MSR74

PICO Portal

Data extraction from full-text PDFs using Large Language Models for systematic reviews

Eitan Agai[1], Alon Agai[1]

1 - PICO Portal, St. Petersburg, FL, USA

Introduction

Rapid adoption of evidence-based decision-making in medicine, public health, and other fields has fueled a surge in evidence synthesis.

Traditional evidence synthesis methods are time-intensive and resource-demanding.

Advances in Artificial Intelligence (AI) and Large Language Models (LLMs) offer transformative potential to address these challenges by automating key tasks.

Problem and Question

Issues with LLM hallucinations, lack of model transparency, and inconsistent outputs hinder reliability and scalability.

Can a multi-layered and AI-assisted approach to data extraction increase efficiency while maintaining accuracy during systematic review?

Methods

We used an example systematic review with 193 included studies about digital health interventions.

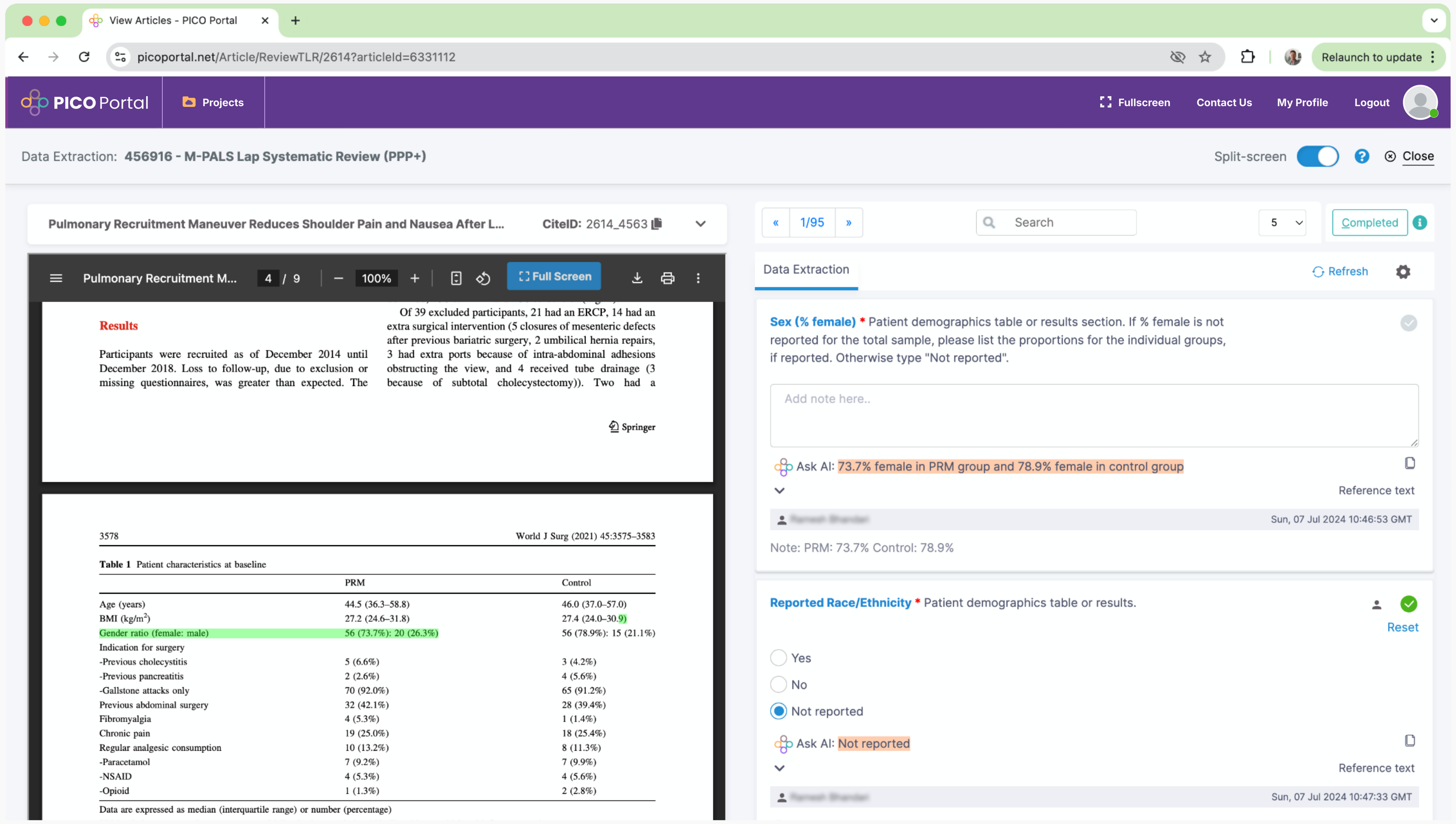
Our workflow parsed each PDF to text, encoded the content, and applied task-specific prompt engineering with OpenAI GPT-4o. The model returned answers and highlighted supporting passages, allowing us to extract 40 data elements such as demographics/study characteristics, intervention details, and outcome data.

Human methodologists extracted the data assisted by the AI suggestions; we checked the accuracy of the AI by comparing the final human response to the AI recommended answer.

Results

Overall, the AI reached 72.3% accuracy in its suggestions, with individual questions ranging from 52.9% to 99.3%. No custom prompt engineering was applied.

Productivity increased from an average of 80 extracted elements per hour for human-only extraction to 195 extracted elements per hour for AI-assisted extraction, an estimated time savings of 59%. This approach also appeared to reduce fatigue, allowing methodologists to work for an additional 1-2 hours without breaks.


View Articles - PICO Portal
picoportal.net/Article/ReviewTLR/2614?articleId=6331112
Relaunch to update
PICO Portal
Projects
Fullscreen
Contact Us
My Profile
Logout

Data Extraction: **456916 - M-PALS Lap Systematic Review (PPP+)**

Pulmonary Recruitment Maneuver Reduces Shoulder Pain and Nausea After L... **CiteID:** 2614_4563

Results

Participants were recruited as of December 2014 until December 2018. Loss to follow-up, due to exclusion or missing questionnaires, was greater than expected. The

Of 39 excluded participants, 21 had an ERCP, 14 had an extra surgical intervention (5 closures of mesenteric defects after previous bariatric surgery, 2 umbilical hernia repairs, 3 had extra ports because of intra-abdominal adhesions obstructing the view, and 4 received tube drainage (3 because of subtotal cholecystectomy)). Two had a

Springer

3578 World J Surg (2021) 45:3575–3583

Table 1 Patient characteristics at baseline

	PRM	Control
Age (years)	44.5 (36.3–58.8)	46.0 (37.0–57.0)
BMI (kg/m^2)	27.2 (24.6–31.8)	27.4 (24.0–30.9)
Gender ratio (female: male)	56 (73.7%): 20 (26.3%)	56 (78.9%): 15 (21.1%)
Indication for surgery		
-Previous cholecystitis	5 (6.6%)	3 (4.2%)
-Previous pancreatitis	2 (2.6%)	4 (5.6%)
-Gallstone attacks only	70 (92.0%)	65 (91.2%)
Previous abdominal surgery	32 (42.1%)	28 (39.4%)
Fibromyalgia	4 (5.3%)	1 (1.4%)
Chronic pain	19 (25.0%)	18 (25.4%)
Regular analgesic consumption	10 (13.2%)	8 (11.3%)
-Paracetamol	7 (9.2%)	7 (9.9%)
-NSAID	4 (5.3%)	4 (5.6%)
-Opioid	1 (1.3%)	2 (2.8%)

Data are expressed as median (interquartile range) or number (percentage)

Sex (% female) * Patient demographics table or results section. If % female is not reported for the total sample, please list the proportions for the individual groups, if reported. Otherwise type "Not reported".

Add note here..

Ask AI: 73.7% female in PRM group and 78.9% female in control group

Reference text

[redacted] Sun, 07 Jul 2024 10:46:53 GMT

Note: PRM: 73.7% Control: 78.9%

Reported Race/Ethnicity * Patient demographics table or results.

Reset

○ Yes
○ No
● Not reported

Ask AI: Not reported

Reference text

[redacted] Sun, 07 Jul 2024 10:47:33 GMT

Figure: AI-assisted data extraction (right) with interactive PDF (left).

Conclusions

LLM technologies, combined with human oversight, have demonstrated the ability to reduce time, costs, and decision fatigue.

This approach represents a scalable solution to accelerate evidence syntheses across diverse fields including clinical research, public health policy making, and education.

Acknowlegments: We would like to acknowledge the EPC team at Minnesota University for allowing us to use their systematic review as an example.example.

REFERENCES [1] Khan et al. 2025 (DOI: 10.1101/2024.09.20.24314108); [2] Konet et al. 2023 (DOI: 10.1002/jrsm.1732); [3] Gartlehner et al. 2025 (DOI: 10.1101/2025.03.20.25324350); [4] Lieberum et al. 2025 (DOI:10.1016/j.jclinepi.2025.111746); [5] Gartlehner et al. 2024 (DOI: 10.1002/jrsm.1710); [6] Khraisha et al. 2024 (DOI: 10.1002/jrsm.1715); [7] Schmidt et al. 2024, ALTARS; [8] James et al. 2023 (DOI: 10.21203/rs.3.rs-3288515/v1); [9] Honghao et al. 2024 (DOI: 10.1001/jamanetworkopen.2024.12687); [10] Tyler et al. 2023 (DOI: 10.1101/2023.11.19.23298727); [11] Jacobsen et al. 2022 (DOI: 10.1186/s12874-022-01649-y); [12] Angelika et al. 2024 (DOI: 10.1101/2024.07.16.24310483).