Can LLMs Perform Data Analysis?*

My subtitle if needed

Evelyn Hughes

June 11, 2025

LLMs have become widely known and popular for their ability to perform a large variety of tasks. Recently, there has been a growing interest in determining their ability to perform more technical tasks such as data analysis. (With the increase in LLM use for data analysis tasks, it becomes increasingly important to understand and quantify their abilities in this area.) Currently, there is a lack of insight into this area given the lack of an efficient "evaluation pipeline". In this paper, we first create such an "evaluation pipeline", and then use it to evaluate popular models on their ability to perform data analysis.We find _____ . As LLMs become increasingly involved in data analysis, it becomes important to understand their capabilities.Being able to efficiently quantify LLMs ability in these tasks improves both transparency and trust, both fundamental to science.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

*Code and data are available at: https://github.com/EveHughes/Can-LLMs-Perform-Data-Analysis.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Results

Model	Replicable	Tests	Prose	References	Introduction	Data	Abstract	Title	Figures
”ChatGPT 3.0”	1	0	2	0	1	0	1	2	2
”ChatGPT 4.0”	1	1	2	0	2	2	1	2	1
”Gemini”	2	2	0	1	0	0	1	1	0
”Claude”	1	2	2	0	1	2	0	2	0

Figure 1: Bills of penguins

(1) shows the scores of each model on the rubric, with categories aggregated.

(2) shows the scores of each model aggregated across categories

Model	Code	Paper	Total
"ChatGPT 3.0"	1	8	9
"ChatGPT 4.0"	2	10	12
"Gemini"	4	3	7
"Claude"	3	7	10

Figure 2: Bills of penguins

4 Discussion

4.1 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.

4.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

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

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.