petaurievidence

ISPOR Europe Research Poster:

Performance of large language model clinical data extraction by data domain: A rapid systematic review

Included records, n=31

Agai E. MSR74 Data Extraction from Full-Text PDFs Using Large Language Models for Systematic Reviews. Value in Health. 2025;28(6):S288.

Aggarwal S, Kumar S, Topaloglu O. MSR174 Leveraging Chat-GPT for Conducting Systematic Literature Reviews. Value in Health. 2024;27(12):S472.

Aronitz E, Habib JB, Olsen C, Hou K, Ferko N. MSR88 Key Considerations in the Use of Large Language Models for Data Extraction in Health Economics and Outcomes Research. Value in Health. 2025;28(6):S291.

Benbow E, et al. MSR68 Variability and Improvements of Answers Generated with Different Versions of Large Language Models. Value in Health. 2024;27(6):S272.

Bravo À, Elissa C, Shalaby N, Atanasov P. MSR35 Is GPT-4o Capable of Automating Detailed Data Extraction for Systematic Literature Reviews (SLRs)? Value in Health. 2024;27(12):S444-5.

Campbell F, et al. Exploring the use of a large language model for data extraction in systematic reviews: a rapid feasibility study. Prague Global Evidence Summit. 2024.

Cribbs K, Baisley W, Blackmore L, Lahue B. MSR9 Can Artificial Intelligence Tools Enhance Data Abstraction During Systematic Literature Reviews? Value in Health. 2024;27(6):S261.

Datta S, et al. MSR103 Optimizing Systematic Literature Reviews in Endometrial Cancer: Leveraging AI for Real-Time Article Screening and Data Extraction in Clinical Trials. Value in Health. 2024;27(6):S279.

Gartlehner G, et al. Data extraction for evidence synthesis using a large language model: A proof-of-concept study. Research Synthesis Methods. 2024;15(4):576-89.

Gartlehner G, et al. AI-Assisted Data Extraction with a Large Language Model: A Study Within Reviews. medRxiv: 2025.2003. 2020.25324350.

Goring S, Goring SJ. PT12 Large-Language Models to Complement and Augment Literature Review: Hi! How Can I Help You? Value in Health. 2025;28(6):S307.

Guerra I, Gallinaro J, Rtveladze K, Lambova A, Asenova E. MSR92 Can Artificial Intelligence (AI) Large Language Models (LLMS) Such as Generative Pre-Trained Transformer (GPT) Be Used to Automate Literature Reviews? Value in Health. 2023;26(12):S410-1.

Hanegraaf P, Mosselman J, de Jong R, Boersma C, van der Schans J. MSR5 Accuracy of Automated Data Extraction for Systematic Literature Reviews. Value in Health. 2023;26(12):S393.

Khan MA, et al. Collaborative large language models for automated data extraction in living systematic reviews. Journal of the American Medical Informatics Association. 2025;32(4):638-47.

Khraisha Q, Put S, Kappenberg J, Warraitch A, Hadfield K. Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. Research Synthesis Methods. 2024;15(4):616-26.

Klijn S, et al. MSR18 Improving the Performance of Generative AI to Achieve 100% Accuracy in Data Extraction. Value in Health. 2024;27(6):S262-3.

Konet A, et al. Performance of two large language models for data extraction in evidence synthesis. Research Synthesis Methods. 2024;15(5):818-24.

Lai H, et al. Language models for data extraction and risk of bias assessment in complementary medicine. NPJ Digital Medicine. 2025;8(1):74.

Lambova A, et al. MSR135 Using Large Language Models (LLMs) for Data Extraction in Literature Reviews: An Enhanced Approach. Value in Health. 2024;27(12):S464.

Lee SH, Chacko A, Yankovsky A. MSR17 Leveraging Large Language Models (LLMs) for Classifying Peer Reviewed Publications for Literature Review. Value in Health. 2024;27(6):S262.

Li Y, et al. MSR234 Enhancing Systematic Literature Reviews With GenAI: Development, Applications, and Performance Evaluation. Value in Health. 2024;27(12):S484.

Li Y, et al. Enhancing systematic literature reviews with generative artificial intelligence: development, applications, and performance evaluation. Journal of the American Medical Informatics Association. 2025;32(4):616-25.

Liu J, et al. AI-driven evidence synthesis: data extraction of randomized controlled trials with large language models. International Journal of Surgery. 2025;111(3):2722-6.

Livieratos A, et al. SA16 Superior Performance of Generative AI After Application of Mixture-of-Agents LLMs in Outcomes Research. Value in Health. 2025;28(6):S398.

Lootus M, et al. MSR41 The Use of Large Language Models for Systematic Literature Review Automation: An Evaluation of Quality and Time Savings. Value in Health. 2025;28(6):S281-2.

Luedtke NF, Shung D. Mo1181 AUTOMATED EXTRACTION OF RANDOMIZED CONTROLLED TRIAL DATA USING LARGE LANGUAGE MODELS: A PILOT STUDY WITH VEDOLIZUMAB META-ANALYSIS. Gastroenterology. 2024;166(5):S-967.

Motzfeldt Jensen M, et al. ChatGPT-4o can serve as the second rater for data extraction in systematic reviews. PloS One. 2025;20(1):e0313401.

Purewal A, Fautsch K, Klasova J, Hussain N, D'Souza RS. Human versus artificial intelligence: evaluating ChatGPT's performance in conducting published systematic reviews with meta-analysis in chronic pain research. Regional Anesthesia & Pain Medicine. 2025;rapm-2024-106358.

Reason T, et al. Artificial intelligence to automate network meta-analyses: four case studies to evaluate the potential application of large language models. PharmacoEconomics Open. 2024;8(2):205-20.

Reason T, et al. P22 Disrupting Health Economics: Automating Network Meta-Analyses With AI and Large Language Models. Value in Health. 2023;26(12):S6.

Schmidt L, et al. Exploring the use of a large language model for data extraction in systematic reviews: a rapid feasibility study. arXiv preprint arXiv:2405.14445. 2024.

Records included for further analyses, n=15

Cribbs K, Baisley W, Blackmore L, Lahue B. MSR9 Can Artificial Intelligence Tools Enhance Data Abstraction During Systematic Literature Reviews? Value in Health. 2024;27(6):S261.

Datta S, et al. MSR103 Optimizing Systematic Literature Reviews in Endometrial Cancer: Leveraging AI for Real-Time Article Screening and Data Extraction in Clinical Trials. Value in Health. 2024;27(6):S279.

Gartlehner G, et al. Data extraction for evidence synthesis using a large language model: A proof-of-concept study. Research Synthesis Methods. 2024;15(4):576-89.

Gartlehner G, et al. AI-Assisted Data Extraction with a Large Language Model: A Study Within Reviews. medRxiv: 2025.2003. 2020.25324350.

Goring S, Goring SJ. PT12 Large-Language Models to Complement and Augment Literature Review: Hi! How Can I Help You? Value in Health. 2025;28(6):S307.

Guerra I, Gallinaro J, Rtveladze K, Lambova A, Asenova E. MSR92 Can Artificial Intelligence (AI) Large Language Models (LLMS) Such as Generative Pre-Trained Transformer (GPT) Be Used to Automate Literature Reviews? Value in Health. 2023;26(12):S410-1.

Khan MA, et al. Collaborative large language models for automated data extraction in living systematic reviews. Journal of the American Medical Informatics Association. 2025;32(4):638-47.

Konet A, et al. Performance of two large language models for data extraction in evidence synthesis. Research Synthesis Methods. 2024;15(5):818-24.

Lai H, et al. Language models for data extraction and risk of bias assessment in complementary medicine. NPJ Digital Medicine. 2025;8(1):74.

Lambova A, et al. MSR135 Using Large Language Models (LLMs) for Data Extraction in Literature Reviews: An Enhanced Approach. Value in Health. 2024;27(12):S464.

Lee SH, Chacko A, Yankovsky A. MSR17 Leveraging Large Language Models (LLMs) for Classifying Peer Reviewed Publications for Literature Review. Value in Health. 2024;27(6):S262.

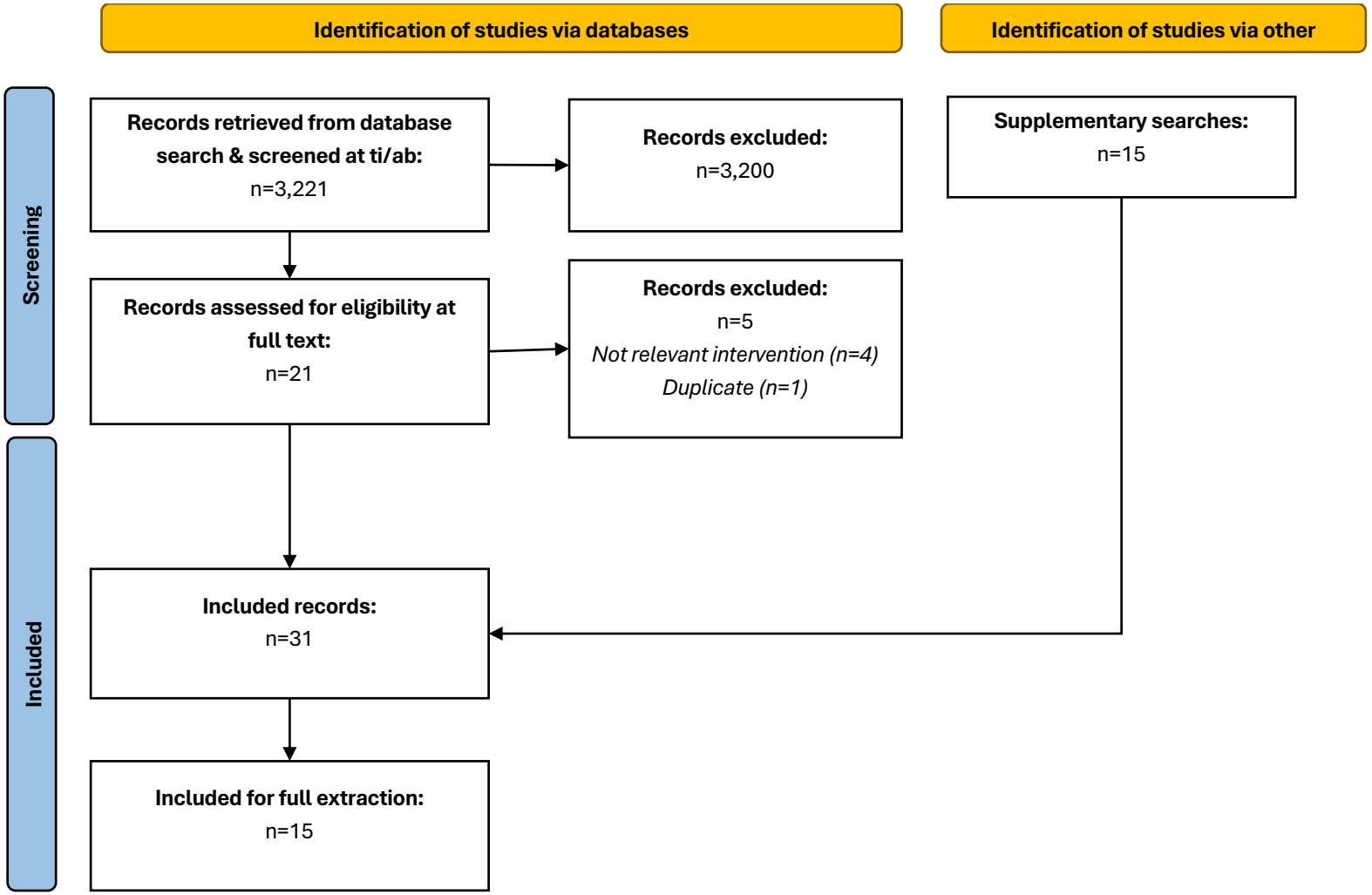
Li Y, et al. Enhancing systematic literature reviews with generative artificial intelligence: development, applications, and performance evaluation. Journal of the American Medical Informatics Association. 2025;32(4):616-25.

Liu J, et al. AI-driven evidence synthesis: data extraction of randomized controlled trials with large language models. International Journal of Surgery. 2025;111(3):2722-6.

Lootus M, et al. MSR41 The Use of Large Language Models for Systematic Literature Review Automation: An Evaluation of Quality and Time Savings. Value in Health. 2025;28(6):S281-2.

Motzfeldt Jensen M, et al. ChatGPT-4o can serve as the second rater for data extraction in systematic reviews. PloS One. 2025;20(1):e0313401.

PRISMA



Abbreviations: ti/ab, title and abstract