Self-reported LLM usage and results on a data science project: Evidence from a Canadian undergraduate data science course*

Rohan Alexander Luca Carnegie

August 4, 2024

To help understand the effect of Large Language Models (LLMs) on data science practice we examine the extent to which LLM usage is correlated with the mark that a student gets on a final paper in a classroom data science setting. We find no clear relationship between more extensive LLM usage and the student's mark. Despite the classroom setting used for evaluation, the particular activity of interest reflects the work done by professional data scientists. Our finding suggests the need for more extensive work evaluating how LLMs can be integrated into the data science workflow in a way that provides value.

1 Introduction

There are many aspects of trustworthy data science, including education, culture, and workflow to name a few. One aspect of special importance is the tools that we use to do data science: the computers, programming languages, settings, and environments. The wide release of user-friendly Large Language Models (LLMs), especially OpenAI's ChatGPT, is a rare instance of a new tool being made available.

Like all new tools, LLMs have created both excitement and apprehension. ChatGPT's public release on November 30, 2022, brought LLMs into the mainstream conversation. By now many people, especially educators and students, have some experience with LLMs, in both personal and professional contexts.

*Code and data are available at: https://github.com/lcarnegie/llms-achievement. We thank Nathalie Moon for developing and sharing her survey questions. This research is currently under review by the University of Toronto's University Research Ethics Board. Comments can be sent to: rohan.alexander@utoronto.ca.

In the context of teaching statistics and data science LLMs could be useful for many tasks. For instance, there is considerable interest in the potential of chatbots and personalized tutors [**ADD CITE**]. ChatGPT has also been the catalyst for many interesting and important conversations around academic integrity [**ADD CITE**], the development of critical thinking skills [**ADD CITE**], and what effective learning looks like [**ADD CITE**].

In this paper we are interested in better understanding LLMs as a tool for producing trustworthy data science. We study how they were used by students in an upper-year undergraduate data science course and whether students who used LLMs tended to have higher scores than those who did not.

The tasks involved in a trustworthy data science workflow can be generally broken down into a number of key competencies (Adhikari, DeNero, and Jordan 2021; Gibbs and Taback 2021). The potential for LLMs to positively affect students' academic performance in data science can be clearly inferred from already-demonstrated effects in a number of those competencies on adjacent professional fields. The first is programming, which is done when cleaning, analyzing, visualizing data often using languages like R or Python. The second is writing, which is primarily done when communicating results.

In terms of general programming, Peng et al. (2023) found a positive impact of GitHub Copilot (an LLM-powered programming assistant) on productivity. Specifically, in an experiment involving 95 freelance programmers, they found that that programmers with access to GitHub Copilot completed a standardized programming task 56% faster than the control group. Programmers with less experience saw the greatest improvements in productivity.

Dell'Acqua et al. (2023) focus on management consulting tasks and provide evidence that LLMs can improve writing productivity. In a field experiment involving consultants from Boston Consulting Group they found that the use of OpenAI's GPT-4 led to a 25% increase in delivery speed of business tasks (most involving some writing), as well as a 40% increase in human-rated performance on those tasks. Similar to computer programming, these productivity increases were most pronounced for those with below average performance, with their output increasing by 43%.

On the other hand, Valenzuela et al. (2024) argue that LLMs lead to a loss of serendipity (which leads to less original work) and de-skilling (primarily with respect to programming ability), among other consequences. These outcomes could negatively affect students' effective learning of data science.

Ellis and Slade (2023) take a more optimistic perspective and argue that LLMs are just another technology that will impact statistics and data science education. Similarly, Tu et al. (2024) acknowledge that LLMs can streamline many parts of a data science workflow. With that in mind, they suggest that data-scientists-in-training should shift their self-perspectives from primarily being an analyst to primarily being a manager responsible for strategic oversight of the analysis.

At the high school level Lazar et al. (2023) conducted a informal survey of secondary school teachers and students on their opinions of ChatGPT, and found that while LLMs could: help creativity, provide academic support when teachers were unavailable, and model certain types of writing well; teachers were also aware of LLMs potential to limit students' learning in certain ways through over-reliance. Beyond academic integrity concerns, teachers had similar concerns to Valenzuela et al. (2024) about de-skilling and an overall loss of agency in writing and critical thinking.

Cahill and McCabe (2024) surveyed undergraduate political science students on their attitudes toward and usage of AI tools. They found that the use of ChatGPT was widespread. However, they also found that many students lacked confidence in using AI for academic purposes. In particular, only 11% 'strongly agreed' that they know how to use AI to improve their writing. Students had nuanced views on appropriate AI use. In particular, respondents found that using it to write whole papers as inappropriate, while using it for basic tasks like general assistance, writing feedback and basic data visualization was perceived more appropriately.

To understand the current state of LLMs as a tool for trustworthy data science, this paper focuses on the association between student academic performance and their LLM usage. Specifically, we examine the relationship between students' grades and self-reported measures of student LLM usage, as well as student attitudes toward LLMs in general. This is based on final papers and a survey, conducted in a third-year undergraduate data science course at the University of Toronto. By examining how students interact with and perceive LLMs as tools, and how these variables translate into student outcomes, the effects of LLM integration in data science can be more precisely determined and recommendations made for future development.

The remainder of this paper is structured as follows: Section 2 visualizes and analyzes survey data and coursework from students. Section 3 specifies a model used to investigate the relationship. Section 4 describes and analyzes the model's results. Section 5 discusses the implications of the findings for data science education and future research and practice at the intersection of LLMs and trustworthy data science.

2 Data

2.1 Background

To investigate students' usage and attitudes towards LLMs and how they related to their academic performance, a dataset containing their usage/attitudes, coursework, and academic performance was constructed. This was based on three components:

1. an optional survey;
2. self-reported LLM usage; and
3. student marks on the final paper.

All data are from the cohort of students taking STA302 "Methods of Data Analysis I" in the Winter 2024 semester at the University of Toronto. This course had 275 students initially enrolled which, reflecting a normal rate of attrition for undergraduate statistics courses at the University of Toronto, reduced to 154 students by the end of the semester. Assessment was heavily based on three papers submitted over the course of the 12 week semester.

The student marks that we analyze are based only on the final paper, which is done individually. By this stage, uninterested students have typically dropped the course, and students are familiar with course expectations. A typical paper submission is 10-20 pages, and requires students to conduct original research to answer a research question of interest to them. It reflects the skills typically used by a professional data scientist. Students are expected to develop a research question of interest to them, identify or collect data to answer the question, conduct statistical analysis, and write a short paper. Examples of final papers (shared with consent) include: Yu (2024); Su (2024); and Rochwerg (2024).

By the time they are working on their final paper, students have submitted and received feedback on two previous papers with similar requirements and rubrics to that of the final paper. Each paper has the same basic structure and expectations. Before the final paper is due students have received feedback on all their previous work in the class (including their past papers) and there is an, optional, two-day period of peer review.

By virtue of pre-requisites of this course the typical student is an upper-year undergraduate. Coding and writing are major parts of the course. Students are welcome to use R or Python, but the majority code in R because that is the programming language currently mostly taught in pre-requisite courses. All writing must be in English. The aim of the course for students to create a public portfolio of work they can use to apply for jobs.

Throughout the semester students were encouraged to use LLMs. Formal instruction was provided twice during the semester, with a masterclass taught by a computer science faculty member on the ethics of using LLMs (see Horton et al. (2024) for details), and another masterclass taught by a TA on writing with LLMs.

Data was collected from students through an optional end-of-course survey. Appendix A provides all the questions asked in the survey. Whether or not they consented to their data being used, all respondents received a 1% increase in their final course grade for their participation. Consenting responses were then matched to their final paper mark, as well as the GitHub repository for their final paper. The responses were anonymized by removing any personal references to the students themselves including names, emails, student numbers, and GitHub links.

All data cleaning and analysis was done using the `R` statistical programming language (R Core Team 2023), and especially uses the `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `reshape2` (Wickham 2007), and `readxl` (Wickham and Bryan 2023) libraries.

2.2 Survey data

There were 146 responses to the survey. Of these, 119 respondents provided authorization for their data to be collected and used. Four of those respondents submitted the survey twice, and after removing their second response, 115 responses remained. Of those, 15 respondents did not include a statement on LLM usage in the README of the GitHub repository of their final paper, leaving 100 responses that were of use and were merged based on student name.

All but one respondent completed the survey within 45 minutes (Figure 1a). That one respondent took more than 4,000 minutes to complete the survey, suggesting they took a break while filling it out. That respondent is included in the analysis dataset, but without that respondent, the average time to complete the survey was 11 minutes, and the standard deviation was 8 minutes.

There is a wide distribution of self-reported GPAs (Figure 1b). Seven of the 100 respondents included in the analysis dataset did not report their GPA. The majority of responses cluster around a B (3.0/4.0), and the average is 3.06 with a standard deviation of 0.55. One factor that may affect the range is that the course is required for programs in the Statistics, Mathematics and Computer Science Departments. Self-reported GPAs introduce the possibility of reporting bias. For instance, respondents may have provided their cumulative GPA, their most recent term's GPA, or could have misreported it entirely.

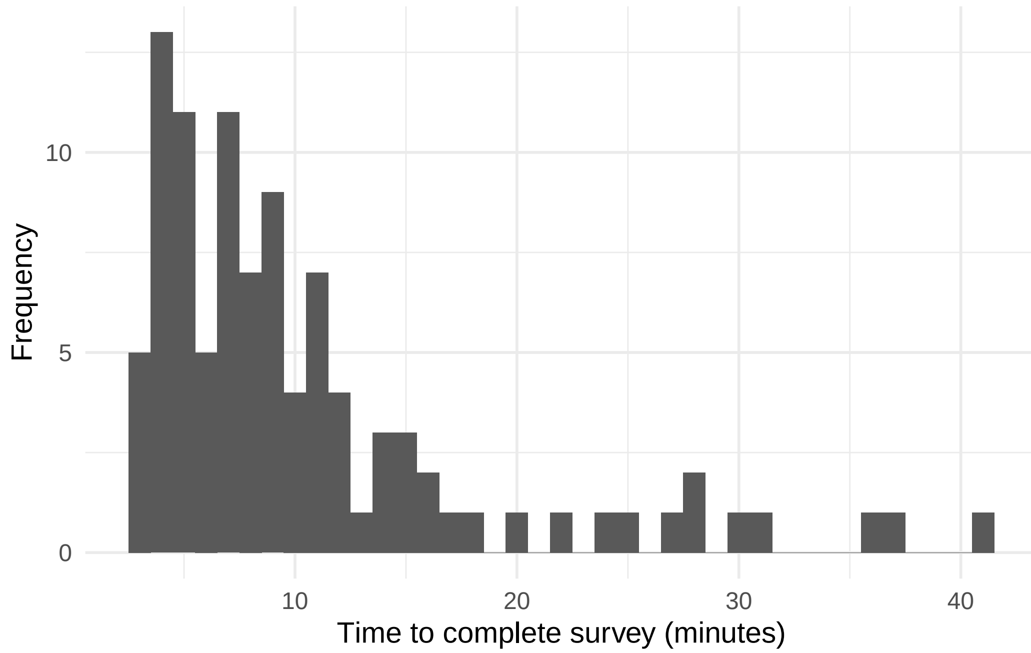
Students from a range of years took the course, however the majority of respondents were in their 3rd or 4th year of study (Table 1). The course's prerequisite of general statistics, which is a two-course sequence typically completed by students in their second year, would make it difficult to take this course earlier than 3rd year.

Table 1: "What year are you?"

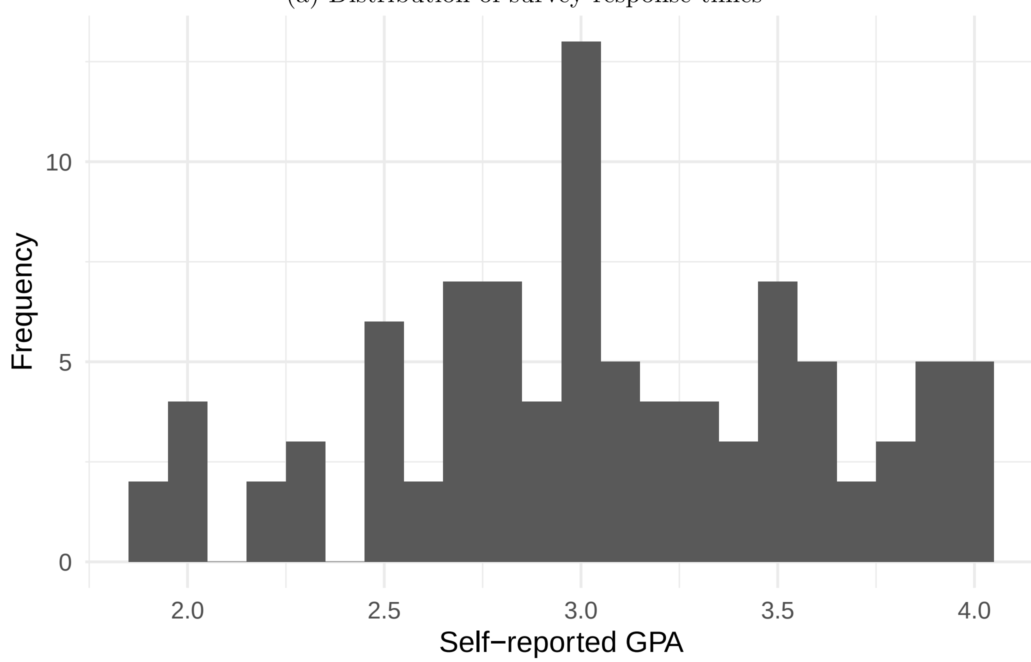
2nd	3rd	4th	5th or over
4	52	42	2

Respondents had a varied self-perception of their coding and writing abilities (Figure 2). Most respondents believed it is important to be good at writing, but many are either indifferent or do not like to write (Figure 2a). Respondents also do not find writing to be particularly easy, which could be associated with the reported relative antipathy toward writing. Most respondents were at least somewhat confident in their own writing abilities, but a substantial contingent felt otherwise. Although few respondents felt that they were confident in their writing ability, more felt that they were able to catch their mistakes, which could indicate a disconnect between how respondents perceive their work and how the work was evaluated.

Respondent self-perceptions regarding coding proficiency and importance varied (Figure 2b). There was strong consensus on the perceived importance of coding skills, with a majority of respondents strongly affirming this belief. However, respondents' self-assessed coding abilities

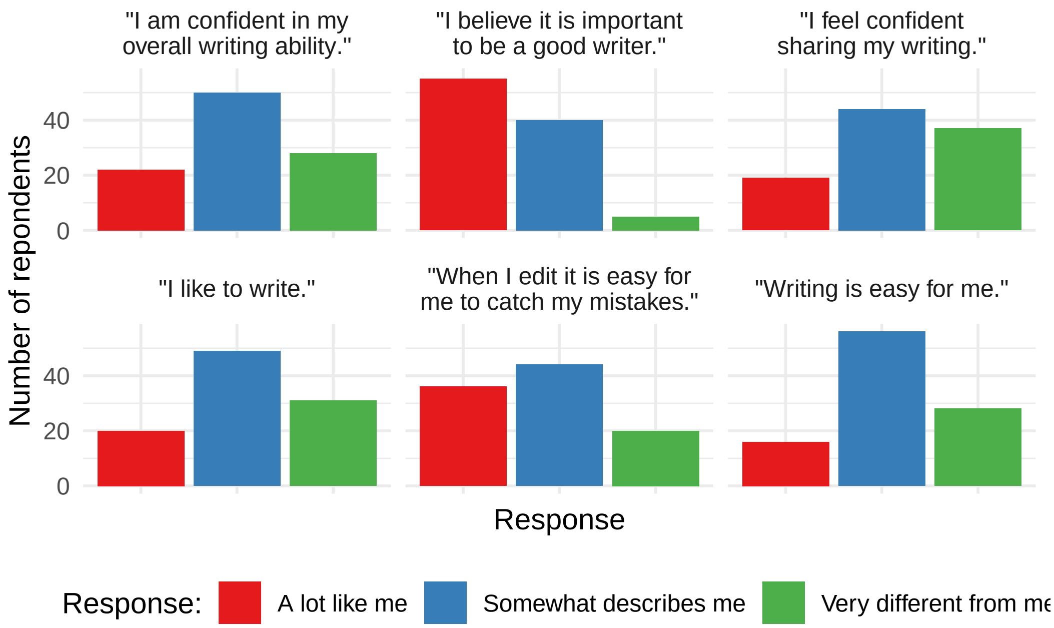


(a) Distribution of survey response times

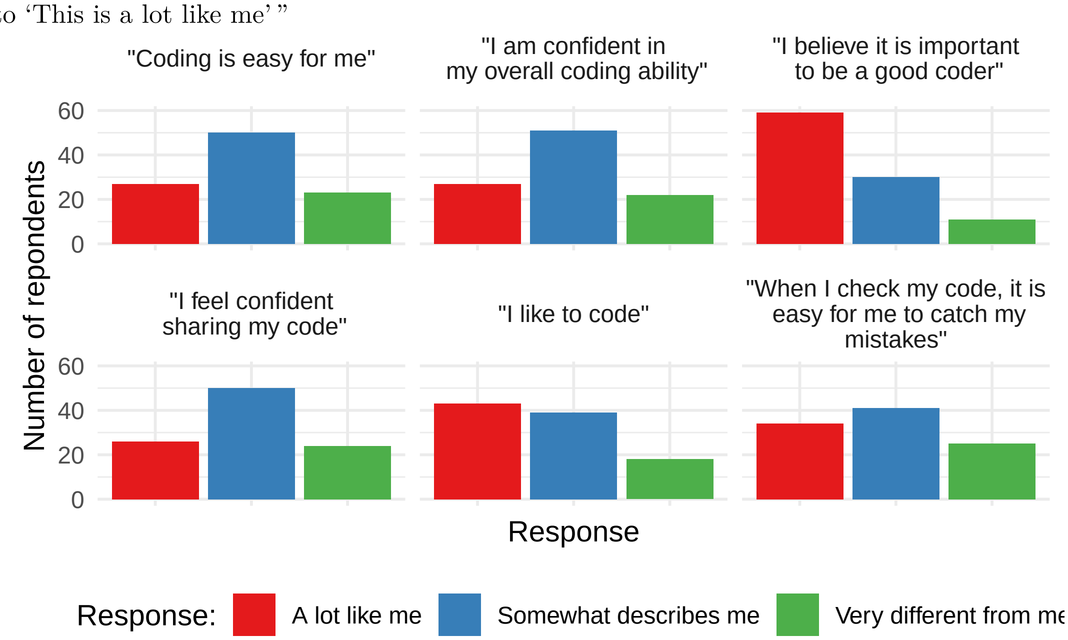


(b) “What is your GPA?”

Figure 1: Distribution of respondents’ survey response times and GPA



(a) "Please rate how much each statement describes you, on a scale from 'This is very different to me' to 'This is a lot like me' "



(b) "Please rate how much each statement describes you, on a scale from 'This is very different to me' to 'This is a lot like me' "

Figure 2: Self-perception of coding and writing abilities

Table 2: Familiarity with, and appropriateness of, generative AI

(a) 'How familiar are you with using generative AI tools such as OpenAI's ChatGPT or equivalents?'

Extent of AI familiarity	Number
Not familiar	2
Somewhat familiar	56
Very familiar	42

(b) 'To what extent do you think using generative AI tools such as ChatGPT by OpenAI (or equivalents) is ethical and appropriate for schoolwork?'

Ethical and appropriate for school?	Number
Appropriate	80
Inappropriate	9
It depends	11

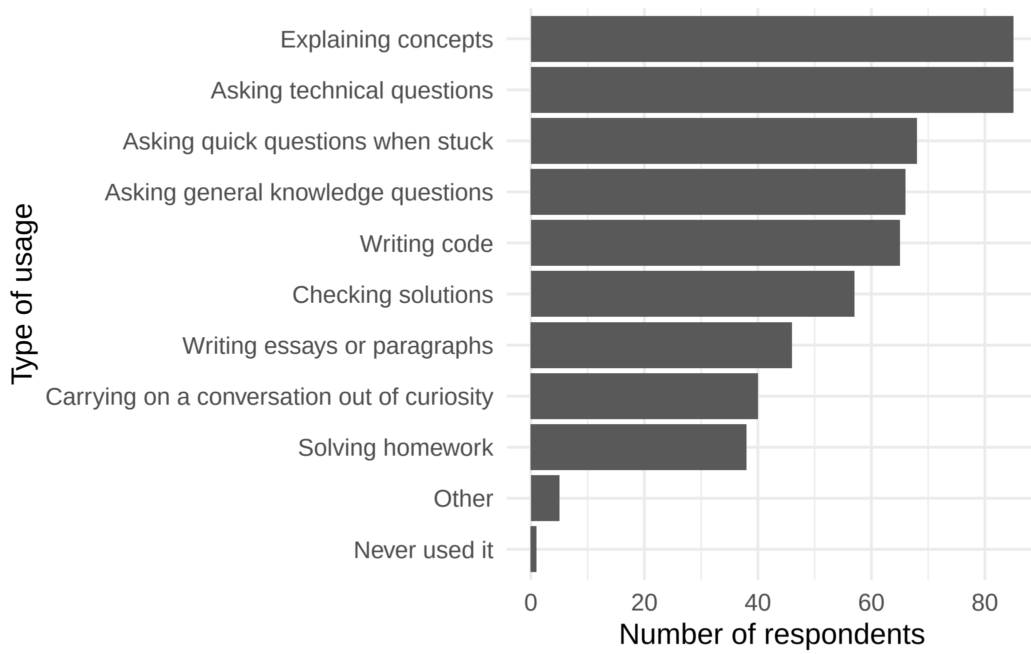
and enjoyment are more heterogeneous, with a substantial proportion reporting moderate rather than high levels of ease and enjoyment in coding tasks. These findings suggest a complicated relationship between respondents' recognition of coding's importance and their personal experience with coding.

Overall there was a moderate level of confidence among respondents in their overall coding ability, willingness to share code, and capacity to identify errors (Figure 2b). Notably, respondents express slightly higher confidence in detecting their own coding mistakes compared to general coding ability or code sharing. These patterns suggest that while respondents have developed some coding self-efficacy, there is still considerable potential for enhancing their perceived competence and comfort across various coding-related activities.

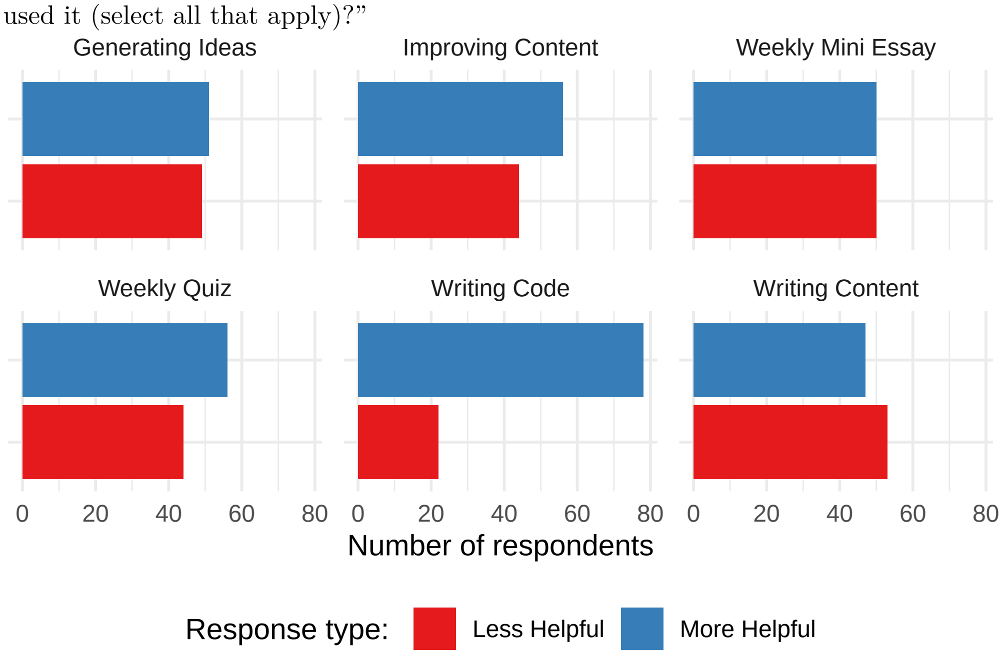
The majority of respondents were at least "somewhat familiar" with generative AI such as ChatGPT (Table 2a). Though respondents' self-perceptions around writing and coding are varied, there was strong consensus that the use of generative AI tools is appropriate within an academic setting (Table 2b). Most respondents who responded "It depends" generally found artificial intelligence tools to be appropriate, though with certain guidelines and rules governing their use.

To understand the role of LLMs in learning, students more granularly identified their usage by checking off various pre-defined use cases in the survey (Figure 3a). Technical questions and explaining concepts were the two top use cases among students in the course. More than half of students also used LLMs for quick questions, general knowledge, writing paper code, and checking solutions. Just less than half used it to write paper content, which could suggest that students do not feel confident using it to improve writing.

Respondents were also asked to rate the helpfulness of LLMs on various tasks assigned during the course on a 4-point scale of "Did not use" to "Very Helpful" (Figure 3b). To simplify



(a) "If you have used generative AI tools such as OpenAI's ChatGPT or equivalents, in what ways have you used it (select all that apply)?"



(b) "How helpful did you find generative AI tools such as ChatGPT by OpenAI (or equivalents) for each component of STA302?"

Figure 3: Use and usefulness of generative AI

Table 3: Comparing Incidences of Positive and Negative Words in Student.

(a) Top 10 Positive Words

word	n
helpful	9
fine	7
improve	5
recommend	5
beneficial	3
creative	3
effectively	3
enhance	3
nice	3
properly	3

(b) Top 10 Negative Words

word	n
complex	3
critical	3
difficult	3
error	2
errors	2
hard	2
mistakes	2
worry	2
abuse	1
break	1

the presentation, responses were grouped into two main categories: "Less Helpful" and "More Helpful." The "Less Helpful" category combines responses where students found the AI either "Not helpful" or did not use it for the task, while the "More Helpful" category includes responses where the AI was considered "Somewhat helpful" or "Very helpful".

Respondents differed in terms of how they used LLMs in the course (Figure 3b). While most tasks were roughly split between respondents finding LLMs helpful or not, respondents found them most helpful in generating code. In the context of the course, this meant generating R code for transforming, analyzing, and visualizing data. To a lesser extent, respondents also found LLMs to be helpful in improving the existing writing they had, while at the same time not favouring it for writing content from scratch.

2.3 Open-ended survey responses

One question asked students to elaborate on whether they thought generative AI tools such as ChatGPT were ethical and appropriate for schoolwork. This was an open-response question. We summarized the responses to that question using text analysis and the `tidytext` package (Silge and Robinson 2016).

2.4 LLM usage and final paper marks

Two other components were merged with the survey responses: self-reported LLM usage on the final paper, and final paper mark.

Students were encouraged to use LLMs to complete their papers. Each paper required the students to disclose their usage through a statement in the GitHub repository README for the paper. Even students who did not use generative AI at all were required to state this in the README. For students who did use generative AI, there was an additional requirement that they save the logs of their usage in a txt file which was also included in their GitHub repository.

Those README statements were gathered and parsed using OpenAI's ChatGPT 4o model (as at 26 July). The following prompt was used: 'The following statement is about to what extent LLMs were used by a student. Please characterize it as one of: "None", "Minimal", "Somewhat", "Extensive", "Unsure". Respond with only one of those options.' All classifications were then manually checked by hand for reasonableness.

We find a varied extent of self-reported LLM usage (Table 4). 41 respondents were classified as having made extensive use of LLMs, while 28 were classified as having made somewhat use. 31 respondents were classified as having made minimal or no use of LLMs. In the analysis dataset we combine those two classifications because only 8 respondents were classified as having minimal usage.

Table 4: Self-reported LLM usage in final paper

Self-reported LLM usage	Number
Extensive	41
Minimal	8
None	23
Somewhat	28

The third, and final, component is the mark, in percentages, on the final paper (Figure 4). The overall mean was 78% and standard deviation was 17 percentage points. However there was considerable differences by the extent of LLM usage. For extensive use, the mean was 82% and the standard deviation was 15 percentage points. For somewhat usage, the mean was 77% and the standard deviation was 16 percentage points. For minimal usage, the mean was 74% and the standard deviation was 23 percentage points. Finally for no usage, the mean was 75% and the standard deviation was 19 percentage points.

3 Model

The goal of our modelling strategy is to better understand how respondents result on their final paper is associated with their self-reported LLM usage. The result on the final paper is a proportion and some respondents got full marks and so we use zero-one-inflated beta regression. Here we briefly describe the model that we use, which follows Kurz (2023). Diagnostics are included in Appendix B.

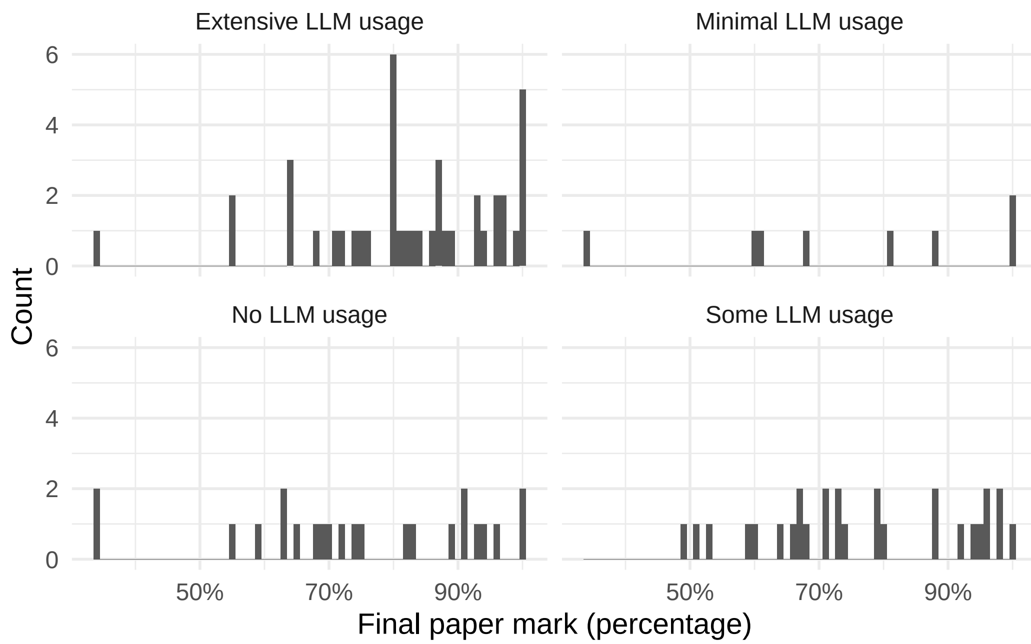


Figure 4: Mark on final paper, in percentages, and LLM usage in final paper

Define y_i as the percentage received on the final paper. Then β_1 is the characterization of self-reported LLM usage and β_2 is self-reported GPA.

$$
\begin{aligned}
y_i &\sim \text{Beta}(\mu_i, \phi) && (1) \\
logit(\mu_i) &= \beta_0 + \beta_1 \times \text{LLM usage}_i + \beta_2 \times \text{GPA}_i && (2) \\
\beta_0, \beta_1, \beta_2 &\sim \text{Normal}(0, 1) && (3) \\
\phi &\sim \text{Gamma}(4, 0.1) && (4)
\end{aligned}
$$

We estimate the model in `R` (R Core Team 2023) using `brms` (Bürkner 2017).

The expected relationship between LLM usage and final mark is unclear. It may be that stronger students used LLMs in a more sophisticated way or that they did not need to use them. However, the relationship between GPA and final mark is expected to be positive.

4 Results

Our results are summarized in Table 5 and Figure 5. We especially draw on `modelsummary` (Arel-Bundock 2022).

Table 5: Coefficient estimates and mean absolute deviation (MAD)

	Base model	Including self-reported GPA
b_Intercept	1.61	−1.84
	(0.15)	(0.40)
b_llm_usageNoneorminimal	−0.46	−0.58
	(0.21)	(0.17)
b_llm_usageSomewhat	−0.35	−0.32
	(0.22)	(0.19)
phi	5.49	10.39
	(0.76)	(1.48)
b_what_is_your_gpa		1.17
		(0.13)
Num.Obs.	93	93
R2	0.052	0.507
ELPD	55.2	82.9
ELPD s.e.	6.4	8.1
LOOIC	−110.4	−165.8
LOOIC s.e.	12.8	16.1
WAIC	−110.4	−165.9
RMSE	0.15	0.11

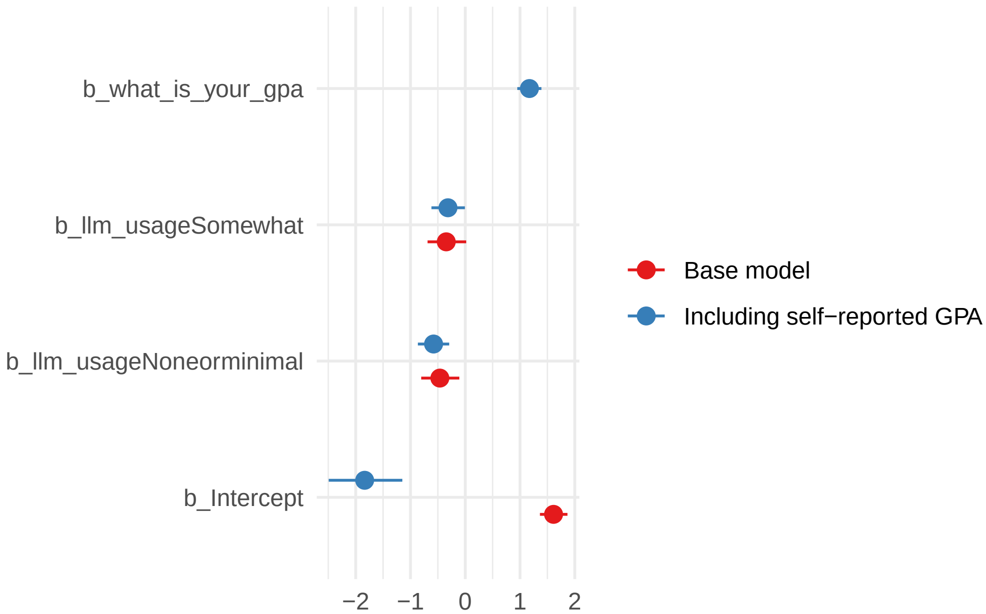


Figure 5: Coefficient estimates and 90 per cent credibility intervals

Interestingly, we find a slightly negative effect, relative to self-reported extensive use, of using LLMs less (Figure 5). As expected, self-reported GPA is positively associated with final mark.

As always with regression, average estimates over the full dataset. It may be that the results are different for different tranches of respondents. For instance, looking at the 28 students who received an A+ for the final paper, 13 of them had extensive LLM usage.

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

We don't know whether it made them faster.

Appendix

A Survey questions

1. After carefully reading the informed consent document, please indicate below whether you consent to have your anonymized responses included in the research study?
    - Yes, I authorize the use of the data collected about me for the STA302 course survey to be used. I will be compensated 1% of my course grade for completing the survey.
    - No I do not want my data included in the research study, but I want to complete the survey. I will be compensated 1% of my course grade for completing the survey.
    - I do not want to complete this survey. I realize that I am forfeiting the corresponding course credit.
2. What is your full name on Quercus?
3. What is your Student ID?
4. What year are you?
5. What is your specialization?
6. What is/are your major/s?
7. What is/are your minor/s?
8. What is your GPA?
9. Please rate how much each statement describes you, on a scale from "This is very different to me" to "This is a lot like me" ["This is very different to me"; "This somewhat describes me"; "This is a lot like me"]
    - Writing is easy for me
    - I like to write
    - I believe it is important to be a good writer.
    - When I edit it is easy for me to catch my mistakes.
    - I feel confident sharing my writing.
    - I am confident in my overall writing ability.
10. Please rate how much each statement describes you, on a scale from "This is very different to me" to "This is a lot like me". When answering, please consider whichever programming language you are most familiar with. ["This is very different to me"; "This somewhat describes me"; "This is a lot like me"]
    - Coding is easy for me
    - I like to code
    - I believe it is important to be a good coder.
    - When I check my code it is easy for me to catch my mistakes.
    - I feel confident sharing my code.
    - I am confident in my overall coding ability.

11. How familiar are you with using generative AI tools such as OpenAI's ChatGPT or equivalents?
    - Very familiar
    - Somewhat familiar
    - Not familiar
    - Other
12. Have you used any generative AI tools such as OpenAI's ChatGPT or equivalents for any reason (personal or educational)?
    - Yes
    - No
    - Other
13. If you have used generative AI tools such as OpenAI's ChatGPT or equivalents, in what ways have you used it (select all that apply)?
    - Asking technical questions
    - Carrying on a conversation out of curiosity
    - Asking general knowledge questions
    - Solving homework
    - Checking solutions
    - Asking quick questions when stuck
    - Explaining concepts
    - Writing essays or paragraphs
    - Writing code
    - Never used it
    - Other
14. To what extent do you think using generative AI tools such as ChatGPT by OpenAI (or equivalents) is ethical and appropriate for schoolwork?
    - Appropriate
    - Inappropriate
    - Other
15. Please elaborate on your answer above.
16. Did you use any generative AI tools such as OpenAI's ChatGPT or equivalents for STA302?
    - Yes
    - No
    - Other
17. How helpful did you find generative AI tools such as ChatGPT by OpenAI (or equivalents) for each component of STA302? ["Not helpful"; "Somewhat helpful"; "Very helpful"; "I did not use generative AI for this component"]

- Weekly quiz
- Weekly mini-essay
- Papers: Generating ideas
- Papers: Writing code
- Papers: Writing content
- Papers: Improving content

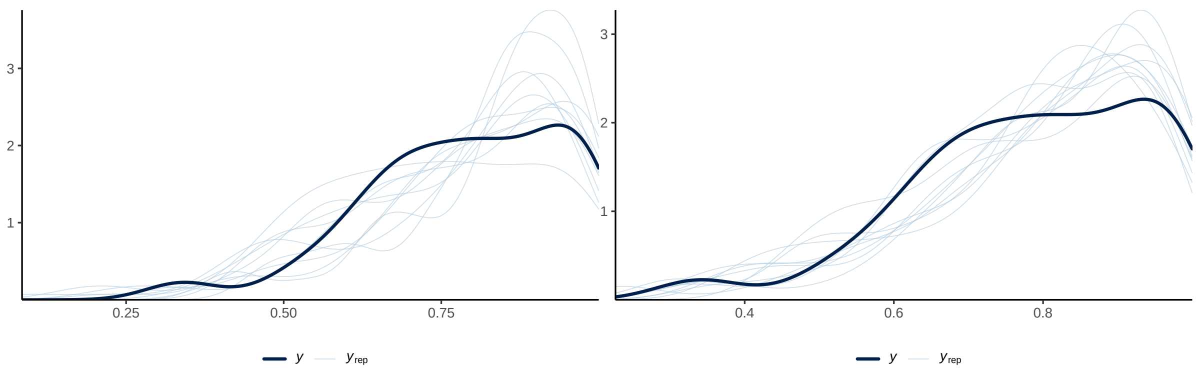
18. What is your recommendation for how generative AI tools such as ChatGPT by OpenAI (or equivalents) should be used in the course in future?
19. (Optional) Any other comments?

Table 6: Model comparison

	elpd_diff	se_diff
fit2	0.0	0.0
fit1	-27.7	6.7

B Model details

B.1 Posterior predictive check



(a) Base model

(b) Including self-reported GPA

Figure 6: Posterior predictive checking

B.2 Diagnostics

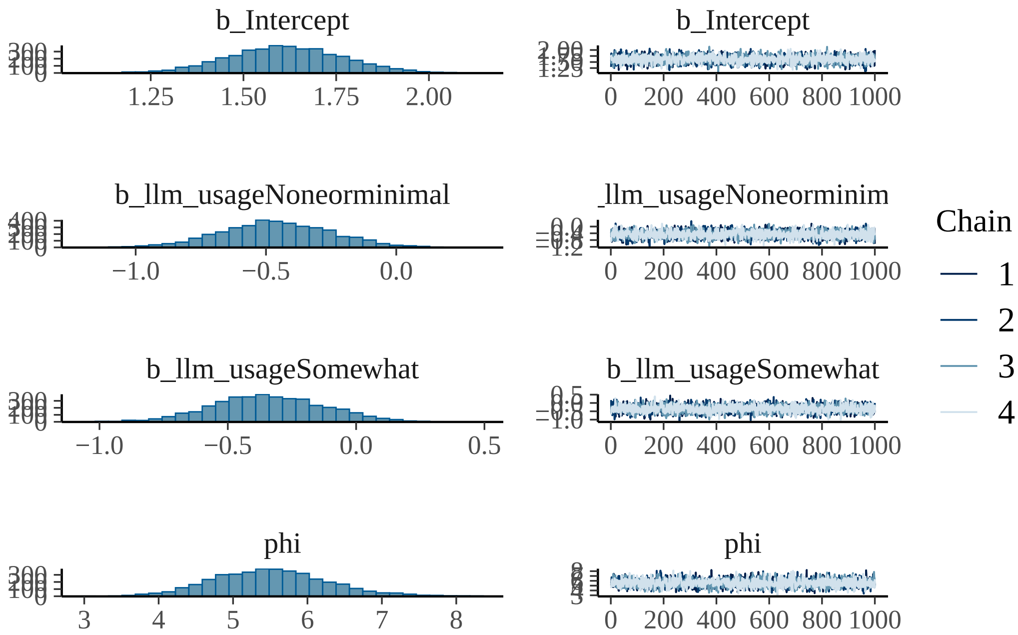


Figure 7: Base model diagnostics

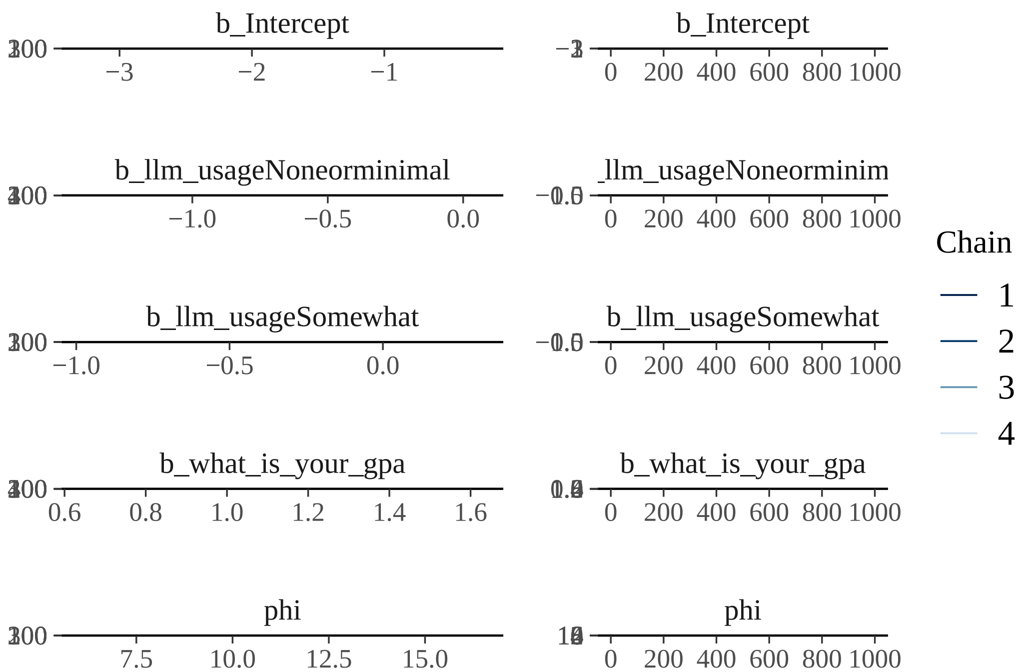


Figure 8: Model including self-reported GPA diagnostics

References

Adhikari, Ani, John DeNero, and Michael I. Jordan. 2021. "Interleaving Computational and Inferential Thinking: Data Science for Undergraduates at Berkeley." *Harvard Data Science Review* 3 (2). https://doi.org/10.1162/99608f92.cb0fa8d2.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bürkner, Paul-Christian. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

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.

Ellis, Amanda R., and Emily Slade. 2023. "A New Era of Learning: Considerations for ChatGPT as a Tool to Enhance Statistics and Data Science Education." *Journal of Statistics and Data Science Education* 31 (2): 128–33. https://doi.org/10.1080/26939169.2023.2223609.

Firke, Sam. 2023. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. https:

//CRAN.R-project.org/package=janitor.
Gibbs, Alison L., and Nathan Taback. 2021. "The Building Blocks of Statistical Education in the Data Science Ecosystem." *Harvard Data Science Review* 3 (2).
Horton, Diane, David Liu, Sheila A. McIlraith, Steven Coyne, and Nina Wang. 2024. "Do Embedded Ethics Modules Have Impact Beyond the Classroom?" In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education v. 1*. SIGCSE 2024. ACM. https://doi.org/10.1145/3626252.3630834.
Kurz, Solomon. 2023. "Causal Inference with Beta Regression," June. https://solomonkurz.netlify.app/blog/2023-06-25-causal-inference-with-beta-regression/.
Lazar, Nicole, James Byrns, Danielle Crowe, Meghan McGinty, Angela Abraham, Mike Guo, Megan Mann, et al. 2023. "Perils and Opportunities of ChatGPT: A High School Perspective." *Harvard Data Science Review* 5 (4).
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/.
Rochwerg, Benny. 2024. "Evidence of Racial Profiling by the Austin Police Department in 2020." https://github.com/bennyrochwerg/profiling/blob/main/paper/paper.pdf.
Silge, Julia, and David Robinson. 2016. "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *JOSS* 1 (3). https://doi.org/10.21105/joss.00037.
Su, Emily. 2024. "Characteristics of Top Songs Has Changed from Pandemic Brain: An analysis of songs on Billboard's Year-End Hot 100 singles list (2014 to 2023)." https://github.com/moonsdust/top-songs/blob/main/paper/paper.pdf.
Tu, Xinming, James Zou, Weijie Su, and Linjun Zhang. 2024. "What Should Data Science Education Do With Large Language Models?" *Harvard Data Science Review* 6 (1). https://doi.org/10.1162/99608f92.bff007ab.
Valenzuela, Ana, Stefano Puntoni, Donna Hoffman, Noah Castelo, Julian De Freitas, Berkeley Dietvorst, Christian Hildebrand, et al. 2024. "How Artificial Intelligence Constrains the Human Experience." *Journal of the Association for Consumer Research* 9 (3): 241–56. https://doi.org/10.1086/730709.
Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.
Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.
Wickham, Hadley, and Jennifer Bryan. 2023. *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl.
Yu, Hannah. 2024. "Fake News vs Fox News: The Influence of Media Preferences on Voting Behavior in the 2020 U.S. Presidential Election Among Party Voters." https://github.com/hannahyu07/Fox-News/blob/main/paper/Fake_News_vs_Fox_News.pdf.