petaurievidence

ISPOR Europe Research Poster:

Traditional versus generative AI: A rapid systematic review assessing accuracy and efficiency of AI in title/abstract screening

Included studies (n=88)

1. Abogunrin S, Sieiro RR, Lane M. Generative artificial intelligence: an effective alternative for screening titles and abstracts in systematic literature reviews. Presented at ISPOR Europe 2024 [Available from: https://www.ispor.org/heor-resources/presentations-database/presentation/euro2024-4018/144789]. Value in Health. 2024;27(12):S2.
2. Aggarwal S, Kumar S, Topaloglu O. Leveraging Chat-GPT for conducting systematic literature reviews. Presented at ISPOR Europe 2024 [Available from: https://www.ispor.org/heor-resources/presentations-database/presentation/euro2024-4018/144789]. Value in Health. 2024;27(12):S2.
3. Akinseloyin O, Jiang X, Palade V. A question-answering framework for automated abstract screening using large language models. Journal of the American Medical Informatics Association. 2024;31(9):1939-52.
4. Attri S, Kaur R, Singh B, Rai P. MSR57 Transforming Systematic Literature Reviews: Unleashing the Potential of GPT-4: A Cutting-Edge Large Language Model, to Elevate Research Synthesis. Value in Health. 2024;27(6 Supplement):S270.
5. Bhagat A, Moon D, Khan H, Kochar P, Kanakagiri S, Kaur R, et al. MSR42 Adoption of Artificial Intelligence in Systematic Reviews. Value in Health. 2022;25(12 Supplement):S358.
6. Borowiack E, Sadowska E, Nowak A, Brozek J. MSR61 AI Support Reduced Screening Burden in a Systematic Review with Costs and Cost-Effectiveness Outcomes (SR-CCEO) for Cost-Effectiveness Modeling. Value in Health. 2023;26(6 Supplement):S288.
7. Burns JK, Etherington C, Cheng-Boivin O, Boet S. Using an artificial intelligence tool can be as accurate as human assessors in level one screening for a systematic review. Health information and libraries journal. 2021;18.
8. Cao C, Sang J, Kloosterman R, Cecere M, Gorla J, Saleh R, et al. Prompting is all you need: LLMs for systematic review screening. medRxiv. 2024;03.
9. Carey N, Harte M, Mc Cullagh L. A text-mining tool generated title-abstract screening workload savings: performance evaluation versus single-human screening. J Clin Epidemiol. 2022;149:53-9.
10. Chai KEK, Lines RLJ, Gucciardi DF, Ng L. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Systematic Reviews. 2021;10(1) (no pagination)(93).
11. Chen J, Mudumba R, Morriss J, Padula W. EE592 Performance Evaluation of Explainable Artificial Intelligence to Replicate a Systematic Review and Meta-Analysis of Cost-Effectiveness Analysis Findings. Value in Health. 2024;27(12 Supplement):S171-S2.

12. Ciapponi A, Bardach A, Glujovsky D, Tarchard R. The efficiency of “Nested Knowledge” to facilitate the conduction of systematic reviews and meta-analyses. Presented at Global Evidence Summit 2024. [Available from: https://abstracts.cochrane.org/2024-prague-global-evidence-summit/efficiency-nested-knowledge-facilitate-conduction-systematic]. 2024.
13. Cichewicz A, Burnett H, Huelin R, Kadambi A. SA3 Utility of Artificial Intelligence in Systematic Literature Reviews for Health Technology Assessment Submissions. Value in Health. 2022;25(7 Supplement):S604.
14. Cichewicz A, Kadambi A, Lavoie L, Mittal L, Pierre V, Raorane R. MSR67 Automation of Abstract Screening: A Case Study of Multiple Systematic Literature Reviews. Value in Health. 2022;25(12 Supplement):S362.
15. Cichewicz A, Kadambi A, Lavoie L, Mittal L, Pierre V, Raorane R. MSR41 Application of Artificial Intelligence As a Decision Support Tool for Abstract Screening: Implications for Time and Cost Savings. Value in Health. 2023;26(6 Supplement):S285.
16. Cichewicz A, Pande A, Borkowska K, Mittal L, Wittkopf P, Slim M. MSR22 Automating Systematic Literature Review (SLR) Updates: A Comparative Validation Study of Artificial Intelligence (AI) Versus Human Screeners. Value in Health. 2024;27(6):S263.
17. Cichewicz A, Pande A, Casanas i Comabella C, Mittal L, Slim M. MSR136 Feasibility of GPT-4-Based Content Extraction to Identify Eligible Titles and Abstracts in a Systematic Literature Review (SLR). Value in Health. 2024;27(12 Supplement):S464-S5.
18. Cichewicz A, Slim M, Deshpande S. MSR163 Artificial Intelligence (AI)-Based Screening: Exploration of Differences in Two Health Technology Assessment (HTA)-Compliant Systematic Literature Reviews (SLRS). Value in Health. 2023;26(12 Supplement):S425.
19. Dai ZY, Wang FQ, Shen C, Ji YL, Li ZY, Wang Y, et al. Accuracy of Large Language Models for Literature Screening in Thoracic Surgery: Diagnostic Study. Journal of Medical Internet Research. 2025;27(no pagination).
20. Datta S, Lee K, Paek H, Mojarad MR, Prabhu V, Zhang J, 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 Supplement):S279.
21. Dennstadt F, Zink J, Putora PM, Hastings J, Cihoric N. Title and abstract screening for literature reviews using large language models: an exploratory study in the biomedical domain. Systematic Reviews. 2024;13(1) (no pagination).
22. DeSilva DTN, Moore BR, Strunk T, Petrovski M, Varis V, Chai K, et al. Development of a pharmaceutical science systematic review process using a semi-automated machine learning tool: Intravenous drug compatibility in the neonatal intensive care setting. Pharmacology Research and Perspectives. 2024;12(1) (no pagination)(e1170).
23. dos Reis AHS, de Oliveira ALM, Fritsch C, Zouch J, Ferreira P, Polese JC. Usefulness of machine learning softwares to screen titles of systematic reviews: a methodological study. Systematic Reviews. 2023;12(1) (no pagination)(68).
24. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, et al. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Systematic Reviews. 2019;8(1):277.
25. Gates A, Gates M, DaRosa D, Elliott SA, Pillay J, Rahman S, et al. Decoding semi-automated title-abstract screening: findings from a convenience sample of reviews. Systematic Reviews. 2020;9(1) (no pagination)(272).

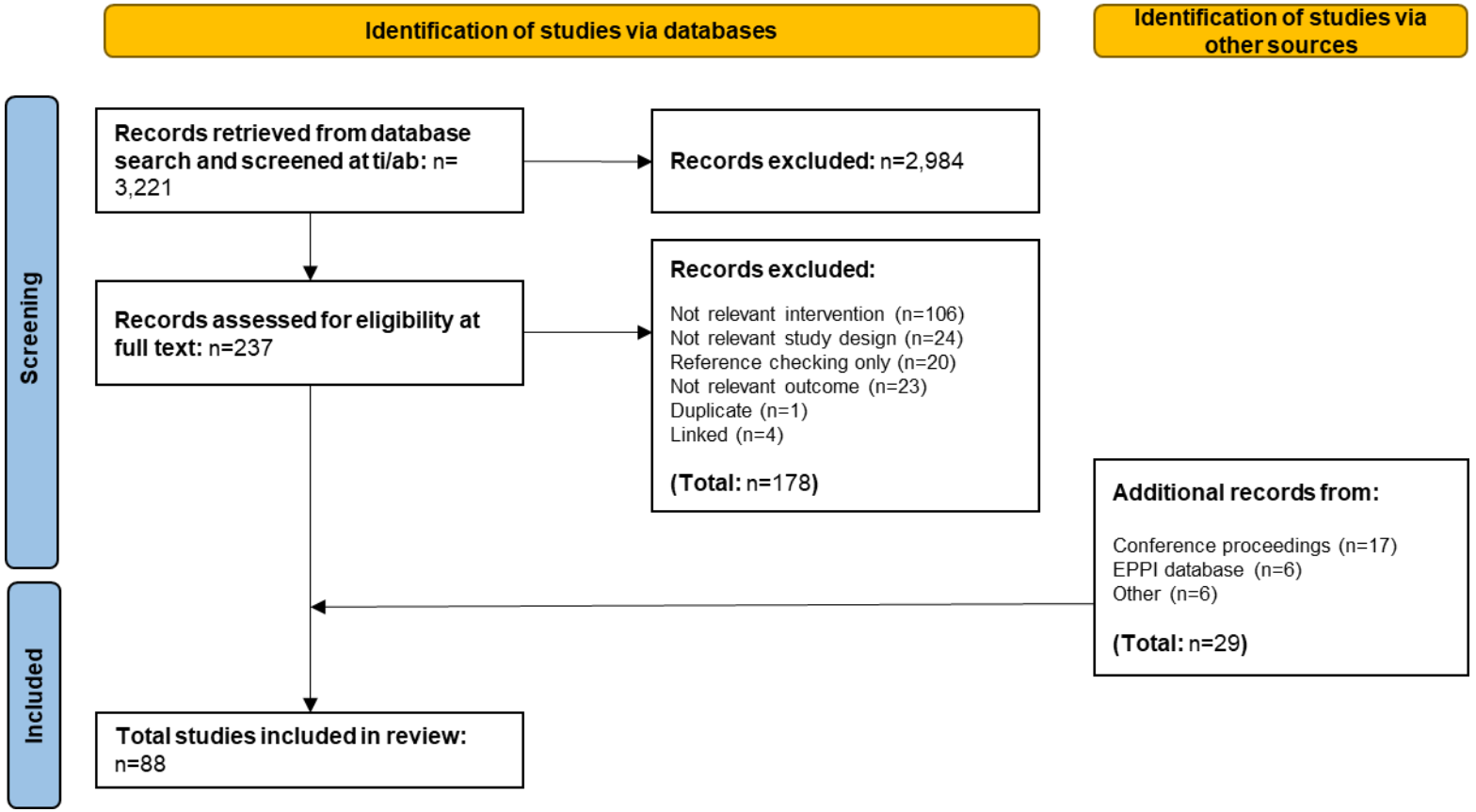
26. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: A comparative evaluation of three tools. Systematic Reviews. 2019;8(1) (no pagination)(278).
27. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr machine learning tool. Systematic Reviews. 2018;7(1) (no pagination)(45).
28. Giles L, Miles G, Sibbring GC. MSR210 Artificial Intelligence (AI) Is Different: Is It Time to Update Systematic Literature Review (SLR) Workflows? Value in Health. 2024;27(12 Supplement):S480.
29. Guo E, Gupta M, Deng J, Park YJ, Paget M, Naugler C. Automated Paper Screening for Clinical Reviews Using Large Language Models: Data Analysis Study. Journal of Medical Internet Research. 2024;26(1) (no pagination)(e48996).
30. Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening - impact on reviewer-relevant outcomes. BMC medical research methodology. 2020;20(1):256.
31. Harmsen W, de Groot J, Harkema A, van Dusseldorp I, De Bruin J, Van den Brand S, et al. Artificial intelligence supports literature screening in medical guideline development: towards up-to-date medical guidelines. (V1.0). Zenodo. 2021.
32. Holland C, Oakes DB, Sarna M, Chai KE, Ng L, Moore HC. Validity of using a semi-automated screening tool in a systematic review assessing non-specific effects of respiratory vaccines. BMC medical research methodology. 2025;25(1):59.
33. Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, et al. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environment International. 2020;138(no pagination)(105623).
34. Hudgens S, Lloyd-Price L, Jafar R, Nourizade M, Burbridge C, Thorlund K. CO83 Screening Articles in a Qualitative Literature Review Using Large Language Models: A Comparison of GPT Versus Open Source, Trained Models Against Expert Researcher Screening. Value in Health. 2024;27(6):S32.
35. Issaiy M, Ghanaati H, Kolahi S, Shakiba M, Jalali AH, Zarei D, et al. Methodological insights into ChatGPT's screening performance in systematic reviews. BMC medical research methodology. 2024;24(1):78.
36. Kamra S, Hyderboini R, Sirumalla Y, Venkateswara Rao J, Chidirala S, Dabral S, et al. MSR70 Pilot Study to Evaluate Efficiency of DISTILLERSR®'S Artificial Intelligence (AI) Tool over Manual Screening Process in Literature Review. Value in Health. 2022;25(7):S532.
37. Kataoka Y, So R, Banno M, Kumasawa J, Someko H, Taito S, et al. Development of meta-prompts for Large Language Models to screen titles and abstracts for diagnostic test accuracy reviews. medRxiv. 2023;01.
38. Kaur R, Rai P, Attri S, Kaur G, Singh DB. MSR15 Revolutionizing Systematic Literature Reviews: Harnessing the Power of Large Language Model (GPT-4) for Enhanced Research Synthesis. Value in Health. 2024;27:S262.
39. Kishore A, Goel A, Sai V, Anam S, Hughes R, Chawla R. MSR88 Validation of a Programmed Probability Score Using the Screening Tool, Sympro to Assist Selection of Studies in Systematic Literature Reviews. Value in Health. 2023;26(12 Supplement):S410.
40. Krag CH, Balschmidt T, Bruun F, Brejnebol M, Xu JJ, Boesen M, et al. Large language models for abstract screening in systematic- and scoping reviews: A diagnostic test accuracy study. medRxiv. 2024;02.

41. Landschaft A, Antweiler D, Mackay S, Kugler S, Ruping S, Wrobel S, et al. Implementation and evaluation of an additional GPT-4-based reviewer in PRISMA-based medical systematic literature reviews. International Journal of Medical Informatics. 2024;189(no pagination).
42. 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.
43. Li J, Kabouji J, Bouhadoun S, Tanveer S, Filion KB, Gore G, et al. Sensitivity and specificity of alternative screening methods for systematic reviews using text mining tools. Journal of Clinical Epidemiology. 2023;162:72-80.
44. Li M, Sun J, Tan X. Evaluating the effectiveness of large language models in abstract screening: a comparative analysis. Systematic Reviews. 2024;13(1) (no pagination).
45. Li Y, Datta S, Rastegar-Mojarad M, Lee K, Paek H, Glasgow J, et al. Enhancing systematic literature reviews with generative artificial intelligence: development, applications, and performance evaluation. Journal of the American Medical Informatics Association : JAMIA. 2025;27.
46. Liu JJW, Ein N, Gervasio J, Easterbrook B, Nouri MS, Nazarov A, et al. Usability and agreement of the SWIFT-ActiveScreener systematic review support tool: Preliminary evaluation for use in clinical research. PLoS ONE. 2024;19(11) (no pagination).
47. Luo R, Sastimoglu Z, Faisal AI, Deen MJ. Evaluating the Efficacy of Large Language Models for Systematic Review and Meta-Analysis Screening. medRxiv. 2024;04.
48. Matsui K, Utsumi T, Aoki Y, Maruki T, Takeshima M, Takaesu Y. Human-Comparable Sensitivity of Large Language Models in Identifying Eligible Studies Through Title and Abstract Screening: 3-Layer Strategy Using GPT-3.5 and GPT-4 for Systematic Reviews. J Med Internet Res. 2024;26:e52758.
49. Metcalf T, Dodd O, Peatman J, Kiessling J, O'Donovan P, Heron L. MSR222 Training Artificial Intelligence for Literature Reviews: Can a Classifier Match a Human Reviewer? Value in Health. 2024;27(12 Supplement):S482.
50. Miles G, Giles L, Kerr BC, Norman B, Sibbring GC. MSR201 Optimising Performance of Generative Artificial Intelligence (GenAI) in Systematic Literature Review (SLR) Screening Using PICOS Criteria. Value in Health. 2024;27(12 Supplement):S478.
51. Mizuno A, Ota E. The Efficacy of GPT-4 in Systematic Reviews within the Cardiovascular Disease Domain. Presented at Global Evidence Summit 2024 [Available from: https://abstracts.cochrane.org/2024-prague-global-evidence-summit/efficacy-gpt-4-systematic-reviews-within-cardiovascular-disease]. 2024.
52. Muthu S. The efficiency of machine learning-assisted platform for article screening in systematic reviews in orthopaedics. International Orthopaedics. 2023;47(2):551-6.
53. Nass P, Rekowska D, Arca E, Nowak A, Sadowska E, Borowiack E, et al. Are the Machines Ready to Take Over? Can Artificial Intelligence Replace a Human Reviewer for Literature Screening and Selection for Systematic Literature Reviews? Value in Health. 2024;27(12 Supplement):S633.
54. Ng J, Szydlowski N, Gill M, Fusco N, Ruiz K. MSR72 An Evaluation of the Rayyan Artificial Intelligence Tool for Systematic Literature Review Screening. Value in Health. 2024;27(6):S273.
55. Oami T, Okada Y, Nakada T-a. Performance of a Large Language Model in Screening Citations. JAMA Network Open. 2024;7(7):e2420496-e.
56. Orel E, Ciglenecki I, Thiabaud A, Temerev A, Calmy A, Keiser O, et al. LiteRev, an Automation Tool to Support Literature Reviews: A Case Study on Acute and Early HIV Infection in Sub-Saharan Africa. medRxiv. 2023:2023.02.20.23286179.

57. Oude Wolcherink MJ, Pouwels XGLV, van Dijk SHB, Doggen CJM, Koffijberg H. Can artificial intelligence separate the wheat from the chaff in systematic reviews of health economic articles? Expert Review of Pharmacoeconomics and Outcomes Research. 2023;23(9):1049-56.
58. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Systematic Reviews. 2016;5(1) (no pagination)(210).
59. Pijls BG. Machine Learning assisted systematic reviewing in orthopaedics. Journal of Orthopaedics. 2024;48:103-6.
60. 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 and Pain Medicine. 2025;(no pagination).
61. Qureshi R, Robinson K, Butler M, Agai E. Machine-learning assisted screening increases efficiency of systematic review. Presented at Cochrane Colloquium 2023 [Available from: https://abstracts.cochrane.org/2024-prague-global-evidence-summit/efficacy-gpt-4-systematic-reviews-within-cardiovascular-disease]. 2023.
62. Rai P, Pandey S, Attri S, Singh B, Kaur R. MSR78 Advancing Systematic Literature Reviews: A Comparative Analysis of Large Language Models (Claude Sonnet 3.5, Gemini Flash 1.5, and GPT-4) in the Automation Era of Generative AI. Value in Health. 2024;27(12):S453.
63. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Systematic Reviews. 2015;4(1) (no pagination)(80).
64. Rathi H, Malik A, Behera DC, Kamboj G. A Comparative Analysis of Large Language Models (LLM) Utilised in Systematic Literature Review. Value in Health. 2023;26(12 Supplement):S6.
65. Reason T, Langham J, Malcolm B, Klijn S, Gimblett A. MSR46 Breaking Through Limitations: Enhanced Systematic Literature Reviews With Large Language Models. Value in Health. 2023;26(12 Supplement):S402.
66. Reddy SM, Patel S, Weyrich M, Fenton J, Viswanathan M. Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence. Systematic Reviews. 2020;9(1) (no pagination)(243).
67. Robinson A, Thorne W, Wu B, Pandor A, Essat M, Stevenson M, et al. Bio-SIEVE: Exploring Instruction Tuning Large Language Models for Systematic Review Automation2023.
68. Royer J, Wu EQ, Ayyagari R, Parravano S, Pathare U, Kisielinska M. MSR131 Prospects for Automation of Systemic Literature Reviews (SLRs) With Artificial Intelligence and Natural Language Processing. Value in Health. 2023;26(12 Supplement):S418.
69. Sanii RY, Kasto JK, Wines WB, Mahylis JM, Muh SJ. Utility of Artificial Intelligence in Orthopedic Surgery Literature Review: A Comparative Pilot Study. Orthopedics. 2024;47(3):e125-e30.
70. Sertel A, Samur S, Yildirim I, Ayer T, Chhatwal J. MSR44 How Accurate Are Large Language Models for Abstract Screening in Systematic Literature Reviews? Value in Health. 2024;27(6):S267.
71. Singh B, Kaur R, Rai P. MSR140 Empowering Systematic Literature Reviews: Utilizing Generative AI for Comprehensive Literature Screening From Titles and Abstracts to Full-Text. Value in Health. 2024;27(12 Supplement):S465.
72. Strachan JA. Designing GTP3 prompts to screen articles for systematic reviews of RCTs. Health Policy and Technology. 2025;14(1) (no pagination).
73. Syriani E, Dávid I, Kumar G. Assessing the Ability of ChatGPT to Screen Articles for Systematic Reviews. ArXiv. 2023;abs/2307.06464.

74. Thurnham J, Kallmes K, Holub KJ. MSR91 Assessing Recall in Abstract Screening: Artificial Intelligence Vs. Human Reviewers. Value in Health. 2024;27(6):S277.
75. Trad F, Charafeddine J, Chkahtoura M, Rahme M, Fuleihan GEH, Chehab A. Streamlining Systematic Reviews in Medical Research: A Novel Application of Large Language Models. Journal of Bone and Mineral Research. 2024;39(Supplement 1):197.
76. Tran VT, Gartlehner G, Yaacoub S, Boutron I, Schwingshackl L, Stadelmaier J, et al. Sensitivity and Specificity of Using GPT-3.5 Turbo Models for Title and Abstract Screening in Systematic Reviews and Meta-analyses. Annals of Internal Medicine. 2024;177(6):791-9.
77. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: Comparative performance of Abstrackr and EPPI-Reviewer. Systematic Reviews. 2020;9(1) (no pagination)(73).
78. Valizadeh A, Moassefi M, Nakhostin-Ansari A, Hosseini Asl SH, Saghab Torbati M, Aghajani R, et al. Abstract screening using the automated tool Rayyan: results of effectiveness in three diagnostic test accuracy systematic reviews. BMC Medical Research Methodology. 2022;22(1):160.
79. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence. 2021;3(2):125-33.
80. van der Pol JA, Huizinga TWJ, Bergstra SA. Is AI-assisted active learning software able to reliably speed-up systematic literature reviews in rheumatology? A real-time comparison of AI-assisted and manual abstract selection. RMD Open. 2024;10(4) (no pagination).
81. van Dijk SHB, Brusse-Keizer MGJ, Bucsán CC, van der Palen J, Doggen CJM, Lenferink A. Artificial intelligence in systematic reviews: promising when appropriately used. BMJ Open. 2023;13(7):e072254.
82. Venkata SK, Velicheti S, Jamdade V, Ranganathan S, Achra M, Banerjee KK, et al. MSR99 Application of Artificial Intelligence in Literature Reviews. Value in Health. 2023;26(12 Supplement):S412.
83. Wang S, Scells H, Zhuang S, Potthast M, Koopman B, Zuccon G, editors. Zero-Shot Generative Large Language Models for Systematic Review Screening Automation. Advances in Information Retrieval; 2024 2024//; Cham: Springer Nature Switzerland.
84. Wilson L, Robinson KA. Streamlining systematic reviews: Using machine learning to enhance screening efficiency. Presented at Cochrane Colloqium 2024 [Available from: https://abstracts.cochrane.org/2024-prague-global-evidence-summit/streamlining-systematic-reviews-using-machine-learning-enhance]. 2024.
85. Yammine R, Chakhtoura MT, Rahme M, Ajjour SI, Chehab A, El-Hajj Fuleihan G. Performance of a Machine Learning Model in the Selection of Articles for Systematic Reviews. Journal of Bone and Mineral Research. 2024;39(Supplement 1):333.
86. Zdorovtsova N, Castanon A, Marsland A, Bray BD. Automating Systematic Literature Review Abstract Screening Using Large Language AI Models: A Development and Validation Study. Value in Health. 2024;27(12 Supplement):S616.
87. Zhan J, Suvada K, Xu M, Tian W, Cara KC, Wallace TC, et al. Accelerating the pace and accuracy of systematic reviews using AI: a validation study. medRxiv. 2024;11.
88. Zhao W, Lai H, Ge L. Enhancing Systematic Review Efficiency: A Feasibility Study of Large Language Models in Title and Abstract Screening. Presented at Global Evidence Summit 2024 [Available from: https://abstracts.cochrane.org/2024-prague-global-evidence-summit/enhancing-systematic-review-efficiency-feasibility-study-large]. 2024.

PRISMA



Abbreviations: ti/ab, title and abstract.