New College of Florida Professor Bernhard Klingenberg Co-Authors Study: Can AI Replace the Statistician?



SARASOTA, FL, UNITED STATES, September 18, 2026 / EINPresswire.com/ -- As Artificial Intelligence (AI) and Large Language Models (LLMs) become deeply integrated into business and research workflows, a pressing question has emerged across industries: Can AI replace the trained statistician? According to a new study co-authored by Bernhard Klingenberg, Professor of Statistics and Director of the graduate program in Applied Data Science at New College of Florida, the answer is "not with the general-purpose frontier AI models we investigated. The accuracy of the output depends too much on how the prompt is worded rather than the data supplied".

The peer-reviewed study, titled "Can AI Replace the Statistician? Evidence From a Controlled Experiment," was published in the International Statistical Review. Partnering with University of South Florida researchers Wolfgang Jank, Daniel Zantedeschi, and Sonal Prabhune, the team evaluated the true statistical reasoning capabilities of five prominent AI models. The experiment tested the models across three distinct phrasing scenarios and four complex data tasks.

The prompt scenarios reflected how users with varying degrees of statistical knowledge would query AI in the absence of a trained statistician who would help with the translation of questions into more formal statistical language and also help evaluate the output of the AI model. If LLMs could reliably perform statistical reasoning, the implications for scientific research, industry analytics and statistical education would be profound.

Key Research Findings

Analyzing the data (without the help of AI) of a controlled experiment where the ground truth

was known to the researchers but withheld from the AI models, several critical insights regarding how AI processes data were obtained:

- Prompt Sensitivity Over Reasoning: The study revealed that AI models perform linguistic pattern matching rather than genuine statistical analysis. A model's accuracy can swing by 30 to 80 percentage points based entirely on how a question is worded, even when analyzing the exact same data.

- The "Confidence-Competence Inversion": Paradoxically, models often provide the most cautious, inaccurate answers to users who express vagueness in their prompt—the exact users who need AI assistance the most.

- The "Mirror Paradox": Surprisingly, using more precise statistical language in a prompt sometimes made things worse. GPT-4 performed better with vague, ill-formulated prompts and worse when using more appropriate statistical terminology, while GPT-4o showed the exact opposite pattern. This likely reflects differences in each model's architecture and training.

- Failure on Complex Tasks: The studied LLMs fail on tasks requiring quantitative precision, such as effect size estimation, scoring as low as 15% accuracy overall for some models.

- Value as Screening Tools: Despite limitations in rigorous computation, LLMs achieved 76% to 81% accuracy on basic pattern-recognition tasks, indicating they can be useful for initial data screening before a human expert takes over.

"Our research exposes a blind spot in how AI handles data," says Dr. Klingenberg. "People assume that if you hand an AI a dataset and ask a question, it will do the statistical analysis. “Instead,” adds Prof. Jank, “we found that AI behaves more like a linguist than a statistician. It gives drastically different answers based purely on how you phrase the prompt, rather than what the data actually say.”

One caveat of their study, both professors mention, is that AI models develop at a rapid pace and are adding more capabilities with every iteration, so the findings of the study will have to be reexamined

Master of Science in Applied Data Science at New College of Florida

The Master of Science in Applied Data Science program is explicitly designed to bridge the gap between technical theory and real-world application. Ranked among top 10 national programs by Fortune, the curriculum focuses on the four fundamental pillars of data science: Statistics, Machine Learning, Databases, and Computing. Every course in the 3 or 4 semester program incorporates an applied project, emphasizing not just technical execution but also the critical interpretation and clear communication of results.

About New College of Florida

Founded in 1960, New College of Florida is the state’s designated public honors college. Recognized nationally for its academic excellence, rigorous inquiry, and commitment to free expression, New College offers more than 50 undergraduate majors, graduate programs in Applied Data Science, Marine Mammal Science, and a new Master's program in Educational Leadership, and a growing NAIA athletics program on a 151-acre campus on Sarasota Bay.

James Miller
New College of Florida
+1 850-445-0773
email us here

This press release can be viewed online at: https://www.einpresswire.com/article/943349659

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.