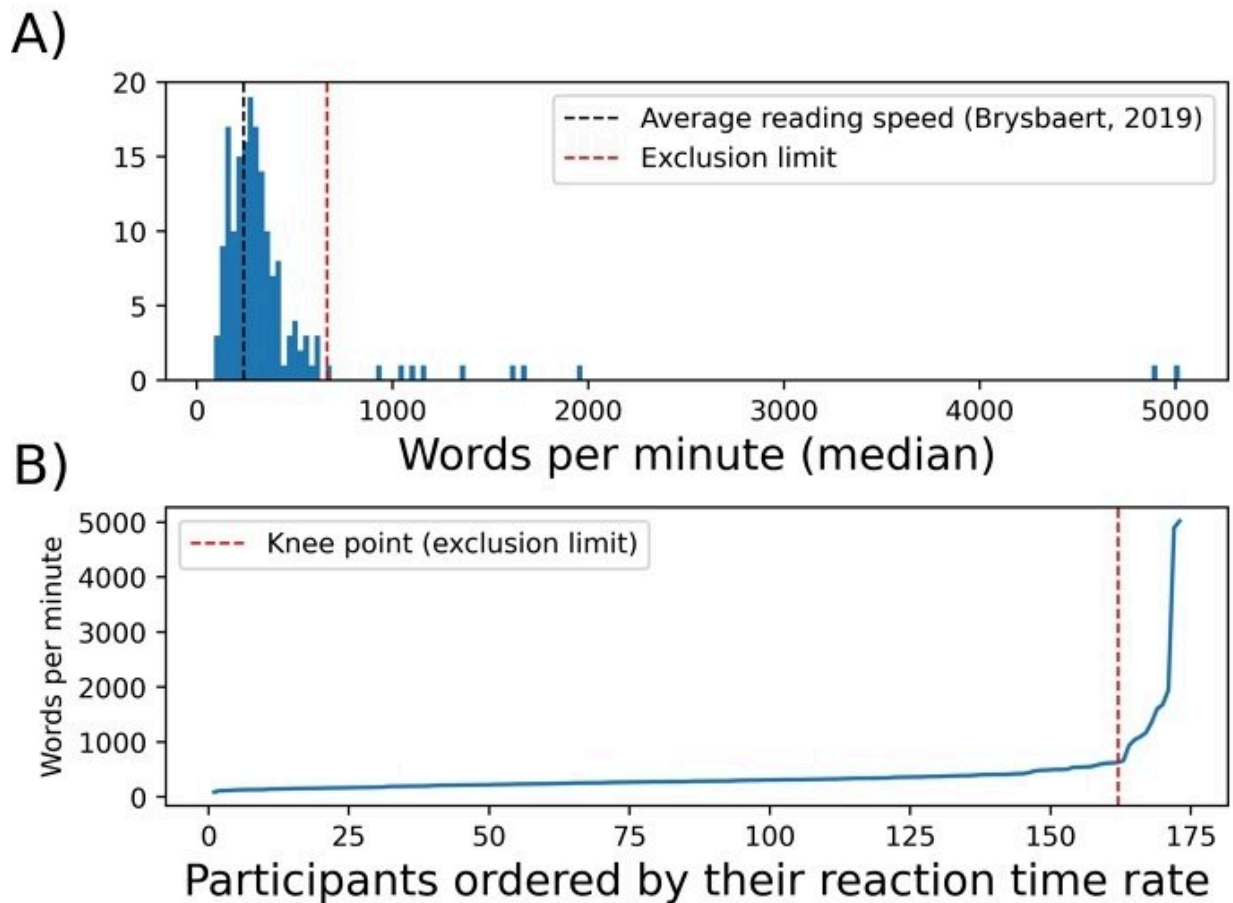# Chat AIs can role-play humans in surveys and pilot studies

May 9 2023

A) Shows the participants' median reaction times (see main text for explanation) to the 20 stimuli as a histogram. Participants with reaction times faster than the exclusion limit were excluded from the final analyses. B) The reaction times plotted from slowest to fastest. The exclusion limit is at the knee point of the curve. Credit: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). DOI: 10.1145/3544548.3580688

Studying people in human-computer interaction (HCI) research can be slow. That's why researchers at the Finnish Center for Artificial Intelligence (FCAI) recently harnessed the power of large language models (LLMs), specifically GPT-3, to generate open-ended answers to questions about video game player experience.

These AI-generated responses were often more convincing, as rated by humans, than real responses. These synthetic interviews offer a new approach to gathering data quickly and at low cost, which may help in fast iteration and initial testing of study designs and data analysis pipelines. Any findings based on AI-generated data, however, should also be confirmed with real data.

The researchers, based at Aalto University and the University of Helsinki, discovered some subtle differences in different versions of GPT-3 that affected the diversity of AI-generated responses. But a more discouraging implication is that data from popular crowdsourcing platforms may now automatically be suspect, as AI-generated responses are hard to distinguish from real ones.

Amazon's Mechanical Turk (MTurk), for example, can host surveys or research tasks for HCI, psychology, or related scientific areas and pay users for participation, but "now that LLMs are so easy to access, any self-reported data from the internet cannot be trusted anymore. The [economic incentives](#) can drive malicious users to employ bots and LLMs to generate high-quality fake responses," says Aalto University associate professor Perttu Hämäläinen.

The implications of synthetic data for anonymity, privacy and [data protection](#) in the medical field and similar domains are clear. However, in the realm of HCI, or science more widely, synthetic interviews and

artificial experiments raise questions about the trustworthiness of crowdsourcing approaches that seeks to gather user data online.

"LLMs cannot and should not replace real participants, but synthetic data may be useful for initial exploration and piloting of research ideas," suggests Hämäläinen. "It may be time to abandon platforms like Mturk for collecting real data and go back to lab studies."

The research was published in the *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

Provided by Aalto University