Artificial Intelligence and the Literature Search

Gaelen P. Adam, MLIS MPH PhD

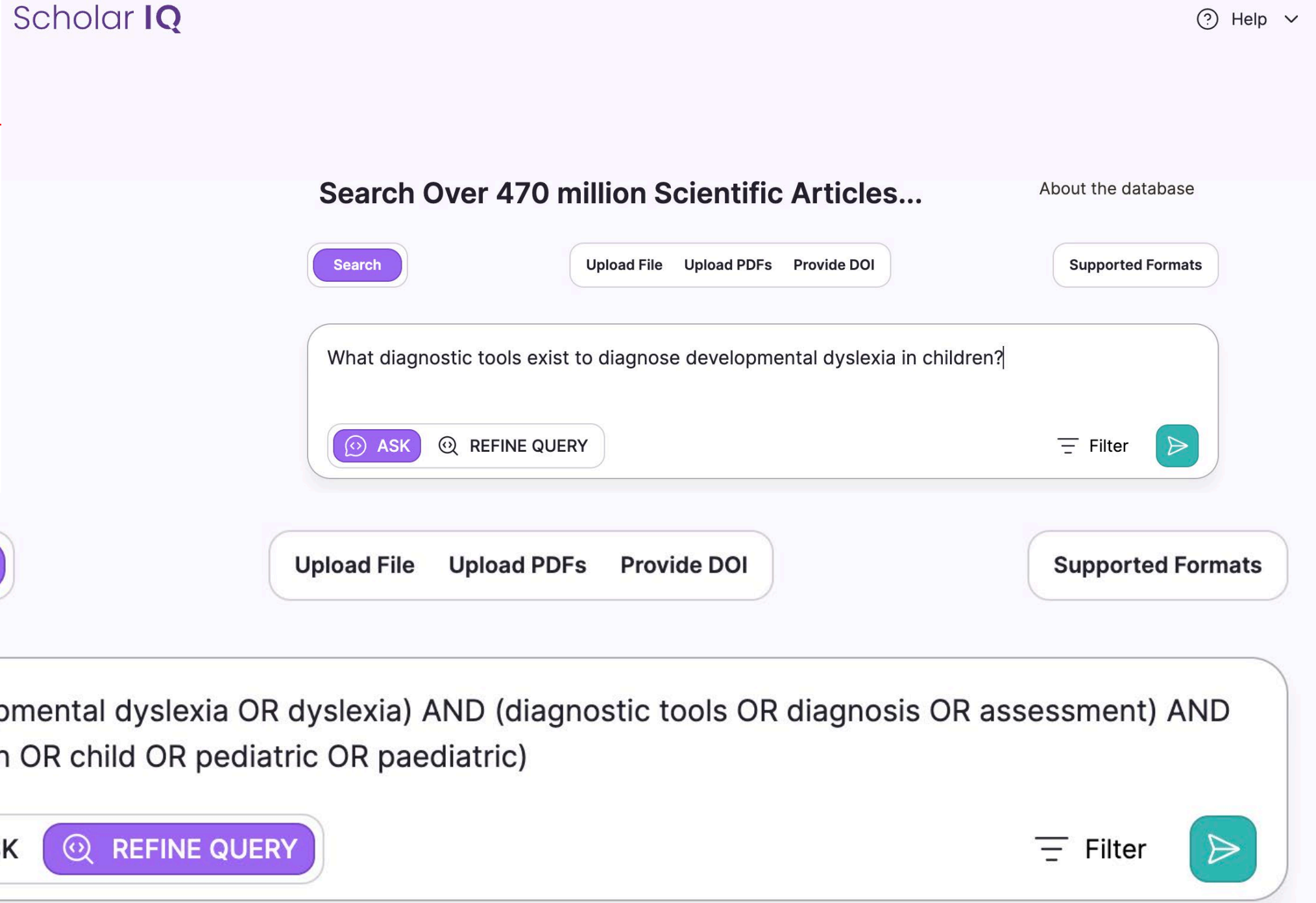
May 1, 2026

Disclosures

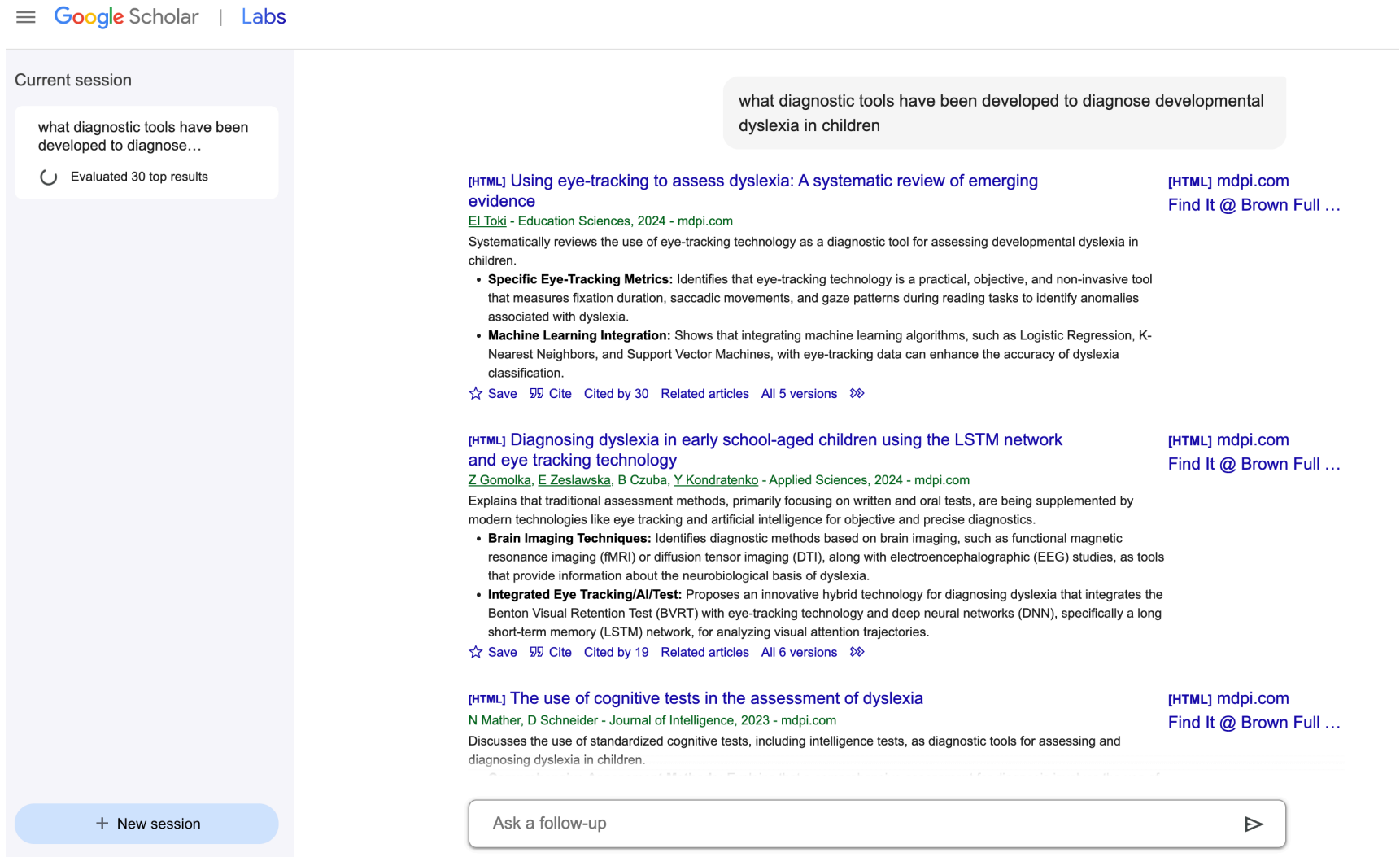
- I have no financial or other conflicts of interest
- I used AI in creating this presentation, though active-learning screening in SRDR+ (now VADRR), and help with r code for the figures
- This presentation is based on a living rapid review conducted under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD*
- Full report and all updates available at:
  https://effectivehealthcare.ahrq.gov/products/machine-learning-tools/white-paper

*The findings and conclusions presented are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ, or the U.S. Department of Health and Human Services.

Search


Scholar IQ
Help
Search Over 470 million Scientific Articles...
About the database
Search
Upload File
Upload PDFs
Provide DOI
Supported Formats
What diagnostic tools exist to diagnose developmental dyslexia in children?
ASK
REFINE QUERY
Filter
Search
Upload File
Upload PDFs
Provide DOI
Supported Formats
(developmental dyslexia OR dyslexia) AND (diagnostic tools OR diagnosis OR assessment) AND (children OR child OR pediatric OR paediatric)
ASK
REFINE QUERY
Filter

Search tools



Key Question and Inclusion Criteria

Key Question: What tools that use artificial intelligence (AI) or machine learning (ML) to semi- or fully automate any aspect of evidence synthesis product (ESP) production have had their performance evaluated, and how well do these tools perform?

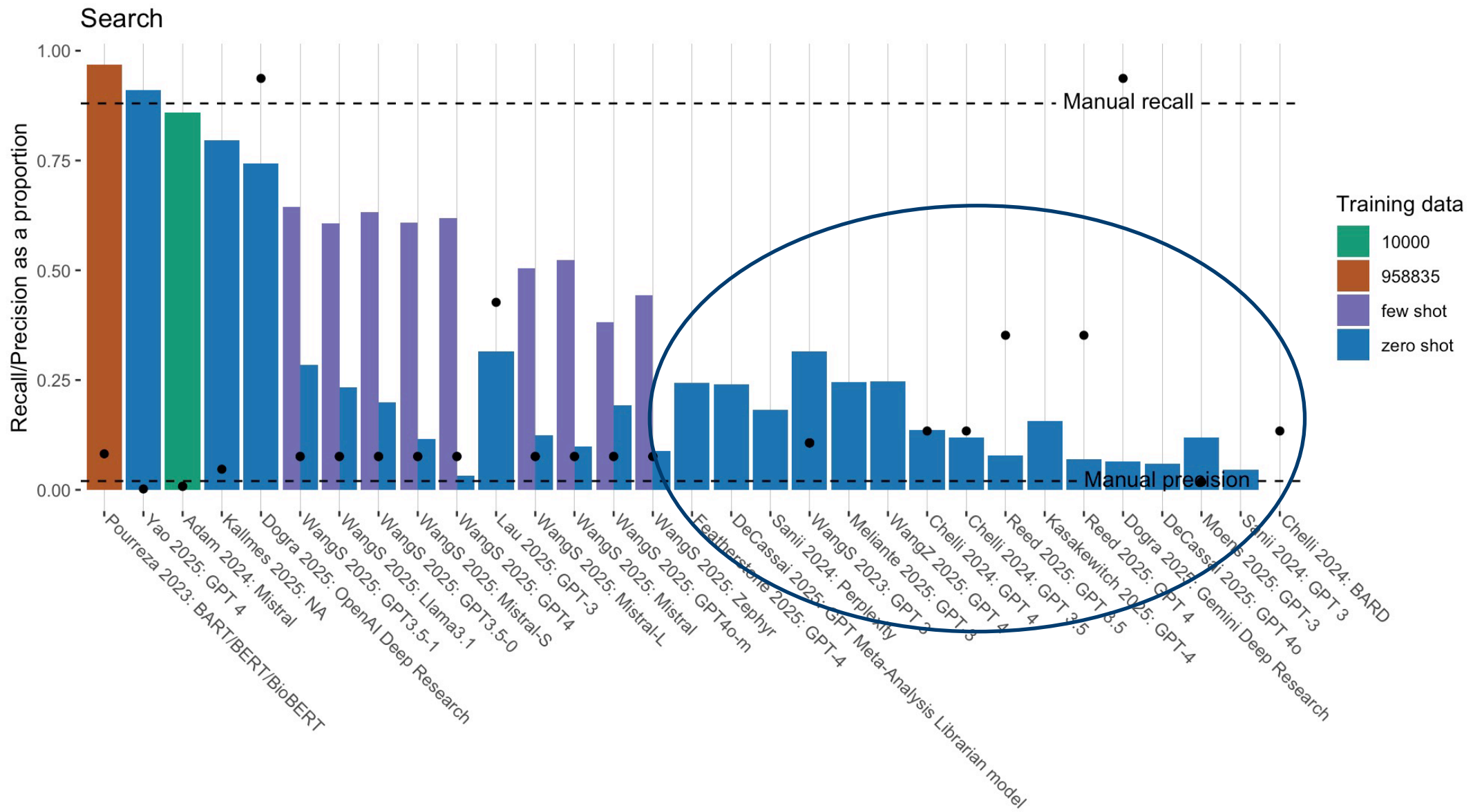
Domain: Any task in ESP production
Tools: Publicly available systems, methods, instruments, models, applications, software, and packages that use AI or ML to semi- or fully automate ESP production
Comparators: Reference standard as defined by the study
Outcomes: Quantitative performance

Standard AHRQ methodology for systematic reviews, as well as EPC guidance for rapid reviews and living reviews. Latest search October 2025.

Literature Identification (25 Studies)

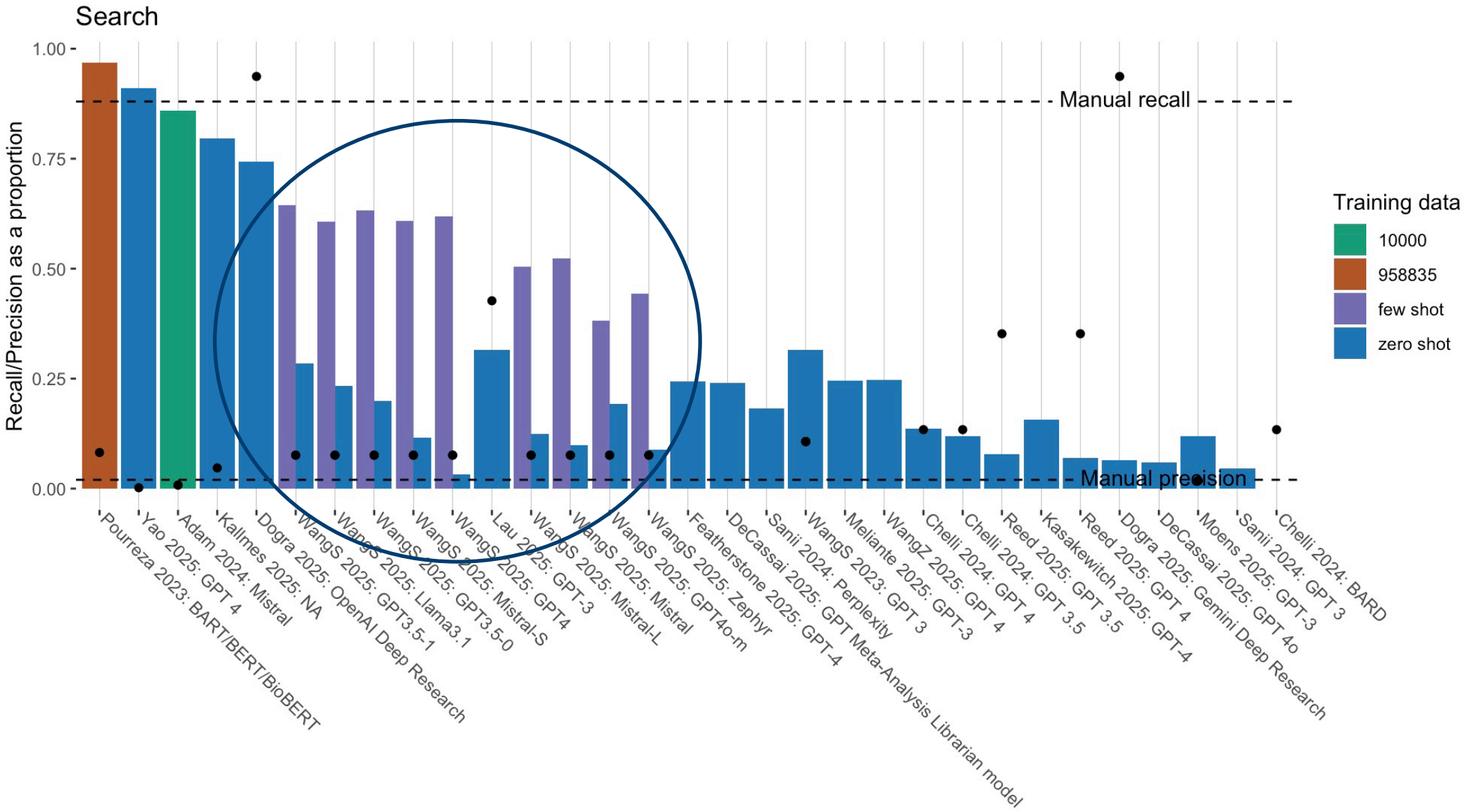
Evidence Synthesis Task	Number of studies (tools)	Comparator	Accuracy measure, median (range)	Burden measure, median (range)
Search	18 (22)	Human-developed search strategies	Recall: 17% (0% to 97%)	Precision: 7% (0% to 94%)
Study design classification	7 (7)	Human identification or PubMed's "publication type" tag	Recall: 98% (96% to 100%)	Precision: 92% (4% to 100%)

Database search



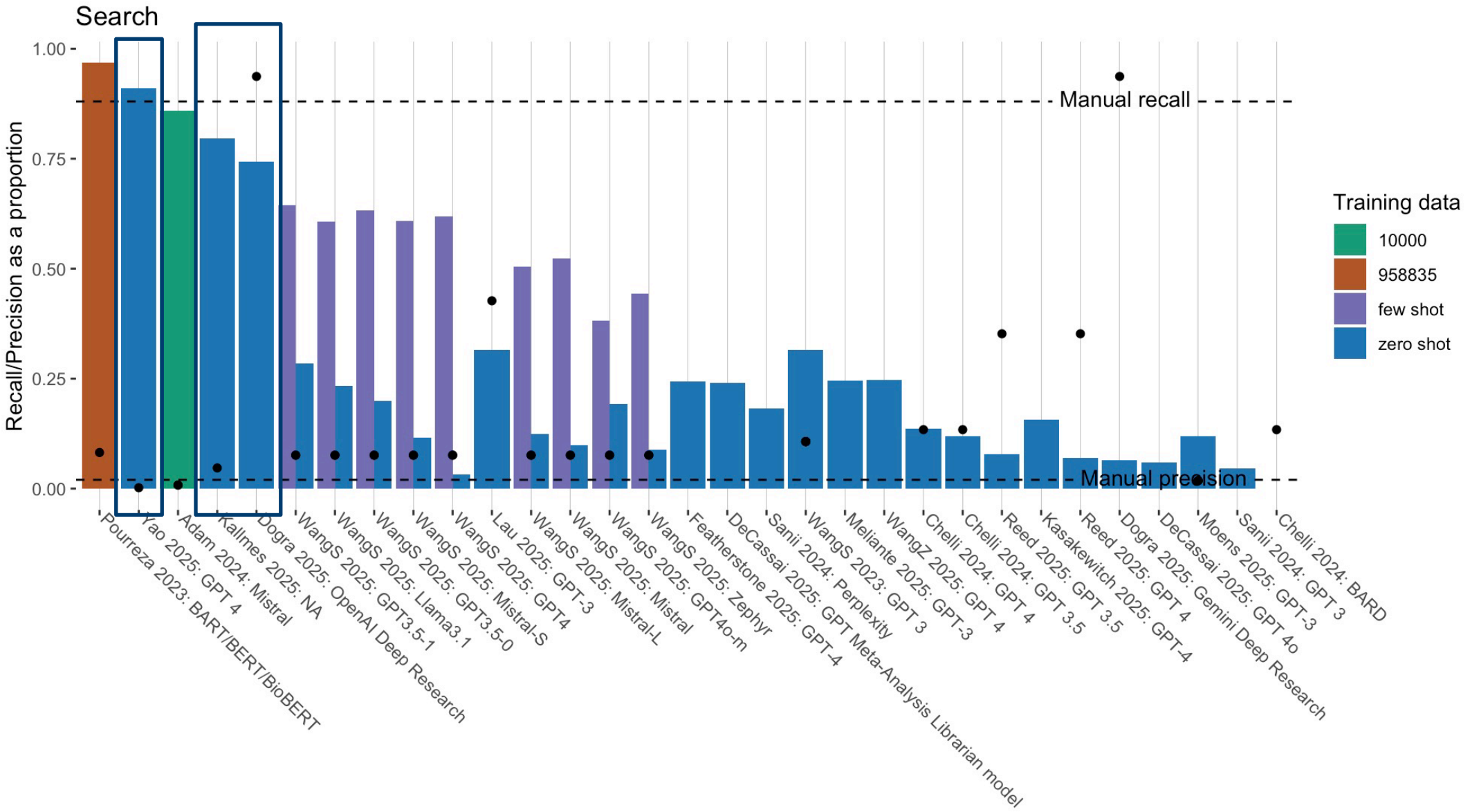
School of Public Health
BROWN UNIVERSITY

Database search



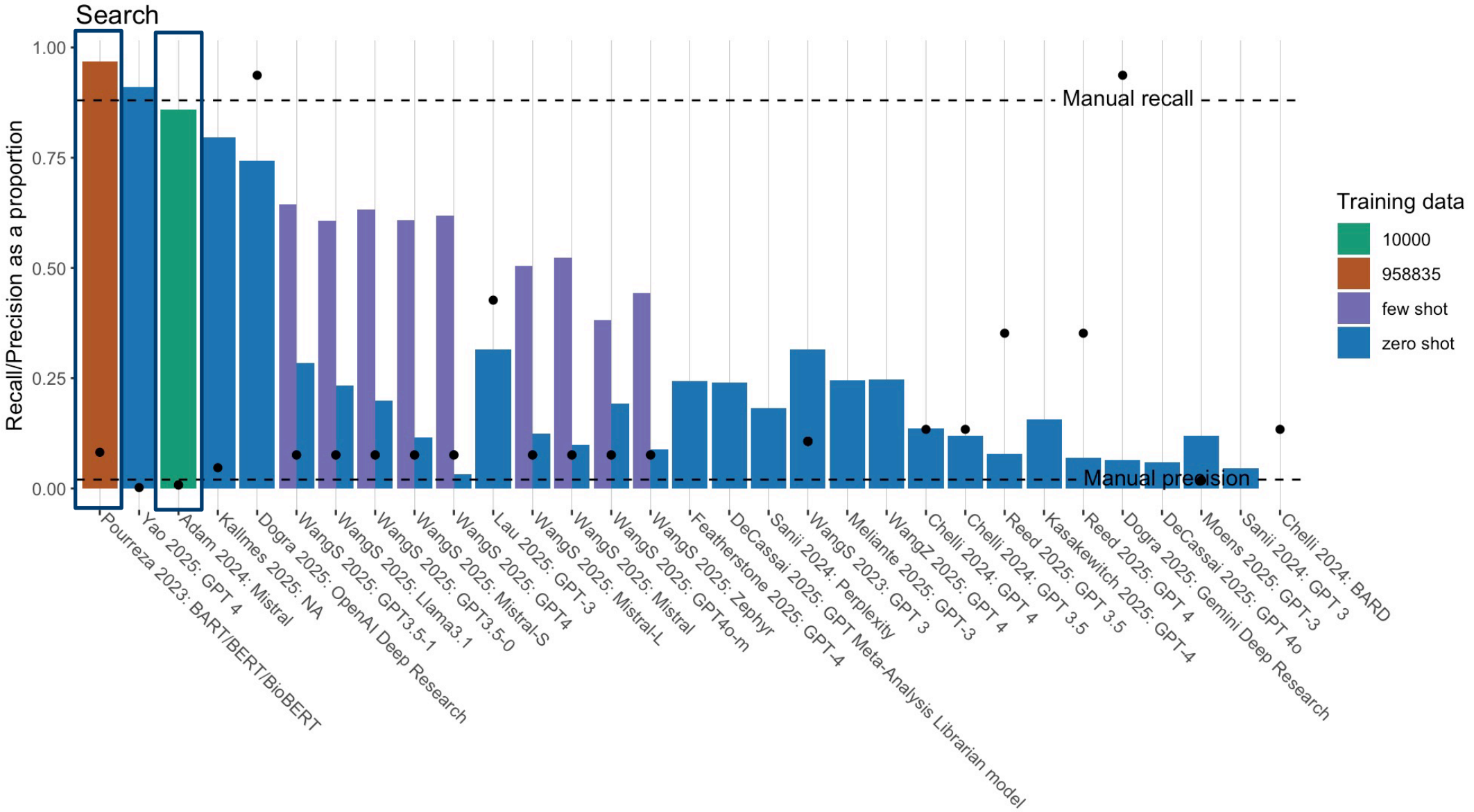
School of Public Health
BROWN UNIVERSITY

Database search



School of Public Health
BROWN UNIVERSITY

Database search



School of Public Health
BROWN UNIVERSITY

Implementation issues

- Platforms are beginning to adopt these tools, making them more accessible.
- But with limitations
- First, we need to get the technology right

Transparency, Replication, Evaluation

- Tools should adhere to the evidence retrieval standards for evidence synthesis
  - Literature searches should be “human evaluable” and “reproduceable”
  - Literature searches should be comprehensive (or at least not biased)

- Search-automation methods can be evaluated, but only retrospectively (based on known included studies)

- Where to from here?