Study says AI data contaminates vital human input

June 20 2023, by Peter Grad

Credit: Unsplash/CC0 Public Domain

At the turn of this century, Jeff Bezos popularized the use of mechanical turks—low-paid workers working remotely with perhaps thousands of others on tiny parts of larger computer projects—to ensure a human

perspective on mostly simple tasks that proved perplexing to computers. He termed this blending of human and digital brain power "artificial artificial intelligence."

About a quarter million people are employed through Amazon's Mechanical Turk marketplace, just one of many sources providing such services.

This week, researchers at Swiss-based university EPFL reported that turks who had provided important human input are now relying on AI-generated content to complete their tasks. They dubbed this phenomenon "artificial artificial artificial intelligence."

The term may evoke smiles, but the researchers say the findings raise serious concerns.

Workers tapping into AI generators to fulfill their tasks "would severely diminish the utility of crowdsourced data," researcher Veniamin Veselovsky said. The paper, "Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks," was published on the *arXiv* pre-print server June 13.

While large language models excel at processing training data, human input is still superior for certain tasks. Humans label data entered into models, describe images and respond to CAPTCHA screens more efficiently than computers can.

"It is tempting to rely on crowdsourcing to validate large language model outputs or to create human gold-standard data for comparison," Veselovsky said. "But what if crowd workers themselves are using LLMs … in order to increase their productivity, and thus their income, on crowdsourcing platforms?"

Such input would contaminate the data pool, and if left unaddressed, could throw the reliability of AI-based operations into question.

The term "turk" is derived from an 18th-century chess master "robot" that defeated players throughout Europe. Napoleon and Benjamin Franklin were among the defeated. The unsuspecting players never knew a human chess expert was hidden under the machine's planks.

Crowdsourcing with modern-day turks has become a billion-dollar industry. Its reputation has been tarnished over the notoriously low wages some companies pay their workers. Turks earn as little as $2 to $5 per hour.

But the industry is being threatened by the abrupt adoption of large language models. A ChatGPT 3.5 turbo model tackling classification assignments was found to perform significantly better than crowd workers at about one-twentieth the cost, according to a recent study.

Workers will face increased pressure to produce more and do it faster, and this in turn could lead to those workers relying more on AI resources.

Based on a limited study of the use of large language models by workers at MTurk, Amazon's crowd sourcing operation, the EPFL researchers estimated that 33% to 46% of worker assignments were completed with the aid of large language models.

"Large language models are becoming more popular by the day, and multimodal models, supporting not only text, but also image and video input and output, are on the rise," Veselovsky said. "With this, our results should be considered the 'canary in the coal mine' that should remind platforms, researchers and crowd workers to find new ways to ensure that human data remain human."

More information: Veniamin Veselovsky et al, Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks, *arXiv* (2023). DOI: 10.48550/arxiv.2306.07899