



A Comparative Analysis of the Effectiveness of Artificial Intelligence Systems in Response to Search Commands

Paige Elkin & Faye Jones



2000799
2000583

Introduction

Research Question - How do artificial intelligence (AI) assistants compare to manual literature search in finding relevant research papers?

Hypothesis - AI in the form of large language learning models will be able to return more relevant research articles than manual search.

FSU's Database OneSearch - "discovery search tool for books, articles, media, and more. This tool searches our local FSU collections as well as our online resources." [1]

Large Language Model - "large deep learning models that are pre-trained on vast amounts of data... learn to understand basic grammar, languages, and knowledge" [2]

Backtracking CTE Pathways

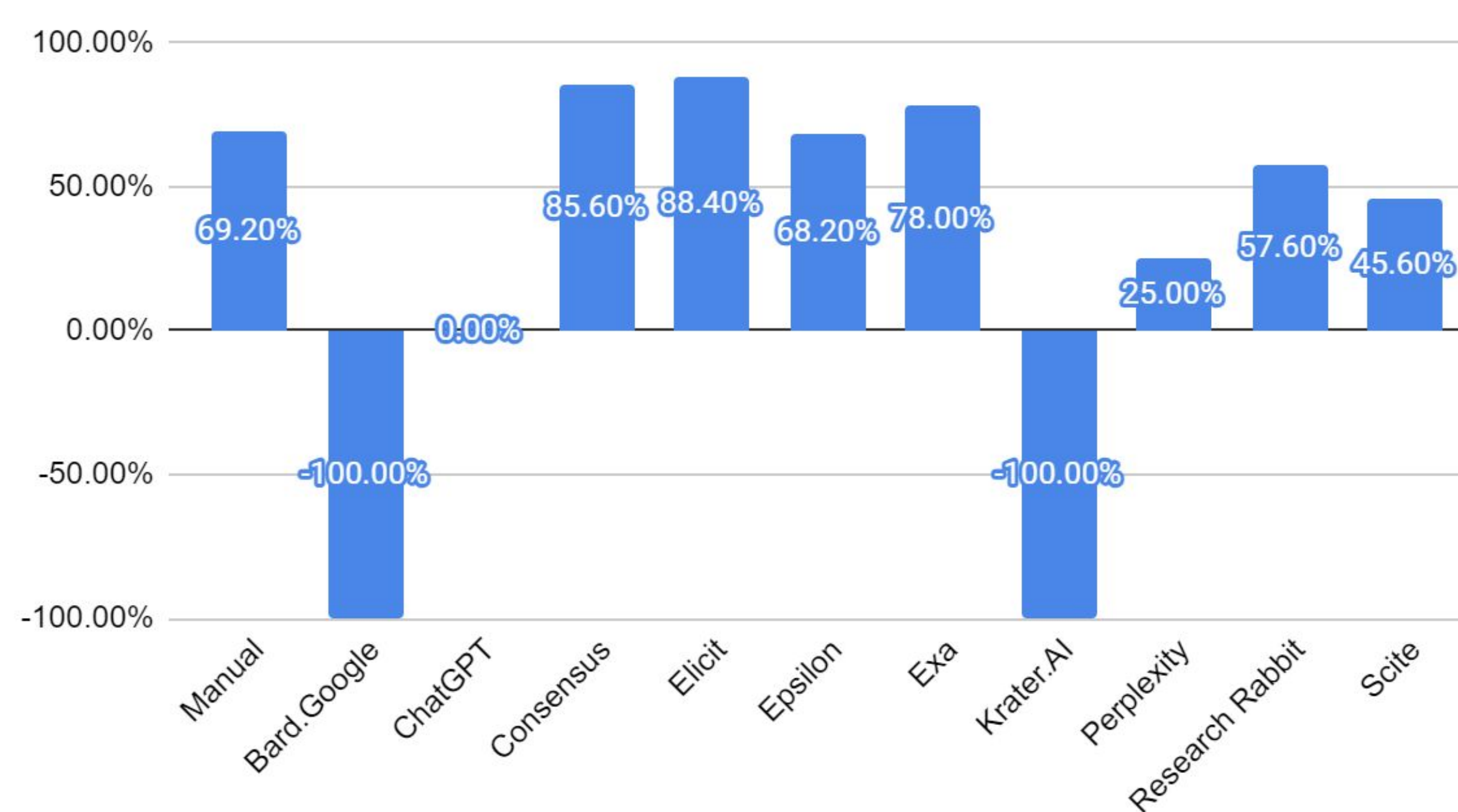
The literature found through manual and AI-aided searches will be used in the literature review for Backtracking CTE Pathways.

Backtracking CTE Pathways: Identifying and Investigating Pathways and Critical Junctures in Two-Year Information Technology Programs aims to

- Understand the pathways students take through college and into careers in IT
- Gather data about former students and information from surveys and interviews
- Analyze students who completed programs, didn't complete programs, or transferred institutions
- Reveal opportunities and hindrances students face as they pursue credentials in information technology

Contacts: Marcia Mardis (mmardis@fsu.edu), Faye Jones (fjones@fsu.edu)

Average Success for Each Method



Method

Queries

Manual Query: FSU OneSearch

Any Field Contains (hindrance* OR barrier*) AND Any Field Contains (2-year OR postsecondary OR post-secondary) AND Any Field Contains (STEM OR CTE) SEARCH

Any Field Contains (facilitator* OR opportunit*) AND Any Field Contains (2-year OR postsecondary OR post-secondary) AND Any Field Contains (STEM OR CTE) SEARCH

AI Aided Search Query

Find me the best research articles about hindrances and facilitators to academic success in 2-year post secondary CTE and STEM programs. Only show papers that were peer-reviewed and published after 2012

Focus Attach Pro

Scoring

Rubric included score of 0, 0.5, or 1 depending on the degree to which each category below was addressed:

- Hindrances or facilitators • Academic success • 2 year or post-secondary programs
- STEM or CTE programs • Peer reviewed, published after 2012

Article scores were then averaged to get an overall representation of the effectiveness of each search method or program (0=lowest; 5=highest)

View the data collection, analysis, and visualization process



- Bard** Large language model || Note: Renamed Gemini 2/9/2024 [3]
- ChatGPT** Language model trained to produce text, optimized with human feedback [4]
- consensus** Uses language models, source material from Semantic Scholar [5]
- Elicit** Searches Semantic Scholar, uses language models [6]
- Epsilon** Passes top 100 Semantic Scholar papers to GPT-4 [7]
- EXA AI** Knowledge API for large language models [8]
- krater.ai** Major generative AI apps in a single user interface [9]
- perplexity** Search engine using Generative AI and large language model [10]
- ResearchRabbit** Citation-based literature mapping tool [11]
- scite_** Large language models backed by database of Smart Citations [12]

Conclusions

- AI marketed for research performed at the same level or above manual searches
- General use AI often misunderstood the query. Even when the query was understood, it performed below the success of manual search
- Low correlation between number of articles returned and overall relevancy of articles

Implications

- Improvement to AI Systems
- Guide to using AI in academia
- Literature applied to Backtracking CTE Pathways project

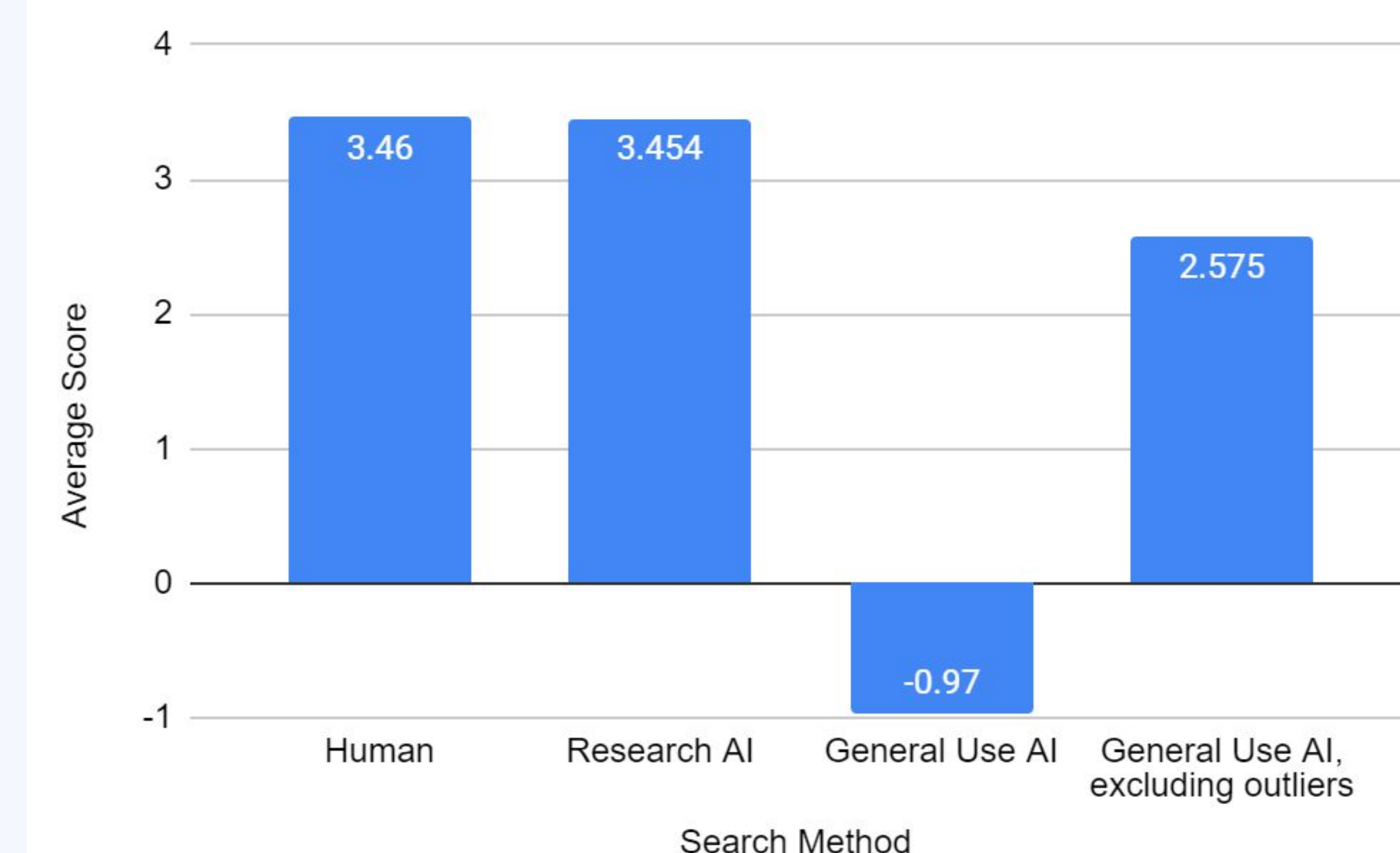
Suggestions for Future Research

- Retry queries as AI programs continue to be trained
- Test other research topics and queries
- Use more sophisticated features offered on the sites further than the general chat feature
- Evaluate other features than articles returned
 - Ease of use
 - Price

View References



Average Score vs. Search Method



DISCLAIMER: It is important to note that each AI was used on a preliminary basis, many without further search than the original prompt. Additionally, AI as a technology is constantly being trained and updated, so it is likely that the performance of these assistants have improved since the original searches.