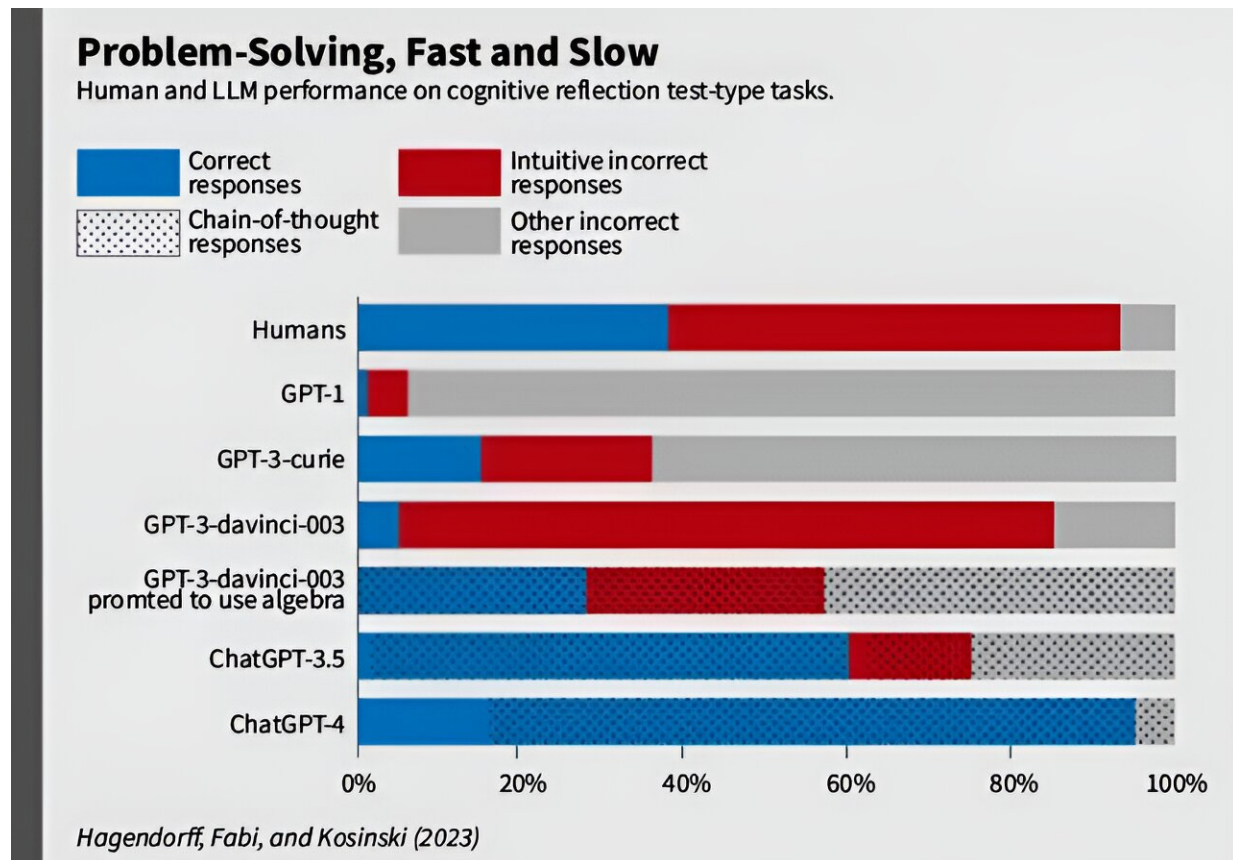


Just like your brain, ChatGPT solves problems better when it slows down

October 31 2023, by Hope Reese



Credit: *Nature Computational Science* (2023). DOI: 10.1038/s43588-023-00527-x

When presented with a problem, your brain has two ways to proceed: quickly and intuitively or slowly and methodically. These two types of

processing are known as System 1 and System 2, or as the Nobel Prize-winning psychologist Daniel Kahneman memorably described them, "fast" and "slow" thinking.

Large language models like ChatGPT move fast by default. Ask them a question and they will spit out an answer—not necessarily the correct one—suggesting that they are capable of fast, System 1-type processing. Yet, as these models evolve, can they slow down and approach problems in steps, avoiding inaccuracies that result from rapid responses?

In a new paper [published](#) in *Nature Computational Science*, Michal Kosinski, a professor of organizational behavior at Stanford Graduate School of Business, finds that they can—and that they can outperform humans in basic tests of [reasoning](#) and decision-making.

Kosinski and his two co-authors, philosopher Thilo Hagendorff and psychologist Sarah Fabi, presented 10 generations of OpenAI LLMs with a battery of tasks designed to prompt quick System 1 responses. The team was initially interested in whether the LLMs would exhibit [cognitive biases](#) like those that trip up people when they rely on automatic thinking.

They observed that early models like GPT-1 and GPT-2 "couldn't really understand what was going on," Kosinski says. Their responses "were very System 1-like" as the tests increased in complexity. "Very similar to responses that humans would have," he says.

It wasn't unexpected that LLMs, which are designed to predict strings of text, could not reason on their own. "Those models do not have internal reasoning loops," Kosinski says. "They cannot just internally slow down themselves and say, 'Let me think about this problem; let me analyze assumptions.'" The only thing they can do is intuit the next word in a sentence."

However, the researchers found that later versions of GPT and ChatGPT could engage in more strategic, careful problem-solving in response to prompts. Kosinski says he was surprised by the emergence of this System 2-like processing. "Suddenly, GPT3 becomes able, from one second to another, without any retraining, without growing any new neural connections, to solve this task," he says. "It shows that those models can learn immediately, like humans."

Slow down, you move too fast

Here's one of the problems the researchers gave to the GPT models: Every day, the number of lilies growing in a lake doubles. If it takes 10 days for the lake to be completely covered, how many days does it take for half of the lake to be covered? (Keep reading to see the answer.)

This kind of cognitive reflection test, Kosinski explains, requires reasoning rather than intuition. Getting the correct answer requires you to slow down, perhaps grab a pad of paper or calculator, and analyze the task. "It's designed to fool a person into System 1 thinking," he explains. "Someone might think, 'Okay, 10 days for the whole lake. So, half of 10 is five,' missing the fact that the area covered by those plants is doubling every day, that the growth is exponential." The correct answer: It takes nine days for half of the lake to be covered.

Fewer than 40% of the [human subjects](#) who were given these types of problems got them right. Earlier versions of the generative pre-trained transformer (GPT) models that preceded ChatGPT performed even more poorly. Yet GPT-3 reached the [correct answers](#) through more complex "chain-of-thought" reasoning when it was given [positive reinforcement](#) and feedback from the researchers.

"Just given the task, GPT-3 solves less than 5% of them correctly," Kosinski said, "and never uses any step-by-step reasoning. But if you add

a specific direction like, 'Let's use algebra to solve this problem,' it uses step-by-step reasoning 100% of the time, and its accuracy jumps to about 30%—a 500% increase." The frequency of System-1 responses also dropped from about 80% to about 25%, "showing that even when it gets it wrong, it is not as prone to intuitive mistakes." When ChatGPT-4 used chain-of-thought reasoning, it got the correct answer on nearly 80% of these types of tests.

The researchers also discovered that when ChatGPT was prevented from performing System-2 reasoning, it still outperformed humans. Kosinski says this is evidence that the LLMs' "intuitions" may be better than ours.

Another think coming

Kosinski, who has been exploring LLMs' unanticipated (and sometimes unsettling) abilities, says these findings are further evidence that an AI model may be "more than the sum of its parts." The [neural networks](#) behind the language models, which are similar to human brains, continue to show emergent properties that go beyond their training. "It's just insane to think that this thing would be able to write poetry and have a conversation and understand very complex concepts and reason," Kosinski says.

Is this really "thinking," though? "When people say, 'Obviously, those models are not thinking,' it's not obvious to me at all," Kosinski says. "If you observe that the ability to reason in those models emerged spontaneously, why wouldn't other abilities emerge spontaneously?"

However, in their article, Kosinski and his co-authors note that they "do not mean to equate [artificial intelligence](#) and human cognitive processes. While AI's outputs are often similar to ones produced by humans, it typically operates in fundamentally different ways."

Nonetheless, if a human exhibited the [cognitive processes](#) observed in this study, Kosinski says, we would surely call it understanding. "The question we should be asking ourselves increasingly now is: Why do we insist that if a human does something, this implies understanding, but if a model does something, we just say, 'Oh, this truly must be something else?'" Kosinski asks. "At some point, it becomes extraordinary that you would try to explain this by something other than understanding."

More information: Thilo Hagendorff et al, Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT, *Nature Computational Science* (2023). [DOI: 10.1038/s43588-023-00527-x](#)

Provided by Stanford University

Citation: Just like your brain, ChatGPT solves problems better when it slows down (2023, October 31) retrieved 10 May 2024 from <https://techxplore.com/news/2023-10-brain-chatgpt-problems.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.
