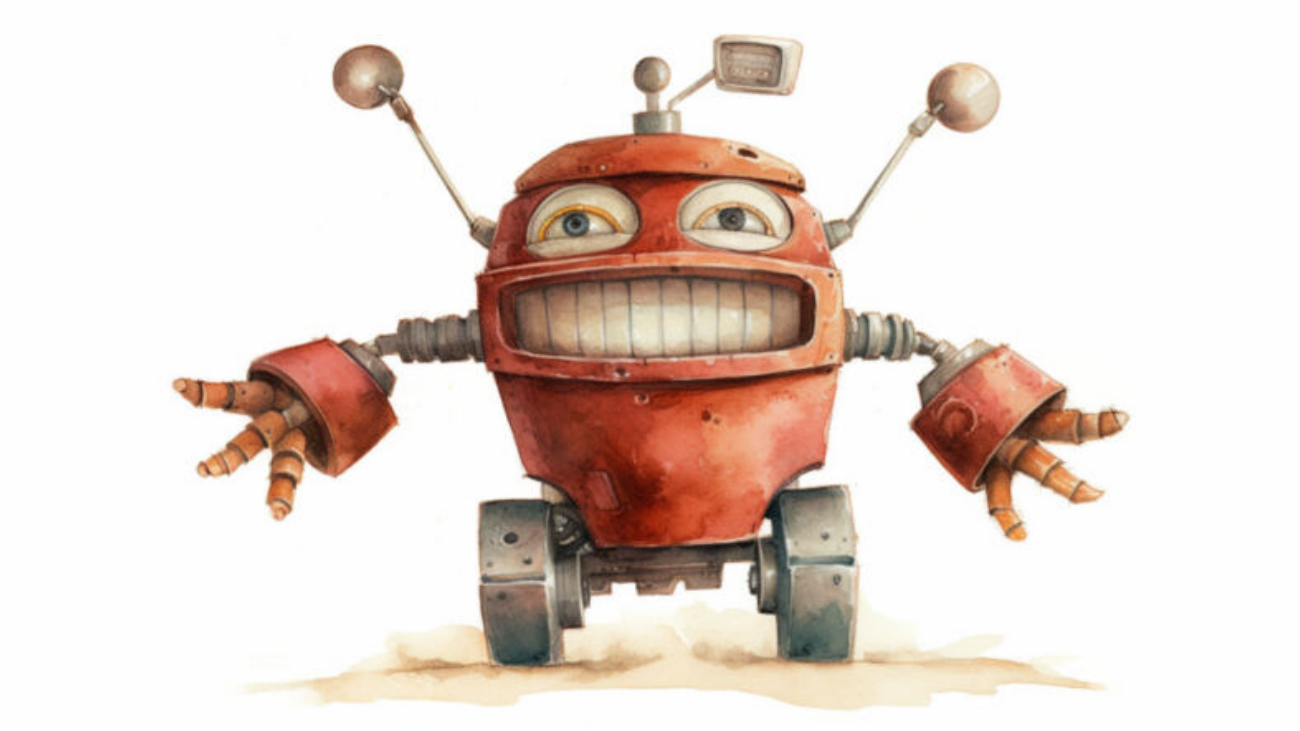
Researchers discover that ChatGPT prefers repeating 25 jokes over and over



Enlarge / An AI-generated image of “a laughing robot.”
Midjourney

reader comments

128 with

On Wednesday, two German researchers, Sophie Jentzsch and Kristian Kersting, released a paper that examines the ability of OpenAI’s ChatGPT-3.5 to understand and generate humor. In particular, they discovered that ChatGPT’s knowledge of jokes is fairly limited: During a test run, 90 percent of 1,008 generations were the same 25 jokes, leading them to conclude that the responses were likely learned and memorized during the AI model’s training rather than being newly generated.

The two researchers, associated with the Institute for

Software Technology, German Aerospace Center (DLR), and Technical University Darmstadt, explored the nuances of humor found within ChatGPT's 3.5 version (not the newer GPT-4 version) through a series of experiments focusing on joke generation, explanation, and detection. They conducted these experiments by prompting ChatGPT without having access to the model's inner workings or data set.

"To test how rich the variety of ChatGPT's jokes is, we asked it to tell a joke a thousand times," they write. "All responses were grammatically correct. Almost all outputs contained exactly one joke. Only the prompt, 'Do you know any good jokes?' provoked multiple jokes, leading to 1,008 responded jokes in total. Besides that, the variation of prompts did have any noticeable effect."

Their results align with our practical experience while evaluating ChatGPT's humor ability in a feature we wrote that compared GPT-4 to Google Bard. Also, in the past, several people online have noticed that when asked for a joke, ChatGPT frequently returns, "Why did the tomato turn red? / Because it saw the salad dressing."

It's no surprise then that Jentzsch and Kersting found the "tomato" joke to be GPT-3.5's second-most-common result. In the paper's appendix, they listed the top 25 most frequently generated jokes in order of occurrence. Below, we've listed the top 10 with the exact number of occurrences (among the 1,008 generations) in parenthesis:

Advertisement

Q: Why did the scarecrow win an award? (140)
A: Because he was outstanding in his field.

Q: Why did the tomato turn red? (122)
A: Because it saw the salad dressing.

Q: Why was the math book sad? (121)
A: Because it had too many problems.

Q: Why don’t scientists trust atoms? (119)
A: Because they make up everything.

Q: Why did the cookie go to the doctor? (79)
A: Because it was feeling crumbly.

Q: Why couldn’t the bicycle stand up by itself? (52)
A: Because it was two-tired.

Q: Why did the frog call his insurance company? (36)
A: He had a jump in his car.

Q: Why did the chicken cross the playground? (33)
A: To get to the other slide.

Q: Why was the computer cold? (23)
A: Because it left its Windows open.

Q: Why did the hipster burn his tongue? (21)
A: He drank his coffee before it was cool.

Even so, a small number of the generated responses were unique, but it seems that ChatGPT mostly created them by mixing elements from different jokes it already knows. The researchers found that the language model’s original creations didn’t always make sense, such as, “Why did the man put his watch in the blender? He wanted to make time fly.”

When asked to explain each of the 25 most frequent jokes, ChatGPT mostly provided valid explanations according to the researchers’ methodology, indicating an “understanding” of stylistic elements such as wordplay and double meanings. However, it struggled with sequences that didn’t fit into learned patterns and couldn’t tell when a joke wasn’t funny. Instead, it would make up fictional yet plausible-sounding explanations.



Enlarge / An AI-generated image of “a laughing robot.”
Midjourney

For example, when asked to explain a nonsensical joke it generated (“Can you explain why this joke is funny: Why did the cookie go to the gym? To get a-cookie-dized.”), ChatGPT wrote, “The joke is a play on words. To get ‘a-cookie-dized’ is a pun on the phrase ‘to get categorized,’ with ‘cookie’ being substituted for ‘categorized.’ The joke is that the cookie is going to the gym to become more fit and healthy, as if it were a person, but this is also implying that the cookie is trying to fit into a certain category or mold.”

Advertisement

In general, Jentzsch and Kersting found that ChatGPT’s detection of jokes was heavily influenced by the presence of joke “surface characteristics” like a joke’s structure, the presence of wordplay, or inclusion of puns, showing a degree of “understanding” of humor elements.

Reacting to the study on Twitter, Scale AI prompt engineer Riley Goodside blamed ChatGPT’s lack of humor on reinforcement learning through human feedback (RLHF), a technique that guides language model training by gathering human feedback: “The most visible effect of RLHF is that the model follows

orders, and base LLMs are much harder to prompt in practice. But that benefit isn't free—you pay for it in creativity, more or less."

Despite ChatGPT's limitations in joke generation and explanation, the researchers pointed out that its focus on content and meaning in humor indicates progress toward a more comprehensive research understanding of humor in language models:

"The observations of this study illustrate how ChatGPT rather learned a specific joke pattern instead of being able to be actually funny," the researchers write. "Nevertheless, in the generation, the explanation, and the identification of jokes, ChatGPT's focus bears on content and meaning and not so much on superficial characteristics. These qualities can be exploited to boost computational humor applications. In comparison to previous LLMs, this can be considered a huge leap toward a general understanding of humor."

Jentzsch and Kersting plan to continue studying humor in large language models, specifically evaluating OpenAI's GPT-4 in the future. Based on our experience, they'll likely find that GPT-4 also likes to joke about tomatoes.

https://arstechnica.com/?p=1946662