasm or Reddit posts labeled /s, a shorthand that Reddit users employ to indicate they are not serious.

The point is not to teach the computer to recognize the two separate meanings of any given sarcastic post. Indeed, meaning is of no relevance whatsoever.

Instead, the computer is instructed to search for recurring patterns, or what one researcher calls "syntactical fingerprints"—words, phrases, emojis, punctuation, errors, contexts, and so forth.

On top of that, the data set is bolstered by adding more streams of examples—other posts in the same threads, for instance, or from the same account.

Each new individual example is then run through a battery of calculations until we arrive at a single determination: sarcastic or not sarcastic.

Finally, a bot can be programmed to reply to each original poster and ask whether they were being sarcastic. Any response can be added to the computer's growing mountain of experience.

The [success rate](#) of the most recent sarcasm detectors approaches [an astonishing 90%](#)—greater, I suspect, than many humans could achieve.

So, assuming AI will continue to advance at the rate that took us from Clippy to facial recognition technology in less than two decades, can ironic androids be far off?

What is irony?

But isn't there a qualitative difference between sorting through the "syntactical fingerprints" of irony and actually understanding it?

Some would suggest [not](#). If a computer can be taught to behave exactly like a human, then it's immaterial whether a rich internal world of meaning lurks beneath its behavior.

But irony is arguably a unique case: it relies on the distinction between external behaviors and internal beliefs.

Here it might be worth remembering that, while computational scientists have only recently become interested in irony, philosophers and literary critics have been thinking about it for a very long time.

And perhaps exploring that tradition would shed old light, as it were, on a new problem.

Of the many names one could invoke in this context, two are indispensable: the German Romantic philosopher [Friedrich Schlegel](#); and the post-structuralist literary theorist [Paul de Man](#).

For Schlegel, irony does not simply entail a false, external meaning and a true, internal one. Rather, in irony, two opposite meanings are presented as equally true. And the resulting indeterminacy has devastating implications for logic, most notably the law of non-contradiction, which holds that a statement cannot be simultaneously true and false.

De Man follows Schlegel on this score, and in a sense, universalizes his insight. He [notes](#) every effort to define a concept of irony is bound to be infected by the phenomena it purports to explain.

Indeed, de Man believes all language is infected by irony, and involves what he calls "permanent parabasis." Because humans have the power to conceal their thoughts from one another, it will always be possible—permanently possible—that they do not mean what they are saying.

Irony, in other words, is not one kind of language among many. It structures—or better, haunts—every use of language and every interaction.

And in this sense, it exceeds the order of proof and computation. The question is whether the same is true of human beings in general.

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

Provided by The Conversation

Citation: Irony machine: Why are AI researchers teaching computers to recognize irony? (2022, July 25) retrieved 14 August 2024 from <https://techxplore.com/news/2022-07-irony-machine-ai.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.