

Generative AI and Large Language Models in Reducing Medication Related Harm and Adverse Drug Events – A Scoping Review

Jasmine Chiat Ling Ong^{1,2,3*}, Chen Michael^{3*}, Ning Ng⁴, Kabilan Elangovan^{4,5}, Nichole Yue Ting Tan³, Liyuan Jin^{3,5}, Qihuang Xie⁶, Daniel Shu Wei Ting^{4,5,7}, Rosa Rodriguez-Monguio^{8,9}, David W. Bates^{10,11}, Nan Liu^{12,13,14,15+}

1. Division of Pharmacy, Singapore General Hospital, Singapore
2. Department of Pharmacy, University of California San Francisco, California, USA
3. Duke-NUS Medical School, Singapore, Singapore
4. Singapore Health Services, Artificial Intelligence Office, Singapore
5. Singapore National Eye Centre, Singapore Eye Research Institute, Singapore, Singapore
6. National University of Singapore, School of Pharmacy, Singapore, Singapore
7. Byers Eye Institute, Stanford University, United States of America
8. Department of Clinical Pharmacy, School of Pharmacy, University of California San Francisco, California, USA
9. Medication Outcomes Center, University of California San Francisco, California, USA
10. Harvard T H Chan School of Public Health, Boston, MA, USA
11. Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA
12. Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore
13. Duke-NUS AI + Medical Sciences Initiative, Duke-NUS Medical School, Singapore, Singapore
14. Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
15. Institute of Data Science, National University of Singapore, Singapore, Singapore

*Contributed Equally

+Corresponding Author

Correspondence to:

Nan Liu, Centre for Quantitative Medicine, Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore

Email: liu.nan@duke-nus.edu.sg

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Background: Medication-related harm has a significant impact on global healthcare costs and patient outcomes, accounting for deaths in 4.3 per 1000 patients. Generative artificial intelligence (GenAI) has emerged as a promising tool in mitigating risks of medication-related harm. In particular, large language models (LLMs) and well-developed generative adversarial networks (GANs) showing promise for healthcare related tasks. This review aims to explore the scope and effectiveness of generative AI in reducing medication-related harm, identifying existing development and challenges in research.

Methods: We searched for peer reviewed articles in PubMed, Web of Science, Embase, and Scopus for literature published from January 2012 to February 2024. We included studies focusing on the development or application of generative AI in mitigating risk for medication-related harm during the entire medication use process. We excluded studies using traditional AI methods only, those unrelated to healthcare settings, or concerning non-prescribed medication uses such as supplements. Extracted variables included study characteristics, AI model specifics and performance, application settings, and any patient outcome evaluated.

Findings: A total of 2203 articles were identified, and 14 met the criteria for inclusion into final review. We found that generative AI and large language models were used in a few key applications: drug-drug interaction identification and prediction; clinical decision support and pharmacovigilance. While the performance and utility of these models varied, they generally showed promise in areas like early identification and classification of adverse drug events and support in decision-making for medication management. However, no studies tested these models prospectively, suggesting a need for further investigation into the integration and real-world application of generative AI tools to improve patient safety and healthcare outcomes effectively.

Interpretation: Generative AI shows promise in mitigating medication-related harms, but there are gaps in research rigor and ethical considerations. Future research should focus on creation of high-quality, task-specific benchmarking datasets for medication safety and real-world implementation outcomes.

Introduction

Medication-related harm poses significant health and economic burden globally. The global prevalence of medication related harm was 12%, of which 15% was severe and fatal, causing a mortality rate of up to 4.3 per 1000 patients.¹ In contrast, cardiovascular deaths caused by ischemic heart disease accounted for 1.09 deaths per 1000 patients.² The economic burden of medication-related harm is estimated at \$30.1 billion and 79 billion euros in United States and Europe respectively.³ Medication-related harm, also termed as adverse drug events (ADEs) include preventable or non-preventable harm caused by interventions related to medication use.⁴ Preventable medication error can occur at any step from the physician prescribing medications to the patient receiving the medication. In turn, ADEs are under active surveillance during healthcare delivery to patients with health systems or pharmacovigilance activities.



Figure 1. Archetypical care delivery process and potential points of error.

Advances in artificial intelligence (AI), digitization of health records, and accessibility to electronic patient records have been shown to reduce the occurrence, duration, and severity of ADEs.⁵⁻⁷ An AI-powered system has been reported to reduce inpatient prescribing error by up to 20%.⁸ However, traditional predictive models are still limited by the lack of in-depth clinical reasoning, poor interoperability in electronic health record systems (EHRs), difficulty in detecting rare events or interactions, and paucity of models that leverage unstructured data. Based on existing system, overlooked ADEs would lead to significant healthcare complications while trivial or clinical insignificant effects are over emphasized leading to healthcare administrative burden. Thus, with significant promises to address such unbalanced issue, generative AI (GenAI) and large language models (LLMs) may enable novel approaches previously unfeasible with conventional methods. For instance, preliminary studies have explored the potential of ChatGPT to recognize adverse drug reactions⁹, pharmacovigilance signal detection¹⁰, and automated medication chart review.^{11,12}

This systematic review summarizes the breadth and depth of existing literature on how generative AI have been utilized to reduce ADEs and highlights areas for future investigation.

Concepts	Definitions
Medication related harm	The harm caused by medication if taken incorrectly, monitored insufficiently or as the result of an error, accident or communication problem. ¹³
Adverse drug event	Any injury resulting from medical interventions related to a drug. This includes both adverse drug reactions in which no error occurred and complications resulting from medication errors. ¹⁴
Pharmacovigilance	The science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem. ¹⁵

Generative artificial intelligence	Generative AI: Deep-learning based AI techniques that can be used to create or produce various types of new contents, including text, images, audio, and videos, based on patterns learnt from training data. This includes foundation models and large language models.
---	--

Table 1. Key concepts and definitions

Methods

Search Strategy and Selection Criteria

This systematic review was conducted according to PRISMA guidelines.¹⁶ We searched PubMed, Web of Science, Embase, and Scopus to identify studies published between 1st January 2012 to 18th February 2024, related to application of generative AI in reducing medication related harm. Details of the search terms are provided in [Supplement III].

Studies were included if they were published in English, described the development or application of generative AI in mitigating potential medication related harms in the care delivery process (Figure 1), and were peer-reviewed original research, review and viewpoints, structured reviews of the literature reported in accordance with PRISMA guidelines, conference abstracts, case reports. We excluded studies that utilized solely predictive modeling approaches or investigated ADEs related to dietary supplements or use of medication not prescribed for the individual.

Based on these criteria, abstracts were screened for eligibility by two independent reviewers using a standardized tool. If no exclusion criteria were apparent in the abstract, it was included for manuscript review. Full-text manuscripts were conducted by two independent reviewers. Studies that did not meet the selection criteria were excluded at this stage. In cases of discrepancy between reviewers, eligibility was determined by a third reviewer.

Data Analysis

We used a standardized form to extract pertinent information, including study characteristics, model details, application setting, outcome measures, findings, and reported challenges and limitations [Supplement IV]. We did not perform a critical appraisal of study quality as our primary objective is to characterise the scope of research in this field, identify research trends and gaps. The wide range of study designs and outcomes also precluded the application of a uniform quality assessment criterion.

Role of funding source

There was no funding source for this study.

Results

The search yielded 2203 articles from all databases, with 1734 remaining after removing duplicates. After applying inclusion and exclusion criteria, 14 articles were eligible for this review.

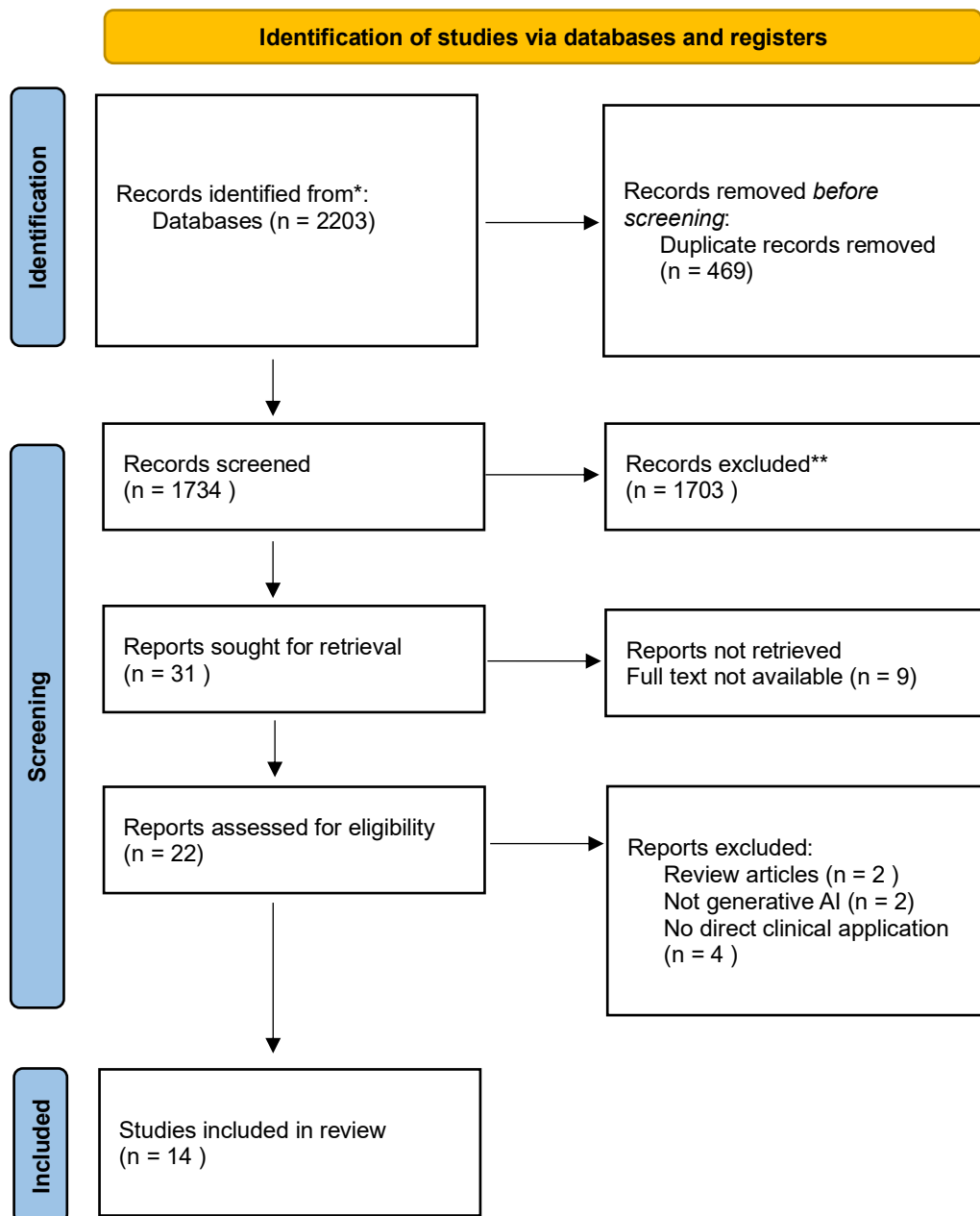


Figure 1: PRISMA flow diagram¹⁷

Study Characteristics

Studies evaluated the performance of GenAI models for various applications, as summarized in Table 2. Four studies focused on the identification, classification, or prediction of drug-drug interactions (DDI).¹⁸⁻²¹ Three studies assessed the performance and utility of GenAI as decision support tools in benzodiazepine deprescribing²², aid dosing calculation of crushed tablets²³ and provision of drug information²⁴. Majority of studies focused on the application of GenAI in adverse event monitoring from specific drug classes and enhancing pharmacovigilance processes. Study designs were predominantly observational and cross-sectional. None of the studies tested models prospectively in their respective

settings of application. The proposed applications of GenAI were broadly distributed across clinical (community, inpatient care) and public health settings.

Domain	Task (n)	Disease / Medication Class (n)	Model Details
Drug-Drug Interaction	Identification / Classification (3)	Non-specific (3)	Screen medical prescriptions for established drug-drug interactions and classify degree of severity based on existing evidence Predict unknown drug-drug interactions
	Prediction (1)	Non-specific (1)	
Adverse Drug Event	Identification / Classification (6)	Hypothyroidism (1) , drugs of abuse (1) , migraine (1) , non-specific (3)	Signal detection, severity classification in post-marketing surveillance
	Information Extraction and Summarization (1)	Non-specific (1)	Extract and summarize information for food-drug interaction from
	Decision Support (2)	Benzodiazepine (1) ; non-specific (1)	Screen for prescriptions eligible for benzodiazepine deprescribing; dose calculation for crushed tablets
Education	Text Generation (1)	Non-specific (1)	Answer enquiries collected from a pharmacovigilance centre

Table 2: Summary of studies included in review

Dataset Characteristics

Pharmacovigilance tasks often used public health databases, including FDA Adverse Event Reporting System (FAERS) Public Dashboard²⁵, Health Canada ADR reporting dashboard²⁶ , China Food and Drug Administration²⁷. Various datasets were used for named entity recognition tasks in ADE detection, such as user-generated content from web platforms (Medicitalia²⁸, WebMD²⁹) and open or restricted access datasets (CoNLL2003³⁰, BioCreative V CDR³¹, n2c2³²). Models built for DDI prediction utilized closed-source or in-house datasets such as DrugBank^{24,33}. Testing datasets are often accrued from prior published studies, such as drug-drug interaction and deprescribing case scenarios^{22,34} with one study using an in-house retrospective cohort of medication prescriptions.¹⁹

Model Types

Proprietary LLMs featured frequently in the reviewed studies, including various versions of ChatGPT and Google Bard. Studies used simple prompts or iterative prompting to generate responses on pretrained LLMs. None of the studies reported the use of additional techniques to enhance model performance, such as retrieval augmented generation or fine-tuning.

Custom-developed models adopt iterations of generative adversarial networks (GAN) and variational autoencoders (VAE) in the studies reviewed. The GTCACS³⁵ approach was a three-step approach to better identify discussion topics from social media texts. GAN achieved dimensionality reduction, keyword clustering and summarization. DeepSAVE³⁶, a deep learning framework used an enriched VAE

for dimensionality reduction through parsimonious modelling of events captured on social media platform. In one study, BART (Bidirectional and Auto-Regressive Transformers) fine-tuned with a small amount of ADR-specific named entities (few-shot learning) was adapted to allow automated identification of diverse ADEs using small volumes of annotated data.³⁷ GAN was adapted in DGANDDI²¹ into a graph attention network that encode drug attributes. DGANDDI was capable of binary and multi-class prediction tasks for drug-drug interactions using an enhanced and augmented multi-dimensional dataset generated by GAN. In a similar fashion, GAN was used to generate artificial features to augment data distribution in an imbalanced spontaneous reporting dataset.²⁷

Generative Model	Type(s) of Data Input	Training Datasets (if used)	Performance Metrics
Large Language Models (ChatGPT-3.5, ChatGPT-4, Bing AI, Google Bard)	Structured, unstructured and combination of both structured and unstructured data	Studies did not use any training datasets (zero-shot inference)	Accuracy, Sensitivity, Specificity, Negative Predictive Value, Positive Predictive Value, readability, cohen's kappa, F-1 statistic, Precision, Recall
Generative Adversarial Network	Unstructured and combination of both structured and unstructured	Spontaneous ADR reports from pharmacovigilance centres, DrugBank, WebMD	Accuracy, AUPR, AUROC, Macro-F1, Precision, Recall, ROUGE, Likert Scale for qualitative human expert evaluation

Table 3: Types of generative models and data characteristics

Model Performance

For tasks that assist clinical decision making, reference standards include information from knowledge databases e.g. Lexicomp® and expert opinion (healthcare providers or pharmacologists). Most commonly reported metrics include accuracy, sensitivity, specificity, F-1 statistic, precision, recall, AUC and AUPRC. Bespoke metrics include qualitative assessment of model responses by human experts, graded on Likert scales for quality, completeness or satisfaction. One study used ChatGPT-4 for qualitative evaluation of model performance, performed in parallel with human expert evaluation.³⁸

Drug-Drug Interaction Classification and Prediction¹⁸⁻²¹

In prediction of potential drug interaction pairs, GAN-based models achieved high accuracy rate. GANs are generative models that learn from the distribution of data or images to create large, realistic synthetic data.³⁹ DGANDDI outperformed baseline methods in both binary and multi-class DDI prediction tasks. In binary prediction, it achieved an accuracy of 96.10%, AUPR of 99.27%, and AUROC of 99.26%. In the multi-class prediction task, DGANDDI attained an accuracy of 95.89%, AUPR of 97.29%, and AUROC of 99.97%.

Proprietary LLMs were used to classify DDIs. One study compared the performance of different LLMs including Microsoft Bing AI, ChatGPT-3.5 and ChatGPT-4 and Google Bard.¹⁸ When Micromedex was used as the reference standard, accuracy of LLMs ranged between 0.469 to 0.788, with Microsoft Bing AI demonstrating the best performance. Sensitivity was comparable across all LLMs, but specificity was significantly lower for ChatGPT-3.5 and ChatGPT-4. In another study, the Google Bard was used to screen prescriptions for drug-drug interactions, demonstrated low degree of agreement with predictions from Lexicomp.¹⁹ There was a nil to slight agreement between interaction risk rating ($\kappa=0.01$), severity rating ($\kappa=0.02$), and reliability ($\kappa=-0.02$). Conversely, ChatGPT (version not reported) was found to be highly accurate in identifying drug-drug interactions in 39 out of 40 DDI pairs tested. When prompted to explain its answer, ChatGPT produced responses that was highly readable.

Decision Support²²⁻²⁴

In decision support applications, a study leveraging GPT-4 for benzodiazepine deprescribing reported high degree of overall agreement between LLM and human expert in identifying cases eligible for deprescribing.²² Agreement on four different deprescribing criteria was varied, ranging 74.7% to 91.3% (lack of indication: $\kappa = .352$, $P < .001$; prolonged use: $\kappa = .088$, $P = .280$; safety concerns: $\kappa = .123$, $P = .006$; incorrect dosage: $\kappa = .264$, $P = .001$). Qualitative analysis of GPT-4 responses found that up to 22% were ambiguous, generic and contained inconsistencies. Another study introduced a web-based calculator developed to guide dosing calculation, particularly in paediatric care where such errors are prevalent.²³ The authors used ChatGPT (version not reported) and Visual Studio to write the underlying HTML code for dose division calculations and webpage interface creation. The webpage's reliability and feasibility were then assessed using retrospective data and validated questionnaires, scoring 88.38 on the System Usability Scale. Accuracy and reproducibility of the calculator was not evaluated.

In the provision of pharmacovigilance related enquiries, ChatGPT-4 responses was compared against responses by pharmacovigilance specialists. The median score (IQR) of the ChatGPT's responses on a 10-point Likert scale was 4.8 (3–7.3), with a specific focus on drug causality scoring lower at 3.7 (3–6.3), and information on medication and proper use scoring slightly better at 5 (3.2–8.3). The authors conclude that chatbot's responses were generally not acceptable, especially in terms of precision and clinical relevance.

Pharmacovigilance^{27,35-37,40}

For signal detection and ADR classification, studies used generative AI for training data augmentation and dimension reduction. These models e.g. GTCACS outperformed non GenAI methods on internal validity measures. The DeepSAVE model, which uses a VAE approach, was tested on a dataset comprising of 104 million user search queries and 800 events. DeepSAVE outperformed existing methods (e.g. disproportionality analysis, DA atop Event Mention Classifier) with the highest F-measure across all validation datasets. A GAN-based classification model developed to automatically evaluate

risk categories of drugs during post-marketing surveillance demonstrated highest accuracy of 97.9% when compared against existing models.

In one study, authors demonstrated few-shot learning with LightNER and BART, the ADR recognition performance in low-resource datasets significantly improved. For instance, the LightNER model fine-tuned using the N2C2 dataset, achieved an F1-score of 61.42%, indicating the model's effective transfer of task knowledge from rich-resource to low-resource settings. GPT-3 was used in another study to generate a comprehensive lexicon of drug abuse synonyms from social media sources. Coupled with automated API queries and simple automated filters (e.g. google filters), the proposed method yielded precision of 0.859 and 0.770, recall of 0.431 and 0.395 for alprazolam and fentanyl respectively.

LLM was harnessed to improve efficiency of pharmacovigilance process in one study, where authors used iterative prompting of GPT-4 to review and summarize food effects on drugs from drug review documents. Final draft summaries generated by GPT-4 were rated by FDA professionals, with 85% rated as factually consistent with reference summaries. This showcases GPT-4's potential to aid in faster and more reliable drug assessment processes.

Discussion

As healthcare systems increasingly prioritize patient safety, the integration of AI has the potential to enhance the detection and prevention of ADEs, and, by extension, reduce the substantial economic cost. Our scoping review revealed three key applications of GenAI in the literature to date: identification and prediction of drug-drug interactions, provision of decision support in medication management, and automation of pharmacovigilance activities.

Effectiveness of Generative AI in Enhancing Safety

Drug-drug interactions make up nearly 3% of all hospital admissions and account for up to 5% of all inpatient medication errors.⁴¹ Harmful DDIs are often only reported from post-marketing surveillance activities, rather than at the clinical trial stage.⁴² Our review included studies that predict potential DDIs pre-clinically. Performance of models augmented by GAN outperforms those trained using traditionally augmented data using a fraction of the original training dataset.^{43,44} GAN can be a useful tool in enhancing prediction accuracy where data is limited. On the other hand, LLMs demonstrated variable performance in screening DDIs from prescriptions. Studies frequently used simple prompting strategies to elucidate response from LLMs, with no additional techniques used to provide contextual knowledge or reduce incorrect responses (or “hallucinations”). Methods such as retrieval augmented generation (RAG) or fine-tuning may allow LLMs to tailor responses to specified tasks through provision of contextual knowledge (e.g. drug-drug interaction database).^{45,46} The advantage of such techniques have been shown in other clinical tasks, including differential diagnosis⁴⁷, evidenced-based decision

support^{46,48} and patient chart review⁴⁹. These techniques however, may rely on well-curated, clinically adjudicated drug-drug interaction datasets that are not often freely available.

As decision support tools, studies adopting LLMs are mainly exploratory in nature. We found a wide range of tasks and purposes (e.g. prescription review, dosage calculation, and answering medication enquiry). These broad applications are enabled by generalist properties of large language models.⁵⁰ LLMs demonstrate capacity to perform tasks with little to no task-specific training, also known as “zero-shot” or “few-shots” learning. In the context of reducing medication harm, LLMs may simulate clinical reasoning and inferential skills across diverse medical disciplines, drug classes and user settings without the need for explicit training. For instance, an LLM trained to screen prescriptions for inappropriate benzodiazepine use may be adapted easily to screen for inappropriate drug use in elderly patients. In addition, LLMs are well poised as medical chatbots, given their text generation capabilities demonstrating high degree of fluency, empathy, and personalization, even outperforming clinicians.⁵¹ These explorations, however, highlight existing challenges to clinical adoption of LLMs. While studies to date are in research phase and no exploration in terms of auto-piloting or co-piloting as modes of clinical integration. Accuracy, reliability and consistency of LLM responses using general purpose LLMs such as ChatGPT precludes its autonomous use in clinical settings.

A promising area of LLM application in enhancing efficiency and impact is in ADE monitoring and pharmacovigilance, where GenAI tools may enhance timeliness and accuracy of ADR detection from specific medication classes. Our review has shown that GenAI enhanced accuracy of signal detection, disease and drug entity recognition over conventional natural language processing tools. LLMs are able to handle a wide breadth of data sources (i.e. electronic health records, online databases, and social media platform), facilitating the detection of rare events and offering a generalist capability that is essential for continuous learning and adaptation.⁵² Automation of specific tasks in pharmacovigilance that is traditionally resource-intensive is a potential avenue for productivity gain with the use of GenAI models.⁵³

Clinical Implications

In our review, we are unable to provide conclusive evidence that GenAI will reduce medication-related harms when applied in clinical settings. Along the continuum of medication use process, GenAI models were actually only adopted for highly selective domains and tasks. While a comprehensive review or metrics of medication-related tasks across multiple domains or tasks across different GenAI models have yet to be concluded. In other words, current generative AI in research are still applied as narrow based AI focused on specific task, yet to be explored or achieve their “generative” potential for medication safety. Non-generative AI models was evaluated in another review, where 78 articles described the application of AI in reducing ADEs.⁵⁴ A variety of AI techniques were described including neural networks and tree-based algorithms in predicting potential ADEs and enhancing early detection.

Utilizing diverse data sources like genetic information and electronic health records, these AI models aimed to inform clinical decisions on safe prescribing and medication management.

Instead of applying generative AI models in tasks that mandate deterministic outputs, we propose that LLMs can be adopted in ways to reduce cognitive workload for healthcare professionals. Healthcare professionals work with high volumes of multi-modal patient data and are required to pay attention to details, synthesize information and make clinical decisions in real time. High cognitive load pose risk for burnout and medical errors.⁵⁵ For example, LLMs can be used to analyse and reduce alert burden in electronic medical records, in medication incident analysis, and summarization in a similar fashion to discharge notes generation.^{12,56} In a study published after we completed literature search, LLMs were used in a co-pilot system to extract key named entities of online submitted prescriptions and assembly into coherent instructions.⁵⁷ This system was shown to reduce near-miss events and improved the efficiency of pharmacy operations in a large-scale online pharmacy. Finally, LLMs can be leveraged upon as a tool in patient education and engagement thereby enhancing patient access to critical medication related information.⁵⁸

A critical evaluation of studies included in our review revealed a lack of adherence to reporting guidelines for AI studies. We did not perform a quality review of the studies in view of the scoping nature of this review and diverse hypotheses of studies included. Checklists and reporting guidelines such as the MI-CLAIM for transparent model reporting⁵⁹, TRIPOD+AI checklist for comprehensive reporting of predictive models⁶⁰ and DECIDE-AI checklist for early stage clinical evaluation of AI-based decision support tools⁶¹ should be adopted in future studies. However, there is still lack of validated reporting tools for LLM-based AI model, though initial efforts have been made to create LLM-specific frameworks.⁶² Evaluation or discussions on model fairness, bias and other ethical considerations such as data privacy were also found to be lacking in the included studies.

Limitations

Our study has several limitations. The heterogeneity across studies regarding application, GenAI tools used and setting prevented a formal assessment of predictive validity for different AI models. Diversity of training and testing datasets used precludes generalizability of findings across different demographic groups. Patient outcomes were not reported in all studies, limiting any conclusions about the role of GenAI and its impact on patient outcomes. We limited our review to only peer reviewed articles. We acknowledge that the field of generative AI and LLM is rapidly evolving and a large number of studies may still be in the preprint stage or archived.

Blueprint for Future Studies

Future research should focus on developing and benchmarking generative AI models against established healthcare standards to further validate their performance and cost-effectiveness to ensure

their safe integration into clinical practice. There is a need to develop expert curated, high quality training datasets with diverse representation from different geographical, ethnic and social groups. Such datasets, when shared, can facilitate and accelerate training of GenAI models adapted for different applications in the medication use process. For example, an expert annotated dataset of incident reports can be used to fine-tune an LLM-based model to predict risk for medication incidents.⁶³ Other areas of high interest include the use of LLMs for real-time monitoring of drug safety and the exploration of GAN the synthetic generation of training data, which can help overcome the limitations posed by rare ADE occurrences.

Conclusion

GenAI and LLMs demonstrate potential in enhancing medication safety and reducing medication-related harm. Published studies reveal potential areas for successful future implementation. However, the current areas that have been addressed are targeted at only some of the key safety issues of medication safety today. Moreover, research rigor and comprehensive ethical evaluation is lacking in the studies to date. Future studies should address gaps in lack of high-quality datasets specific for medication safety tasks. Continuous update of this review is warranted given the burgeoning nature of this field.

Contributors

NL, JCLO and CM conceptualized and designed the systematic review. JCLO, CM, NN, KB, NYTT, LJ and QX performed article screening and data extraction. RR, DWB, DSWT contributed to the first draft of the report with input from NL. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of Interest

We declare no competing interests.

References

1. WHO Integrated Health Services MWH. Global burden of preventable medication-related harm in health care: a systematic review. *Geneva: World Health Organization*. 2023; Licence: CC BY-NC-SA 3.0 IGO.
2. M V, GA M, JV T, V F, GA R. The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health. *Journal of the American College of Cardiology*. 12/20/2022 2022;80(25)doi:10.1016/j.jacc.2022.11.005
3. H LL, PJ P. Twenty-First Century Global ADR Management: A Need for Clarification, Redesign, and Coordinated Action. *Therapeutic innovation & regulatory science*. 2023 Jan 2023;57(1)doi:10.1007/s43441-022-00443-8
4. DW B, DJ C, N L, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. *JAMA*. 07/05/1995 1995;274(1)
5. DW B, D L, A S, et al. The potential of artificial intelligence to improve patient safety: a scoping review. *NPJ digital medicine*. 03/19/2021 2021;4(1)doi:10.1038/s41746-021-00423-6
6. C G, M B, B M, P R, F M, P T. Accessing Artificial Intelligence for Clinical Decision-Making. *Frontiers in digital health*. 06/25/2021 2021;3doi:10.3389/fdgth.2021.645232
7. A C, O A. Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review. *JMIR medical informatics*. 07/24/2020 2020;8(7)doi:10.2196/18599

8. JC S, Q C, JC D, DM R, KB J, RA M. Evaluation of a Novel System to Enhance Clinicians' Recognition of Preadmission Adverse Drug Reactions. *Applied clinical informatics*. 2018 Apr 2018;9(2)doi:10.1055/s-0038-1646963
9. X H, D E, X L, Y Y, J Q, Z L. Evaluating the performance of ChatGPT in clinical pharmacy: A comparative study of ChatGPT and clinical pharmacists. *British journal of clinical pharmacology*. 2024 Jan 2024;90(1)doi:10.1111/bcp.15896
10. X W, X X, Z L, W T. Bidirectional Encoder Representations from Transformers-like large language models in patient safety and pharmacovigilance: A comprehensive assessment of causal inference implications. *Experimental biology and medicine (Maywood, NJ)*. 2023 Nov 2023;248(21)doi:10.1177/15353702231215895
11. D R, P P, R K, H K, C V, Y W. Effectiveness of ChatGPT in clinical pharmacy and the role of artificial intelligence in medication therapy management. *Journal of the American Pharmacists Association : JAPhA*. 2024 Mar-Apr 2024;64(2)doi:10.1016/j.japh.2023.11.023
12. Ong JCL, Jin L, Elangovan K, et al. Development and Testing of a Novel Large Language Model-Based Clinical Decision Support Systems for Medication Safety in 12 Clinical Specialties. 2024/01/29 2024;
13. Medication Without Harm - Global Patient Safety Challenge on Medication Safety. Geneva: World Health Organization LCB-N-SI. <https://iris.who.int/bitstream/handle/10665/255263/WHO-HIS-SDS-2017.6-eng.pdf?sequence=1>
14. DW B, DL B, MB VV, J S, L L. Relationship between medication errors and adverse drug events. *Journal of general internal medicine*. 1995 Apr 1995;10(4)doi:10.1007/BF02600255
15. L H, AC vG. Pharmacovigilance: methods, recent developments and future perspectives. *European journal of clinical pharmacology*. 2008 Aug 2008;64(8)doi:10.1007/s00228-008-0475-9
16. MJ P, JE M, PM B, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed)*. 03/29/2021 2021;372doi:10.1136/bmj.n71
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. 2021-03-29 2021;doi:10.1136/bmj.n71
18. FY A-A, M Z, L G, R A-F, AN B. Evaluating the Sensitivity, Specificity, and Accuracy of ChatGPT-3.5, ChatGPT-4, Bing AI, and Bard Against Conventional Drug-Drug Interactions Clinical Tools. *Drug, healthcare and patient safety*. 09/20/2023 2023;15doi:10.2147/DHPS.S425858
19. DM S, SS S, HB A, AM S, MA M. Screening the Drug-Drug Interactions Between Antimicrobials and Other Prescribed Medications Using Google Bard and Lexicomp® Online™ Database. *Cureus*. 09/09/2023 2023;15(9)doi:10.7759/cureus.44961
20. A J, N P, S S, S M, JK B, H M. The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions. *Cureus*. 03/17/2023 2023;15(3)doi:10.7759/cureus.36272
21. H Y, K L, J S. DGANDDI: Double Generative Adversarial Networks for Drug-Drug Interaction Prediction. *IEEE/ACM transactions on computational biology and bioinformatics*. 2023 May-Jun 2023;20(3)doi:10.1109/TCBB.2022.3219883
22. I B, D B, M D, et al. Clinical decision-making in benzodiazepine deprescribing by healthcare providers vs. AI-assisted approach. *British journal of clinical pharmacology*. 2024 Mar 2024;90(3)doi:10.1111/bcp.15963
23. J S, M R, TM PK, V P, SH C. Dose 4 You: Dose Division Calculator-A Tool to Reduce Calculation Errors. *Hospital pharmacy*. 2024 Apr 2024;59(2)doi:10.1177/00185787231207757
24. F M, W S, C dC, et al. Will artificial intelligence chatbots replace clinical pharmacologists? An exploratory study in clinical practice. *European journal of clinical pharmacology*. 2023 Oct 2023;79(10)doi:10.1007/s00228-023-03547-8
25. @US_FDA. FDA Adverse Event Reporting System (FAERS) Public Dashboard | FDA. 2024;
26. Canada H. Adverse Reaction Database - Canada.ca. 2009-09-24 2009;
27. J W, G F, Z L, P H, Y Z, W H. Evaluating Drug Risk Using GAN and SMOTE Based on CFDA's Spontaneous Reporting Data. *Journal of healthcare engineering*. 08/27/2021 2021;2021doi:10.1155/2021/6033860
28. Consulti medici e specialisti online: più informati, più sani! @medicitalia. <https://www.medicitalia.it>
29. WebMD. Drugs & Medications; Topamax. <https://reviewswbmdcom/drugs/drugreview-14494-topamax-oral>. 2024;
30. Erik F. Tjong Kim Sang FDM. Introduction to the CoNLL-2003 shared task | Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4. *Proceedings of the seventh conference on Natural language learning at HLT-NAACL*. 2003;4 May:142–147. doi:<https://doi.org/10.3115/1119176.1119195>

31. J L, Y S, R J J, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database : the journal of biological databases and curation*. 05/09/2016 2016;2016doi:10.1093/database/baw068
32. S H, Y W, F S, O U. The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. *Journal of the American Medical Informatics Association : JAMIA*. 10/01/2020 2020;27(10)doi:10.1093/jamia/ocaa106
33. Online D. DrugBank Online | Database for Drug and Drug Target Info. 2024;
34. AS B, MA S, A A, EA G, J B, D F. Prevalence and Determinants of Multimorbidity, Polypharmacy, and Potentially Inappropriate Medication Use in the Older Outpatients: Findings from EuroAgeism H2020 ESR7 Project in Ethiopia. *Pharmaceuticals (Basel, Switzerland)*. 08/25/2021 2021;14(9)doi:10.3390/ph14090844
35. GraniGiorgio, LenziAndrea, VelardiPaola. Supporting Personalized Health Care With Social Media Analytics: An Application to Hypothyroidism. research-article. 2021-10-15 2021;doi:4
36. F. Ahmad AA, B. Kitchens, D. Adjero and D. Zeng. Deep Learning for Adverse Event Detection From Web Search. *IEEE Transactions on Knowledge and Data Engineering*. 2022;34(6):2681-2695.
37. Chang C-H, National Sun Yat-sen University DoIM, Kaohsiung, Taiwan, Chang F-Y, Hwang S-Y, Yang CC, Drexel University CoCal, Philadelphia, USA. Prompting for Few-shot Adverse Drug Reaction Recognition from Online Reviews. *IEEE Computer Society*; 2023:168-175.
38. Y S, P R, J W, et al. Leveraging GPT-4 for food effect summarization to enhance product-specific guidance development via iterative prompting. *Journal of biomedical informatics*. 2023 Dec 2023;148doi:10.1016/j.jbi.2023.104533
39. AA, AA. Generative adversarial networks and synthetic patient data: current challenges and future perspectives. *Future healthcare journal*. 2022 Jul 2022;9(2)doi:10.7861/fhj.2022-0013
40. KA C, RB A. Using GPT-3 to Build a Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance. *Biomolecules*. 02/18/2023 2023;13(2)doi:10.3390/biom13020387
41. LL L, DW B, DJ C, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA*. 07/05/1995 1995;274(1)
42. Lu Y, Shen D, Pietsch M, et al. A novel algorithm for analyzing drug-drug interactions from MEDLINE literature. OriginalPaper. *Scientific Reports*. 2015-11-27 2015;5(1):1-10. doi:doi:10.1038/srep17357
43. AA V, FF G, YJ K, et al. Deploying deep learning models on unseen medical imaging using adversarial domain adaptation. *PloS one*. 10/14/2022 2022;17(10)doi:10.1371/journal.pone.0273262
44. AJ SK, RS C, JG C, et al. Evaluation of Generative Adversarial Networks for High-Resolution Synthetic Image Generation of Circumpapillary Optical Coherence Tomography Images for Glaucoma. *JAMA ophthalmology*. 10/01/2022 2022;140(10)doi:10.1001/jamaophthalmol.2022.3375
45. C Z, R S, A C, et al. Almanac - Retrieval-Augmented Language Models for Clinical Medicine. *NEJM AI*. 2024 Feb 2024;1(2)doi:10.1056/aioa2300068
46. Ke Y, Jin L, Elangovan K, et al. Development and Testing of Retrieval Augmented Generation in Large Language Models -- A Case Study Report. 2024/01/29 2024;
47. S R, A R, J N, et al. A retrieval-augmented chatbot based on GPT-4 provides appropriate differential diagnosis in gastrointestinal radiology: a proof of concept study. *European radiology experimental*. 05/17/2024 2024;8(1)doi:10.1186/s41747-024-00457-x
48. S K, M G, MA A, LS C, DL S. Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. *NPJ digital medicine*. 04/23/2024 2024;7(1)doi:10.1038/s41746-024-01091-y
49. Vaid Aea. Using fine-tuned large language models to parse clinical notes in musculoskeletal pain disorders. *The Lancet Digital Health*. 2024;5(12):e855 - e858.
50. Tu T, Azizi S, Driess D, et al. Towards Generalist Biomedical AI. research-article. 2024-02-22 2024;doi:10.1056/Aloa2300138
51. IA B, YV Z, D G, et al. Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions. *JAMA network open*. 08/01/2023 2023;6(8)doi:10.1001/jamanetworkopen.2023.30320
52. M M, O B, ZSH A, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023 Apr 2023;616(7956)doi:10.1038/s41586-023-05881-4
53. M S, J P, P Y, et al. The Use of Artificial Intelligence in Pharmacovigilance: A Systematic Review of the Literature. *Pharmaceutical medicine*. 2022 Oct 2022;36(5)doi:10.1007/s40290-022-00441-z

54. A S, W S, MG A, et al. Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review. *The Lancet Digital health*. 2022 Feb 2022;4(2)doi:10.1016/S2589-7500(21)00229-6
55. DE E, SN G, S N, et al. Evaluating and reducing cognitive load should be a priority for machine learning in healthcare. *Nature medicine*. 2022 Jul 2022;28(7)doi:10.1038/s41591-022-01833-z
56. D VV, C VU, L B, et al. Adapted large language models can outperform medical experts in clinical text summarization. *Nature medicine*. 2024 Apr 2024;30(4)doi:10.1038/s41591-024-02855-5
57. C P, J L, R V, V G, E W, M B. Large language models for preventing medication direction errors in online pharmacies. *Nature medicine*. 04/25/2024 2024;doi:10.1038/s41591-024-02933-8
58. Yang R, Tan, TF , Lu, W , Thirunavukarasu, AJ , Ting, DSW , Liu, N . Large language models in health care: development, applications, and challenges . . *Health Care Sci* 2023;2:255 – 263.
59. B N, G Q, BK B-J, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nature medicine*. 2020 Sep 2020;26(9)doi:10.1038/s41591-020-1041-y
60. GS C, KGM M, P D, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ (Clinical research ed)*. 04/16/2024 2024;385doi:10.1136/bmj-2023-078378
61. B V, M N, B C, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature medicine*. 2022 May 2022;28(5)doi:10.1038/s41591-022-01772-9
62. Ning Y, Teixayavong S, Shang Y, et al. Generative Artificial Intelligence in Healthcare: Ethical Considerations and Assessment Checklist. 2023/11/02 2023;
63. ZSY W, N W, J L, S U. A large dataset of annotated incident reports on medication errors. *Scientific data*. 02/29/2024 2024;11(1)doi:10.1038/s41597-024-03036-2
64. Rights of the Individual. 2023;
65. Organization TWH. Regulation and Prequalification: What is Pharmacovigilance? Accessed 20 Feb 2024,

Supplement I - Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	1
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	3,4
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	3,4, supplement
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	NA
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	5
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	4
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	Supplement
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	5
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	Supplement
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	NA
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	5

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	6
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	6
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	6 – 9
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	6 – 9
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	6 – 9
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	9 – 12
Limitations	20	Discuss the limitations of the scoping review process.	12
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	12
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	12

Supplement II - Population, Concept and Contexts (PCC) for the scoping review

Population	Healthcare professionals who prescribe, dispense or administer medications to patients; patient/medication safety groups within health systems responsible for monitoring, design, development and implementation of patient / medication safety initiatives; groups or health agencies who carry out pharmacovigilance activities*; patients who are the recipients of medication prescriptions.
Concept	Generative AI: Deep-learning based AI techniques that can be used to create or produce various types of new contents, including text, images, audio, and videos, based on patterns learnt from training data. ⁶⁴
	Medication Safety: The freedom from harm (both preventable and unpreventable) that arise from medication use.
Context	Refers to a broad range of healthcare settings including provision of direct clinical care, healthcare administration and operations and public or commercial groups with medication safety as an agenda. This includes settings where medication is prescribed, dispensed, administered to patients and monitored for safety indicators.

*Pharmacovigilance is defined as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine/vaccine related problem⁶⁵.

Supplement III - Search Strategy

Advanced Search Query Syntax				
Component	PubMed	Web of Science	Embase	Scopus
Generative AI and Large Language Models	("Generative AI"[tiab]) OR ("Generative artificial intelligence"[tiab]) OR ("Large language"[tiab]) OR ("GPT-*"[tiab]) OR ("Variational Autoencoder"[tiab]) OR ("generative adversarial"[tiab]) OR ("GAN"[tiab]) OR ("GANs"[tiab]) OR ("bard"[tiab]) OR ("stable diffusion"[tiab]) OR ("midjourney"[tiab]) OR ("dall-e"[tiab]) OR ("Variational Autoencoders"[tiab]) OR ("Variational Autoencoder"[tiab]) OR ("Variational Autoencoders"[tiab]) OR ("BART"[tiab]) OR ("GLaM"[tiab]) OR ("LaMDA"[tiab]) OR ("PaLM"[tiab]) OR ("Gemini"[tiab]) OR ("Mistral"[tiab]) OR ("LLaMA"[tiab]) OR ("Med-PaLM"[tiab]) OR ("Med-LLM"[tiab])	TS=(("Generative AI") OR ("Generative artificial intelligence") OR ("Large language Model") OR ("GPT-*") OR ("Variational Autoencoder") OR ("generative adversarial") OR ("GAN") OR ("bing") OR ("bard") OR ("stable diffusion") OR ("midjourney") OR ("dall-e") OR ("Variational Autoencoders") OR ("Variational Autoencoder") OR ("XLNet") OR ("BART") OR ("GLaM") OR ("LaMDA") OR ("PaLM") OR ("YaLM") OR ("LLaMA") OR ("Gemini") OR ("Mistral") OR ("Mixtral") OR ("LLaMA") OR ("Med-PaLM") OR ("Med-LLM"))	('generative ai':ab OR 'generative artificial intelligence':ab OR 'large language':ab OR 'gpt':ab OR 'generative adversarial':ab OR 'gan':ab OR 'gans':ab OR 'bard':ab OR 'stable diffusion':ab OR 'midjourney':ab OR 'dall-e':ab OR 'variational autoencoder':ab OR 'variational auto-encoder':ab OR 'language model':ab OR 'machine learning':ab OR 'artificial intelligence':ab OR 'xlnet':ab OR 'bart':ab OR 'glam':ab OR 'lamda':ab OR 'palm':ab OR 'yalm':ab OR 'llama':ab OR "gemini" OR "mistral" OR "mixtral" OR "LLaMa" OR "med-PaLM" OR "Med-LLM")	TITLE-ABS-KEY("Generative AI" OR "Generative artificial intelligence" OR "Large language" OR "GPT-*" OR "Variational Autoencoder" OR "generative adversarial" OR "GAN" OR "GANs" OR "bard" OR "stable diffusion" OR "midjourney" OR "dall-e" OR "Variational Autoencoders" OR "Variational Autoencoder" OR "Variational Auto-encoders" OR "Variational Auto encoder" OR "Variational Auto encoders" OR "XLNet" OR "BART" OR "GLaM" OR "LaMDA" OR "PaLM" OR "Gemini" OR "Mistral" OR "Mixtral" OR "LLaMA" OR "Med-PaLM" OR "Med-LLM")
	AND	AND	AND	AND
Adverse Drug Reactions (ADRs)	"Pharmacovigilance"[Mesh] OR "Incident reporting [tiab]" OR "Adverse Drug Reaction Reporting Systems"[Mesh] OR "pharmacovigilance"[tiab] OR "Drug toxicit*[tiab] OR "Side effect"[tiab] OR "Adverse drug event*[tiab] OR "Adverse drug reaction*[tiab] OR "Drug-Related Side Effects and Adverse Reactions"[Mesh] OR "Overdose*[tiab] OR "Drug overdose"[MeSH] OR "Contraindication*[tiab] OR "Poison*[tiab] OR "Drug Eruptions"[Mesh] OR "Drug Hypersensitivity"[Mesh] OR "Drug Hypersensitivity Syndrome"[Mesh] OR "drug reaction*[tiab] OR "adverse drug*[tiab] OR "drug hypersensitivit*[tiab]	TITLE-ABS-KEY "Pharmacovigilance" OR "Drug Incident reporting" OR "Adverse Drug Reaction Reporting" OR "Drug Side effect" OR "Adverse drug event" OR "Contraindication" OR "Drug Eruption" OR "Drug Hypersensitivity"	TS=(("Pharmacovigilance") OR ("Drug Incident reporting") OR ("Adverse Drug Reaction Reporting") OR ("Drug Side effect") OR ("Adverse drug event") OR ("Adverse drug reaction") OR ("Contraindication") OR ("Drug Eruption") OR ("Drug Hypersensitivity"))	pharmacovigilance':ab OR 'adverse drug reaction':ab OR 'hypersensitivity':ab

<p>Drug Interactions</p>	<p>"Drug Interactions"[Mesh] OR drug interaction*[tiab] OR pharmacokinetic interactions[tiab] OR pharmacodynamic interactions[tiab])</p>	<p>TITLE-ABS-KEY "Drug Interaction" OR "Pharmacokinetic interaction" OR "pharmacodynamic interactions" OR "Drug-Drug Interaction"</p>	<p>TS=(("Drug Interaction") OR ("Pharmacokinetic interaction") OR ("pharmacodynamic interactions") OR ("Drug-Drug Interaction"))</p>	<p>("drug interaction":ab)</p>
<p>Medication Use Process</p>	<p>"Medication Errors"[Mesh] OR medication error*[tiab] OR "Prescription Errors"[Mesh] OR prescribing errors[tiab] OR prescription errors[tiab] OR "Contraindications, Drug"[Mesh] OR "omission error[tiab]" OR "Medication Reconciliation"[Mesh] OR medication reconciliation[tiab] OR drug reconciliation[tiab] OR "Polypharmacy" [Mesh] OR "deprescribing" [Mesh] OR "deprescribe*[tiab] OR dispensing error[tiab] OR administration error[tiab] OR transcribing error[tiab] OR transcription error[tiab]</p>	<p>TITLE-ABS-KEY "Medication Error" OR "Prescription Errors" OR "Contraindications" OR "Medication Reconciliation" OR "Polypharmacy" OR "Deprescribing" OR "Dispensing error" OR "Administration error" OR "transcription error"</p>	<p>TS=(("Medication Error") OR ("Prescription Error") OR ("Prescribing Error") OR ("Contraindications") OR ("Medication Reconciliation") OR ("Polypharmacy") OR ("Deprescribing") OR ("Dispensing error") OR ("Administration error") OR ("transcription error"))</p>	<p>("Medication error":ab) OR ("deprescription":ab) OR ("transcription error":ab)</p>

<p>Patient Compliance / Adherence</p>	<p>“Medication adherence”[Mesh] OR adherence[tiab] OR “Patient compliance”[Mesh] OR Medication persistence[tiab] OR drug persistence[tiab] OR drug adherence[tiab] OR drug compliance[tiab] OR “treatment refusal”[Mesh] OR remote monitoring[tiab] OR remote sensing[tiab]</p>	<p>TITLE-ABS-KEY “Medication adherence” OR “Medication compliance” OR Medication persistence”</p>	<p>TS=(("Medication adherence") OR ("Medication compliance") OR ("Medication persistence"))</p>	<p>("medication compliance":ab)</p>
<p>Patient / HCP Education</p>	<p>“Patient education as topic”[Mesh] OR medication counselling[tiab] OR drug counselling[tiab] OR “health education”[Mesh] OR patient interview[tiab]</p>	<p>TITLE-ABS-KEY “Patient education” OR “Medication counselling” OR “Drug counselling” OR “health education” OR “patient interview”</p>	<p>TS=(("Patient education") OR ("Medication counselling") OR ("Drug counselling") OR ("health education") OR ("patient interview"))</p>	<p>((“patient education”) AND (“drug therapy”) OR “prescription”))</p>

Supplement IV - Data abstraction variables

Article Details	Title Author List Year of Publication Journal
Study Characteristics	Study design (Model development, observational, review) Dataset characteristics (source e.g. electronic health records, online database, sample size, image or structured data or unstructured data) Patient group (specific disease state, medication class, public health) Comparator (control group) if any
AI model details	Type of model used (including tools used, software or specific algorithms) Model application (classification, prediction, decision support)
Application area and setting	Medication use process setting (ambulatory, inpatient, clinical trial) Safety domain (ADR, drug interaction etc) Mode of model implementation (trailed or proposed) (autonomous, hybrid or fully human in the loop)
Outcome measures	Key performance metrics for AI model (Validation and external testing if any) Key outcome measure study (patient related outcome) if any
Findings	Reported model performance Reported patient outcome measure if any (reported values and confidence intervals or SD if any)
Discussion	Challenges and limitations
Others	Any other relevant remarks