# [Result, but poetic]: Does LLM Use Improve Data Science Education?*

## Evidence from a Canadian Undergraduate Statistics Course

Rohan Alexander Luca Carnegie

July 15, 2024

Blah Blah Blah Blah

## 1 Introduction

The rapid advancement of Large Language Models (LLMs) like ChatGPT has had massive effects in all areas of society, with notable increases in productivity within some knowledge-based industries. Recent papers have demonstrated evidence of some productivity increases, in industries ranging from consulting to computer programming.

On the lower end of productivity increases, Ben-Michael et al. (2024) conducted a randomized control trial with judges in Wisconsin to assess the impact of AI recommendations on bail hearing decisions. The study found that providing AI recommendations did not significantly improve judges' decision-making accuracy. On the other hand, Peng et al. (2023) investigated the significantly positive impact of GitHub Copilot, an LLM-powered coding assistant, on programmer productivity. They conducted an experiment with 95 professional programmers recruited through Upwork. The study found that programmers with access to Copilot completed a standardized programming task 55.8% faster than the control group. In particular, less experienced programmers appeared to benefit more from the tool.

Examining the middle-ground, a study by Dell'Acqua et al. (2023) examined the impact of AI on management consulting tasks through a field experiment with 758 Boston Consulting Group consultants. They found that AI significantly increased performance across various business tasks, with speed increasing by over 25% and quality by more than 40%. However, the study revealed a "jagged technological frontier" where AI excelled in some tasks but struggled with others. Notably, AI usage appeared to level performance differences across ability levels, with lower-performing consultants benefiting the most, similar to that of Peng et al.

*Code and data are available at: https://github.com/lcarnegie/llms-achievement.

We can see in the current literature that LLMs have had varying levels of impact across different knowledge-intensive industries, with more computer-intensive tasks like programming receiving the highest performance boosts. One area sorely lacking quantitative evidence of performance increases is within higher education, an area where scrutiny of the impact of LLMs and other AI tools has been common. A survey of undergraduate political science students was conducted by Cahill and McCabe (2024), revealing widespread use of AI tools like ChatGPT. However, their study also uncovered a significant gap in students' self-perceived competence in using these tools effectively for academic work.

Student challenges while learning Data Science in particular were explored by Kross and Guo (2019). By polling instructor-practitioners of Data Science, they found several key challenges that data science students face. These included diverse programming/computing backgrounds, the challenges of integrating various technical workflows to conduct authentic analysis, software setup, finding relevant datasets, as well as dealing with the inherent uncertainty of the data science process.

Given the rapid integration of AI tools in educational and professional settings, there is an urgent need to understand how these technologies affect student learning and performance, particularly the rapidly-growing field of data science. This paper aims to investigate the relationship between academic performance, LLM usage, and students' perceptions of AI, using evidence from an exit survey of a third-year undergraduate statistics course at the University of Toronto. By examining how students interact with and perceive AI tools and how this usage correlates with their grades, we seek to understand the potential benefits and pitfalls of AI integration in data science education.

The remainder of this paper is structured as follows: Section 2 analyzes our survey data; Section 3 models our unstructured data to find fun and cool things; Section 4 lists the results; Section 5 discusses the implications of these findings for data science education and for future research and practice in this rapidly evolving field.

## 2 Data

## 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

### 3.1 Model set-up

Define $y_i$ as the number of seconds that the plane remained aloft. Then $\beta_i$ is the wing width and $\gamma_i$ is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

#### 3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance $\theta$.

## 4 Results

Our results are summarized in **?@tbl-modelresults**.

## 5 Discussion

### 5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

## 5.2 Second discussion point

## 5.3 Third discussion point

## 5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

# Appendix

# A Additional data details

# B Model details

## B.1 Posterior predictive check

In ... we implement a posterior predictive check. This shows...

In ... we compare the posterior with the prior. This shows...

## B.2 Diagnostics

... is a trace plot. It shows... This suggests...

... is a Rhat plot. It shows... This suggests...

## References


Ben-Michael, Eli, D James Greiner, Melody Huang, Kosuke Imai, Zhichao Jiang, and Sooahn Shin. 2024. "Does AI Help Humans Make Better Decisions? A Methodological Framework for Experimental Evaluation." *arXiv Preprint arXiv:2403.12108*.

Cahill, Christine, and Katherine McCabe. 2024. "Context Matters: Understanding Student Usage, Skills, and Attitudes Toward AI to Inform Classroom Policies." *PS: Political Science & Politics*, May, 1–8. https://doi.org/10.1017/S1049096524000155.

Dell'Acqua, Fabrizio, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Kross, Sean, and Philip J. Guo. 2019. "Practitioners Teaching Data Science in Industry and Academia: Expectations, Workflows, and Challenges." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://api.semanticscholar.org/CorpusID:102493683.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," no. arXiv:2302.06590 (February). http://arxiv.org/abs/2302.06590.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.