

# Opportunities and Limitations in the Use of AI to Assist With Data Extraction in Systematic Literature Reviews


Eleanor King – Researcher, Hannah Rice – Senior Researcher, Dr Kalliopi (Kelly) Roussi – Senior Researcher

What?  
Data extraction (DE) is the most time-consuming task within the systematic literature review (SLR) process and high accuracy is crucial.

Why?  
We aimed to assess technical factors affecting DE efficiency by human researchers and evaluate how far AI tools can increase DE accuracy and speed.

Who?

- We performed data extraction of 26 studies in 3 different formats (conference abstracts, editable pdfs, non-editable pdfs)
- We tested 3 data extraction platforms (Elicit, Perplexity, and PDF AI)
- We extracted study details and baseline characteristics for each of the studies.

 **Elicit** <https://elicit.com/>, May 2024

Our experience:

- ✓ Allows the selection of bespoke categories
- ✓ Data were extracted accurately and within seconds
- x Manual replication for each study
- x Export to csv only with paid subscription

 **perplexity** <https://www.perplexity.ai/>, May 2024

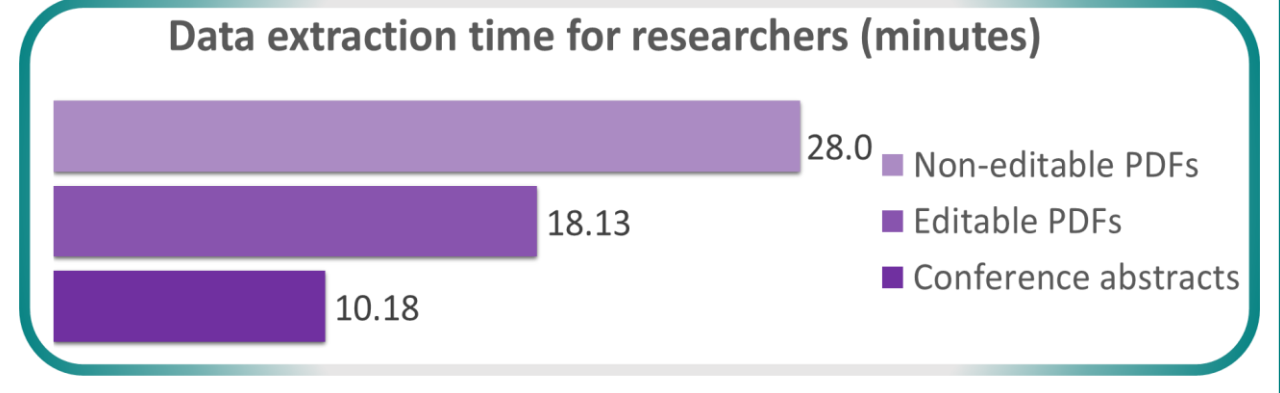
Our experience:

- ✓ A prompt was required to extract data from papers
- x Although some information was extracted correctly, parts of the output were fabricated

 **PDF AI** Microsoft Excel add-in, May 2024

Our experience:

- x The data extraction could not be completed because the provided Excel template, identical to the one used by the researchers, was unreadable by the platform



Author Comments

- Despite the rapid evolution of AI tools, there are still limitations pertaining to their use, mainly associated with the accuracy of the data extraction from the papers and the flexibility of the AI platform.
- Next steps we aim to evaluate the potential use of ChatGPT in data extraction.