

Study shows ChatGPT writes better school essays than students

November 27 2023, by Kathrin Haimerl



Credit: Unsplash/CC0 Public Domain

[In a study](#) published in *Scientific Reports*, a research team from the University of Passau compared the quality of machine-generated content with essays written by secondary school students. The upshot: The AI-

based chatbot performed better across all criteria, especially when it came to language mastery.

The [language](#) model ChatGPT is making enormous progress. After version 3.5 had failed the Bavarian Abitur (a test given at the end of secondary school in Germany) in early 2023, its successor version 4 earned a solid 2 nearly six months later.

A study by the University of Passau has now been able to demonstrate to what extent AI-generated content could revolutionize the school system. The researchers also experimented with the two language model versions.

In a study entitled "A large-scale comparison of human-written versus ChatGPT-generated essays" and published in *Scientific Reports*, they concluded that the machine writes better English essays. They evaluated machine-generated texts and essays written by [secondary school students](#) according to guidelines established by the Ministry of Education of Lower Saxony.

"I was surprised by how clear the outcome was," says Professor Steffen Herbold, who holds the Chair of AI Engineering at the University of Passau and initiated the study. Both Open AI chatbot versions scored higher than the students, with GPT-3 ranking in the middle and GPT-4 achieving the best score. "This shows that schools shouldn't turn a blind eye to these new tools."

Reflecting on AI models

The [interdisciplinary study](#) was carried out by the [computer scientists](#) in collaboration with computer linguist Professor Annette Hautli-Janisz and computer science didactician Ute Heuer. "I find it important to prepare [teachers](#) for the challenges and opportunities coming their way as

artificial intelligence models become increasingly available," says Heuer.

She initiated a training course on "ChatGPT—Opportunity and Challenge" that the research team conducted. This event, which took place in March 2023, was attended by 139 teachers, most of whom teach at German gymnasiums. The teachers were first briefed on selected technological ideas behind general text generators and ChatGPT. The practical stage then specifically involved English-language texts where the training course participants were left unaware of the origin of these texts.

Using questionnaires, the teachers were asked to evaluate the essays presented to them based on grading scales established by the Ministry of Education of Lower Saxony. Content was evaluated based on the criteria topic, completeness, and logic as well as linguistic aspects like vocabulary, complexity, and language mastery. The research team from Passau defined a scale from 0 to 6 for each criterion, with 0 being the worst score and 6 the best.

The machine scores above average in language mastery

One hundred eleven teachers completed the entire questionnaire and evaluated a total of two hundred seventy English language essays. The research team found the biggest difference in language mastery where the machine scored 5.25 (GPT-4) and 5.03 points (GPT-3) respectively, whereas the students scored an average of 3.9 points.

"This does not mean that students have poor English language skills. Rather, the scores achieved by the machine are exceptionally high," underscores Annette Hautli-Janisz, Junior Professor of Computational Rhetoric and Natural Language Processing at the University of Passau.

For Hautli-Janisz, who analyzed the texts from a linguistic perspective together with doctoral student Zlata Kikteva, the study provides further exciting insights into the machine's language development. "We have seen how the models change over time and are able to demonstrate with our studies that they have improved in performing the task we give them."

The researchers have also been able to identify differences between human and machine-generated language. "When we read more AI-generated texts going forward, we'll have to ask ourselves whether and how that affects our [human language](#)," says Hautli-Janisz.

More information: Steffen Herbold et al, A large-scale comparison of human-written versus ChatGPT-generated essays, *Scientific Reports* (2023). [DOI: 10.1038/s41598-023-45644-9](https://doi.org/10.1038/s41598-023-45644-9)

Provided by Universität Passau

Citation: Study shows ChatGPT writes better school essays than students (2023, November 27) retrieved 9 May 2024 from <https://techxplore.com/news/2023-11-chatgpt-school-essays-students.html>

<p>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.</p>
--