SAINTS PERSPECTIVES

Things Futuristic

Talking about AI in human terms is natural—but wrong

When it comes to artificial intelligence, metaphors are often misleading

The Economist, June 22nd 2023

My love’s like a red, red rose. It is the east, and Juliet is the sun. Life is a highway, I wanna ride it all night long. **Metaphor** is a powerful and wonderful tool. Explaining one thing in terms of another can be both illuminating and pleasurable, if the metaphor is apt.

But that “if” is important. Metaphors can be particularly helpful in explaining unfamiliar concepts: imagining the Einsteinian model of gravity (heavy objects distort space-time) as something like a bowling ball on a trampoline, for example. But metaphors can also be misleading: picturing the atom as a solar system helps young students of chemistry, but the more advanced learn that electrons move in clouds of probability, not in neat orbits as planets do.

What may be an even more misleading metaphor—for artificial intelligence (AI)—seems to be taking hold. AI systems can now perform staggeringly impressive tasks, and their ability to reproduce what seems like the most human function of all, namely language, has ever more observers writing about them. When they do, they are tempted by an obvious (but obviously wrong) metaphor, which portrays AI programmes as **conscious** and even intentional agents. After all, the only other creatures which can use language are other conscious agents—that is, humans.

Take the well-known problem of factual mistakes in potted biographies, the likes of which **Chatgpt** and other large language models (LLMS) churn out in seconds. Incorrect birthplaces, non-existent career moves, books never written: one journalist at The Economist was alarmed to learn that he had recently died. In the jargon of AI engineers, these are “hallucinations”. In the parlance of critics, they are “lies”.

“Hallucinations” might be thought of as a forgiving euphemism. Your friendly local AI is just having a bit of a bad trip; leave him to sleep it off and he’ll be back to himself in no time. For the “lies” crowd, though, the humanising metaphor is even more profound: the AI is not only thinking, but has desires and intentions. A lie, remember, is not any old false statement. It is one made with the goal of deceiving others. **Chatgpt** has no such goals at all.

Humans' tendency to **anthropomorphise**[1] things they don't understand is ancient, and may confer an evolutionary advantage. If, on spying a rustling in the bushes, you infer an agent (whether predator or spirit), no harm is done if you are wrong. If you assume there is nothing in the undergrowth and a leopard jumps out, you are in trouble. The all-too-human desire to smack or yell at a malfunctioning device comes from this ingrained instinct to see intentionality everywhere.

It is an instinct, however, that should be overridden when writing about AI. These systems, including those that seem to converse, merely take input and produce output. At their most basic level, they do nothing more than turn strings like **0010010101001010** into **1011100100100001** based on a set of instructions. Other parts of the software turn those 0s and 1s into words, giving a frightening—but false—sense that there is a ghost in the machine.

Whether they can be said to "think" is a matter of **philosophy** and **cognitive science**, since plenty of serious people see the brain as a kind of computer. But it is safer to call what LLMS do "**pseudo-cognition**". Even if it is hard on the face of it to distinguish the output from human activity, they are fundamentally different under the surface. Most importantly, cognition is not intention. Computers do not have desires.

It can be tough to write about machines without metaphors. People say **a watch "tells" the time**, or that a credit-card reader which is working slowly is "thinking" while they wait awkwardly at the checkout. Even when machines are said to "generate" output, that cold-seeming word comes from an ancient root meaning to give birth.

But AI is too important for loose language. If entirely avoiding human-like metaphors is all but impossible, writers should offset them, early, with some suitably bloodless phrasing. "**An LLM is designed to produce text that reflects patterns found in its vast training data**," or some such explanation, will help readers take any later imagery with due scepticism. Humans have

[1] attribute human characteristics or behavior to (a god, animal, or object).

evolved to spot ghosts in machines. Writers should avoid ushering them into that trap. Better to lead them out of it. ■

Related Articles

- **Large, Creative AI Models will Transform Lives and Labour Markets**