International Archives of Medical Sciences & Public Health

E-ISSN : 2798-5172; P-ISSN : 2798-5276X
Volume 7, Number 1, 2026
DOI: 10.53806/iamsph.v7i1.1447
Page: 28–40

EVOLUTION OF ARTIFICIAL INTELLIGENCE IN PATIENT SAFETY ACROSS SOUTHEAST ASIA: A BIBLIOMETRIC ANALYSIS

Muhammad Almanfaluthi[1], **Arlina Dewi**[2]

[1,2]Master of Hospital Administration, Universitas Muhammadiyah Yogyakarta, Indonesia

Article Info

Article history:
Received : February 10, 2026
Revised : March 17, 2026
Accepted : May 15, 2026

Keywords:
Artificial intelligence;
Bibliometrics;
Machine learning;
Patient safety;
Southeast Asia.

ABSTRACT

This study evaluated the evolution of artificial intelligence (AI) and machine learning in patient safety across Southeast Asia (SEA) to identify regional research dynamics and emerging frontiers. A quantitative bibliometric analysis was conducted using 262 primary documents retrieved from the Scopus database (English-language; published up to December 31, 2025). Document selection was guided by an adaptation of the PRISMA-ScR framework. Key indicators included the Mann-Kendall trend test, Pettitt's change-point test, and thematic co-occurrence network mapping. Statistical analysis revealed a significant structural shift in 2015 ($p < 0.01$), marking exponential publication growth and extensive international collaboration (62.2% of documents featuring multi-country co-authorship). Results highlighted a divergence in regional strategies; high-income nations produced high-impact clinical algorithms, while emerging economies prioritized capacity building for resource-constrained systems. Thematic mapping demonstrated a major transitional shift in the literature from traditional, image-based neural networks for diagnostic accuracy toward natural language processing and large language models aimed at addressing clinical documentation and prescribing errors. These publication dynamics map the current academic focus, highlighting the need for future implementation-driven clinical validation and regionally adapted regulatory frameworks.

Corresponding Author:
Muhammad Almanfaluthi,
Master of Hospital Administration,
Universitas Muhammadiyah Yogjakarta
Email: m.luthfi.a@ump.ac.id

1. INTRODUCTION

Ensuring patient safety remains a critical global healthcare challenge. Preventable adverse events (ranging from medication and diagnostic errors to healthcare-associated infections) impose a profound human and economic burden [1]. Concurrently, artificial intelligence (AI) and machine learning (ML) are emerging as transformative tools capable of addressing these clinical vulnerabilities by mitigating human cognitive biases and standardizing care protocols [2].

Historically, foundational AI research in patient safety capitalized on image-based architectures, such as convolutional neural networks, to resolve diagnostic inaccuracies [3]. Recently, this scientific frontier has expanded to include natural language processing (NLP) and large language models (LLMs) to address systemic administrative vulnerabilities, such as clinical documentation burdens and prescribing ambiguities [4], [5]. However, while the technical capabilities of AI are advancing rapidly on a global scale, the trajectory of this research exhibits distinct geographic disparities.

The vast majority of existing literature and deployed AI models originate from high-income nations with mature digital infrastructures. Consequently, global bibliometric analyses often mask the realities of emerging regions, inherently overrepresenting the research priorities of Western healthcare systems [6]. This creates a critical knowledge gap regarding Southeast Asia (SEA), a highly heterogeneous healthcare ecosystem. The region encompasses globally influential clinical AI hubs, such as Singapore, alongside emerging economies like Malaysia and Indonesia, which face distinct infrastructural constraints, workforce shortages, and localized implementation challenges [7].

While AI adoption is accelerating across SEA, a comprehensive understanding of how regional socioeconomic realities shape machine learning research priorities within patient safety remains absent. This study provides the first structured bibliometric evaluation of AI applications specifically within the SEA region. By isolating this geographic network, this research demonstrates how globally developed AI models are adapted to resource-diverse healthcare systems. To address this gap, this study formulated two specific research questions: (1) How has the scientific production of artificial intelligence and machine learning in patient safety evolved over time in Southeast Asia, and what are the emerging technological frontiers? (2) What is the network structure of institutional collaborations and thematic concepts within this domain, and how do these structures reflect the diverging research strategies of regional healthcare systems? By answering these questions, this research aims to guide future localized clinical implementation, policy development, and targeted investments in the region.

2. RESEARCH METHOD

This study employed a quantitative bibliometric analysis. This design was selected because it provides a robust, highly structured, and objective framework for evaluating the evolution of scientific production, identifying key collaborative networks, and mapping the thematic landscape at the intersection of the machine learning and patient safety domains in Southeast Asia [8]. The data were sourced exclusively from Scopus. Scopus was selected over alternatives such as Web of Science or PubMed because of its superior interdisciplinary indexing of both health sciences and computer engineering literature, which is a critical necessity for AI research. Furthermore, Scopus has historically provided broader coverage of journals and conference proceedings affiliated with emerging economies in Southeast Asia than other major international indices. The primary variables observed included bibliographic metadata: publication year, author names, institutional affiliations (filtered to focus on Southeast Asian countries), citation counts, document types, and text data derived from the article titles, abstracts, and keywords.

Data extraction was performed on 3 March 2026. A comprehensive search strategy was executed within the Scopus TITLE-ABS-KEY field to capture literature at the intersection of AI/ML methodologies and specific patient safety domains. To ensure complete reproducibility, the exact Boolean search string utilized was:

> TITLE-ABS-KEY ((”artificial intelligence” OR ”machine learning” OR ”deep learning” OR ”neural network*” OR ”natural language processing” OR ”predictive model*”) AND (”patient safety” OR ”medical error*” OR ”clinical error*” OR ”medication error*” OR ”adverse drug event*” OR ”prescribing error*” OR ”healthcare associated infection*” OR ”hospital acquired infection*” OR ”nosocomial infection*” OR ”surgical error*” OR ”unsafe surgery” OR ”wrong site surgery” OR ”surgical complication*” OR ”unsafe injection*” OR ”needle stick injury” OR ”injection safety” OR ”diagnostic error*” OR ”misdiagnosis” OR ”delayed diagnosis” OR ”failure to diagnose” OR ”venous thromboembolism” OR ”deep vein thrombosis” OR ”pulmonary embolism” OR ”radiation error*” OR ”radiation safety” OR ”radiation overexposure” OR ”unsafe transfusion” OR ”transfusion error*” OR ”adverse transfusion reaction*”))

The query was designed to capture two primary concepts: artificial intelligence/machine learning methodologies and specific patient safety domains (including medication errors, healthcare-associated infections, unsafe surgical strategies, injection safety, diagnostic errors, venous thromboembolism, radiation errors, and unsafe transfusions).

The initial search yielded 13,866 documents in total. To ensure relevance and consistency, the coverage was restricted to literature published in English until 2025. Subsequently, an exclusion process was applied during the screening phase to remove non-primary research and specific document types (Table 1). Following the initial extraction, the dataset underwent rigorous screening and cleaning protocols. First, automated deduplication was performed in the R environment to remove overlapping records. To ensure the clinical relevance of the retrieved literature, two reviewers (MA and AD) independently screened the titles and abstracts of the remaining documents to verify that AI or ML methodologies were explicitly applied to patient safety outcomes. Any discrepancies between the reviewers were resolved through a consensus discussion. Finally, a meticulous data normalization step was conducted to resolve inconsistencies in the bibliographic metadata. Author names and institutional affiliations were standardized utilizing fuzzy string-matching algorithms specifically Levenshtein distance calculations via the *stringdist* and *stringr* packages in R followed by manual validation by the research team to merge structural variants (e.g., consolidating "Natl Univ Singapore" and "National University of Singapore"). A total of 7,797 documents, including books, chapters, and articles, were systematically excluded (Figure 1). While conference proceedings represent a significant volume of early-stage computer science and algorithmic research, they were intentionally excluded. This decision was made to strictly prioritize mature, peer-reviewed clinical applications and translational health informatics research, which serve as stronger indicators of actual patient safety integration. Furthermore, regarding network analysis parameters, self-citations were retained rather than excluded to accurately map the development of continuous regional collaborative networks and sustained institutional output. Raw citation counts were utilized without fractional normalization to reflect absolute academic impact. Ultimately, a regional filter was applied to isolate studies authored by researchers affiliated with Southeast Asia (SEA), yielding a final inclusion dataset of $N = 262$ primary documents for the review.

Table 1: Inclusion and exclusion criteria for document selection

Criterion	Inclusion	Exclusion
Topic	Application of AI/ML within specified patient safety domains.	General AI research without clinical safety applications; clinical safety research without AI methodologies.
Document Type	Primary research articles and comprehensive reviews (Articles, Reviews).	Books, book chapters, conference proceedings, editorials, letters, notes, short surveys, errata, and retracted documents.
Language	English.	Non-English publications.
Time Frame	All years up to and including December 31, 2025.	Publications indexed in 2026 or later.
Regional Affiliation	At least one author was affiliated with an institution in a Southeast Asian (SEA) country (Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam).	No author affiliations from Southeast Asian countries or outside the designated SEA region.

The analytical framework utilised both descriptive statistics and non-parametric statistical modelling, which were selected specifically to evaluate the macro-level translational shifts in clinical AI applications. Key Performance Indicators (KPIs) were calculated to summarise the overall scientific impact, including aggregate document volume, unique author contributions, international collaboration rates, and Compound Annual Growth Rate (CAGR). To evaluate institutional impact alongside raw productivity, the h-index was calculated at the institutional level and integrated into a strategic performance matrix. To ensure statistical rigor in our time-series publication trend analysis, two specific tests were employed: the Mann-Kendall test was utilised to assess the statistical significance and direction of monotonic trends over time, while Pettitt's test was applied to detect critical change points, allowing us to pinpoint the specific years in which structural shifts in publication volume occurred [9]. To explore the thematic landscape, keyword co-occurrence networks were generated to identify how different AI methodologies bridge various clinical and patient safety domains. While traditional bibliometric measures, such as co-citation analysis, bibliographic coupling, and citation burst detection, provide valuable insights into theoretical lineages and author-level networks, this study intentionally omitted these to maintain a strict focus on semantic thematic evolution and temporal growth dynamics [10], [11]. By prioritising co-occurrence and trend detection, the analysis was optimised to map the practical translation of AI methodologies into clinical safety paradigms across the diverse healthcare ecosystem of Southeast Asia.

To ensure a transparent, rigorous, and reproducible methodology, the document screening and selection process was conducted following an adaptation of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines, tailored specifically for bibliometric analysis (Figure 1). The document selection process, detailed in the PRISMA flow diagram (Figure 1), proceeded through three distinct phases: identification, screening, and inclusion. The initial global search identified 13,866 documents. During the screening phase, 7,797 documents were systematically excluded based on document type (e.g., removing books, chapters, conference proceedings, and reviews), leaving an intermediate dataset of 6,069 primary research articles. Finally, during the inclusion phase, a strict regional filter was applied to isolate the geographic network. A total of 5,807 documents were excluded because they lacked author affiliations from Southeast Asian nations, yielding the final, highly targeted inclusion dataset of N = 262 primary documents for review. All data wrangling, statistical analysis, and visualisation procedures were performed using the R programming language in the R-Studio environment [12]. While dedicated bibliometric software platforms such as VOSviewer, CiteSpace, and the Bibliometrix package offer highly accessible graphical interfaces for standard network mapping, a custom script-based approach utilising the R programming language (via the R-Studio environment) was deliberately selected for this study. This approach was chosen because it provides the algorithmic flexibility required to seamlessly integrate advanced non-parametric statistical modeling specifically Mann-Kendall and Pettitt's tests which are not native to standard bibliometric GUIs. Furthermore, R allows highly reproducible data wrangling and custom visual design. Specifically, the dplyr, tidyr, and stringr packages were used for text processing and data manipulation, whereas the trend package was used for statistical computation. High-resolution infographic-style dashboards and area charts were generated using the ggplot2 package in conjunction with the patchwork for grid-based. Because this study relied exclusively on secondary, publicly available bibliographic metadata and did not involve human subjects, animal subjects, or sensitive clinical data, institutional ethical approval was not required layout assembly.

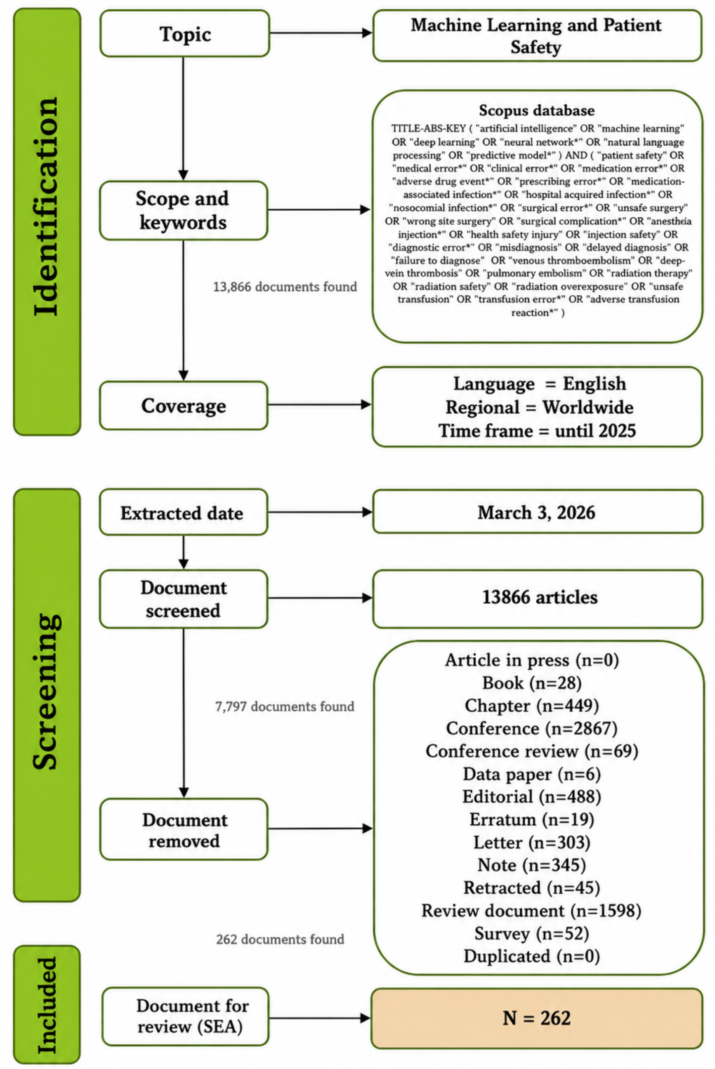

Figure 1: PRISMA Flow Diagram for Bibliometric Document Selection

3. RESULT AND ANALYSIS

The bibliometric analysis of the selected Southeast Asian (SEA) literature identified 262 documents authored by 3,728 unique contributors, representing a highly collaborative research environment with an international collaboration rate of 62.2% (Figure 2). Trend analysis demonstrated a statistically significant monotonic increase in scientific production over the observed period (Mann Kendall Tau = 0.66, $p < 0.001$) (Figure 3). Furthermore, Pettitt's test detected a critical structural shift in 2015 ($p < 0.01$), marking the transition from a dormant phase to a period of exponential growth, resulting in a robust compound annual growth rate (CAGR).

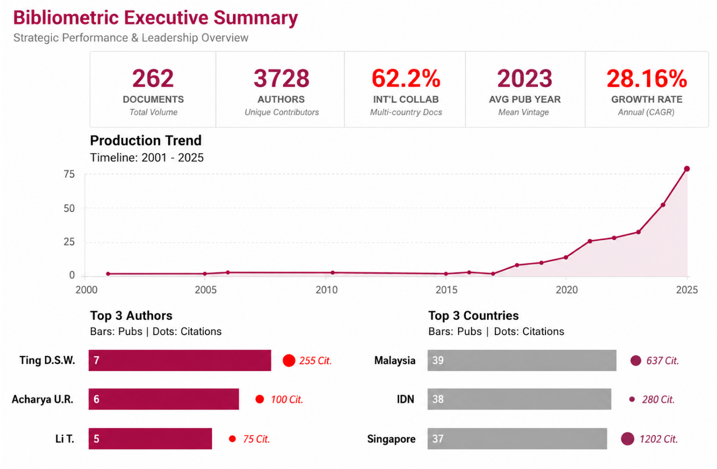

Figure 2: Bibliometric Executive Summary. This dashboard presents the key performance indicators (KPIs) of the dataset

Metrics include total document volume, unique author count, percentage of documents featuring international collaboration (multi-country affiliations), mean publication year, and Compound Annual Growth Rate (CAGR) from 2001 to 2025. The lower panels display the top three authors and countries, with the bar length representing the publication volume and the adjacent dots detailing the total accumulated citations.

The exponential growth in AI-related healthcare research in Southeast Asia (SEA) is consistent with a broader, globally documented paradigm shift toward integrating deep learning frameworks into clinical settings during the late 2010s. This transformation was not merely incremental but represented a fundamental restructuring of how healthcare data are conceptualised, processed, and applied to patient safety outcomes. As noted across multiple domains of clinical informatics, the mid-2010s marked a decisive inflection point at which machine learning (ML) and deep learning (DL) transitioned from experimental laboratory tools to clinically deployable systems [13]. The rapid proliferation of accessible neural network architectures (including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models) has enabled researchers to analyse complex, high-dimensional datasets drawn from electronic health records (EHRs), medical imaging, and real-time physiological monitoring with unprecedented accuracy and scalability [13], [14].

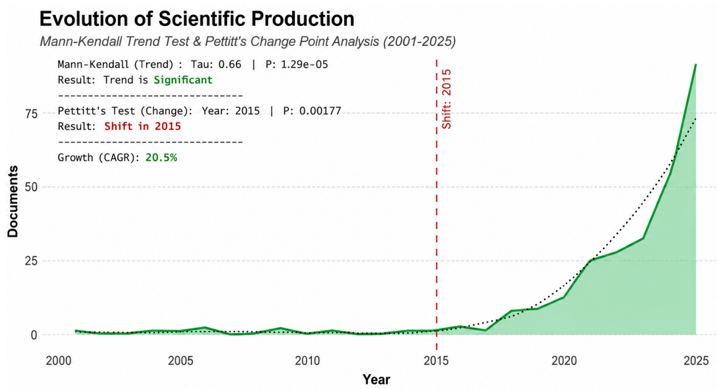

Figure 3: Time series evolution of scientific production

The X-axis represents the publication year (2001–2025), and the Y-axis represents the absolute number of documents published. The solid green line and shaded area denote the annual publication volume, whereas the dotted black line illustrates the smoothed polynomial trend. The vertical red dashed line indicates the structural change point detected by Pettitt's test (Year: 2015, $p < 0.01$), marking a statistically significant shift to exponential growth.

This shift was further catalysed by the maturation of deep learning techniques and the growing availability of large-scale medical data sets. For instance, between 2018 and 2019, the U.S. The Food and Drug Administration (FDA) has approved 23 AI-based medical devices, providing institutional legitimacy and regulatory scaffolding that encourages broader clinical adoption globally [15]. The bibliometric literature on ML in chronic disease management confirms a continuous and steep rise in publications from approximately 2010 onwards, with deep learning models, such as convolutional neural networks and ensemble methods, dominating the recent advancements [16]. Similarly, a comprehensive 33-year bibliometric analysis of AI in type 2 diabetes prediction documented exponential publication growth since 2010, with emerging players such as Singapore and India increasingly contributing to the global research landscape [16]. These trends collectively underscore that the late 2010s represented not only a quantitative increase in research output but also a qualitative transformation in the sophistication and clinical applicability of AI-driven models.

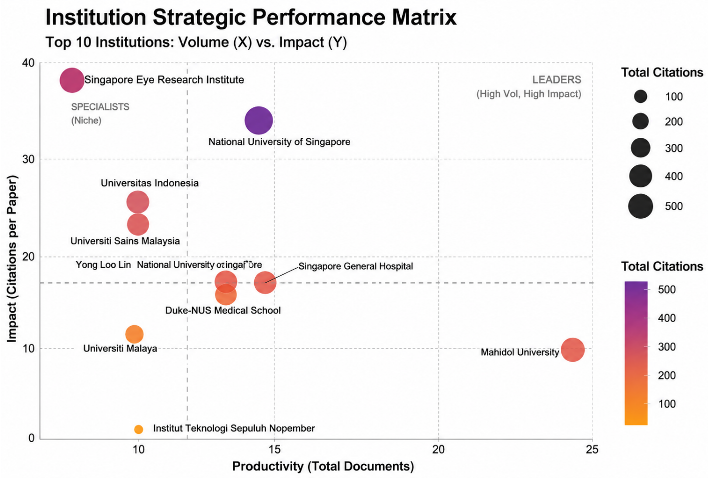

Figure 4: Institutional Strategic Performance Matrix

A scatter plot evaluating the top 10 Southeast Asian institutions. The X-axis measures productivity (Total Documents), while the Y-axis measures impact (Average Citations per paper). Dashed reference lines represent the dataset medians, dividing the matrix into performance quadrants (for example, 'Leaders' in the top-right vs. 'Niche' in the top-left). The bubble size and the corresponding colour gradient (orange to purple) visually encode the total cumulative citations achieved by each institution.

Geographic aggregation highlighted Malaysia, Indonesia, and Singapore as the top three contributing nations in the SEA region. The Institutional Strategic Performance Matrix (Figure 4) plots institutional productivity against citation impact (citations per paper), revealing distinct operational strategies. Mahidol University emerged as the highest-volume producer, driving the regional productivity. Conversely, the Singapore Eye Research Institute and the National University of Singapore occupied the high-impact "niche" quadrant; although their absolute publication volume was lower, they achieved the highest citation rates per paper in the dataset.

The divergence between publication volume and citation impact across Southeast Asian (SEA) nations reflects a nuanced landscape of complementary regional strengths, in which different countries occupy distinct but mutually reinforcing roles within the broader AI healthcare research ecosystem. Singapore's prominence as a hub for methodologically sophisticated, high-impact clinical AI research is well documented across multiple domains, particularly in ophthalmology, where the city-state has consistently ranked among the most productive and influential global contributors [17], [18]. Bibliometric analyses of AI in ophthalmology confirm that Singapore exhibits strong centrality and influence in international research networks, with a notably high h-index relative to its publication volume, indicating that Singaporean institutions produce fewer but disproportionately highly cited publications [18]. This pattern is consistent with Singapore's strategic positioning as a high-income country with mature digital infrastructure, well-funded research institutions, and a regulatory environment conducive to the clinical translation of advanced AI technologies [19].

Singapore's strength in automated ophthalmic screening exemplifies this dynamic. The Singapore National Eye Centre has been identified as one of the most productive institutions globally in AI ophthalmology research [19]. Such programs represent not merely proof-of-concept research but fully operationalized AI systems embedded within national healthcare delivery frameworks, a level of clinical maturity that distinguishes Singapore from its regional peers. The economic evaluation literature further confirms that AI-driven DR screening, for which Singapore has served as a model, represents one of the most rigorously evaluated and cost-effective applications of clinical AI in ophthalmology. Singapore-affiliated researchers, including those at Duke-NUS Medical School, have also contributed foundational methodological work in deep learning for retinal disease detection, which has garnered substantial international citations and shaped the global research agenda [20].

In contrast, emerging research hubs in Malaysia and Indonesia demonstrate a distinct but equally important trajectory, characterised by rapid capacity building, high publication productivity, and a focus on localised implementation challenges. Malaysia has been identified as the leading AI adopter in Southeast Asia (SEA) after Singapore, aligning with its National Transformation 2050 and Industry 4.0 policy frameworks. Malaysian institutions have demonstrated growing engagement with AI-driven cancer screening, including breast and colorectal cancer applications, with studies spanning prospective clinical evaluations and silent trial phases [21]. This paradox of high adoption rates alongside infrastructural limitations suggests that Indonesian researchers and institutions are actively adapting AI tools to resource-constrained hospital systems, prioritising practical and locally relevant applications over methodologically novel contributions [22], [23].

Qualitative evidence from SEA healthcare professionals supports this interpretation. Participants from lower- and middle-income SEA countries, including Indonesia, consistently identified infrastructure barriers, market access concerns, and limited investment as critical impediments to AI adoption, while expressing strong interest in AI's potential to address population health management, service accessibility in remote areas, and operational efficiency [7]. Indonesian participants specifically highlighted the potential of AI-assisted health facilities to serve frontier and remote regions where access to physicians and medical resources is severely constrained, reflecting a research and implementation agenda oriented toward public health equity rather than methodological innovation. Similarly, Malaysia's AI governance policies emphasise human capital development over ethics-focused frameworks, suggesting a pragmatic orientation toward building the workforce and institutional capacity necessary to deploy AI in real-world clinical settings [23]. This divergence in research orientation (between Singapore's technologically mature, citation-intensive clinical AI specialisation and Malaysia and Indonesia's productivity-driven, implementation-focused capacity building) is consistent with broader patterns observed in global AI healthcare research, wherein high-income countries tend to produce foundational, highly cited methodological contributions, while middle-income countries generate larger volumes of applied, context-specific research [7]. Rather than representing a deficit, the high productivity of Malaysian and Indonesian institutions reflects a necessary and complementary function within the regional research ecosystem: generating the locally validated evidence base required to adapt globally developed AI models to the specific disease burdens, healthcare system configurations, and resource constraints of Southeast Asian (SEA) populations [21]. Together, these complementary strengths position SEA as a region capable of advancing the methodological frontier of clinical AI and demonstrating its real-world applicability across diverse, socioeconomic contexts.

Semantic analysis of the dataset (Figure 6) yielded 111 unique thematic terms with a Shannon entropy of 4.19, indicating a highly diverse and information-rich research landscape. "Machine learning," " deep learning", and " artificial intelligence " artificial intelligence functioned as the core technological pillars (For a complete numerical breakdown of keyword frequencies, top contributing authors, and leading journals, please refer to Supplementary Appendix A). The co-occurrence network (Figure 5) mapped this landscape into five distinct clusters (modularity = 0.416, density = 0.134), linking core AI methodologies directly to specific clinical domains, such as COVID-19, stroke, brain tumours, and overarching patient safety.

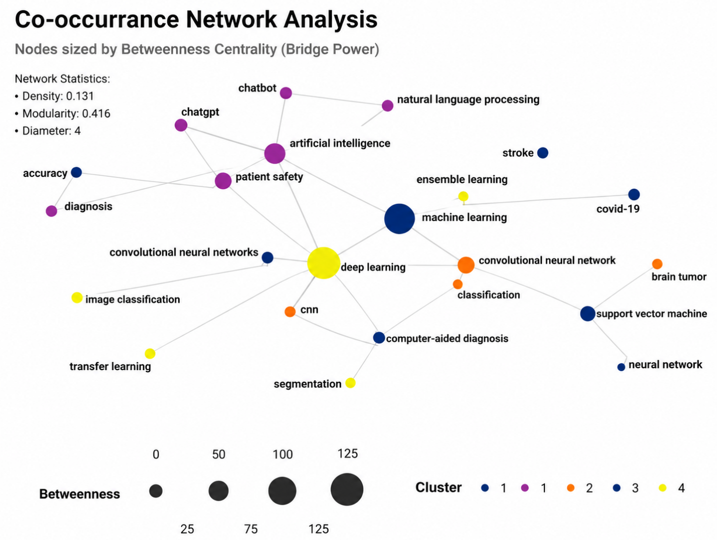

Figure 5: Thematic Co-occurrence Network Analysis

A sociogram mapping the interconnections between the most frequent author keywords. Each node represents a distinct theme. The node size is strictly proportional to the Betweenness Centrality' (Bridge Power), highlighting terms that connect disparate research areas. The edges (grey lines) represent co-occurrences in the same document, with thicker lines indicating higher frequency. Node colours represent algorithmic community clustering (Clusters 1-4), grouping methodologies with their most commonly associated clinical application domains.

The moderate modularity and relatively low density of co-keyword networks in AI healthcare research reflect a field that remains in active maturation, wherein distinct thematic clusters exist but are bridged by shared methodological and clinical concerns. Bibliometric analyses employing modularity-based clustering algorithms consistently reveal that AI healthcare research networks exhibit intermediate modularity scores, indicating the presence of identifiable thematic communities that maintain substantial cross-cluster connectivity [24]. This structural characteristic is diagnostic of an interdisciplinary field in transition: sufficiently differentiated to produce specialised research clusters, yet insufficiently consolidated to exhibit the high modularity associated with fully mature, siloed disciplines [25]. In the context of SEA AI patient safety research, this network topology suggests that artificial intelligence methodologies function as "bridge technologies" computational tools whose applicability transcends individual clinical domains and connects otherwise disparate patient safety challenges across specialties, pathologies, and healthcare system levels.

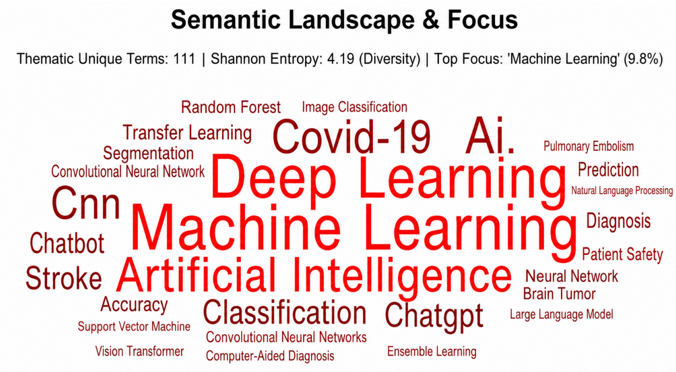

Figure 6: Semantic Landscape and Word Cloud

A visual representation of the 111 unique thematic terms in the dataset. The size of each term is strictly proportional to its absolute frequency of occurrence within the document titles and abstracts. The Shannon Entropy score (4.19) indicates the high informational diversity of the terminology used across the dataset.

The conceptualisation of AI as a bridge technology connecting disparate clinical domains is well supported by the literature on AI integration in healthcare. AI and machine learning frameworks have been documented as cross-cutting methodological infrastructures capable of simultaneously addressing diagnostic accuracy, treatment optimisation, risk stratification, and patient monitoring across multiple clinical specialties [25]. The interdisciplinary nature of AI deployment is particularly evident in radiology, where deep learning systems have been shown to bridge the gap between image acquisition, computer-aided diagnosis, predictive analytics, and workflow optimisation, functions that span multiple traditional clinical disciplines [3]. Similarly, in the domain of diagnostic error reduction, AI tools have been proposed as integrative mechanisms capable of synthesising information from electronic health records, vital signs, laboratory results, and imaging data to improve real-time diagnostic accuracy in obstetrics, infectious diseases, sepsis, and oncology [25]. This cross-domain applicability is precisely what generates the moderate modularity observed in bibliometric network analyses: AI methodologies appear as high-betweenness-centrality nodes connecting otherwise loosely coupled clinical research clusters [24].

The prominence of specific pathologies alongside systemic crises, such as COVID-19, in SEA AI research networks further illustrates this bridging function. The COVID-19 pandemic served as a powerful accelerant for AI adoption across multiple clinical domains, catalysing the development of AI tools for diagnostic imaging, predictive analytics, triage, and epidemiological surveillance within compressed timeframes [26]. This crisis-driven expansion of AI applications reinforced the cross-domain connectivity of AI methodologies within the research literature, as tools developed for pandemic response were rapidly adapted and validated for application to endemic disease burdens, including tuberculosis, dengue, and diabetic complications, which are conditions of particular relevance to SEA populations [25].

Temporal tracking of topic lifespans (Figure 7) illustrates the rapid pace of the domain's evolution. Foundational concepts, such as "neural network", exhibited the most sustained presence over the timeline and served as the continuous backbone of the research. In contrast, "ChatGPT" and "Large Language Models" emerged abruptly as the newest trends, clustering with very recent median publication years alongside advanced techniques like "transfer learning" and "segmentation."

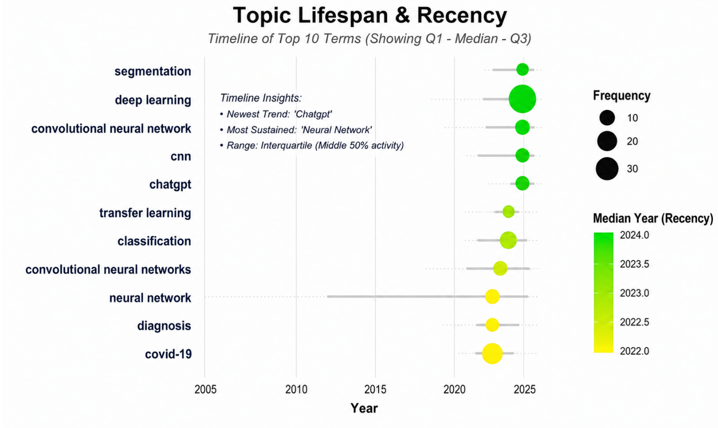

Figure 7: Topic Lifespan and Recency Timeline

Chronological mapping of the top 10 most prominent research themes. The X-axis denotes the publication year. For each term, the horizontal grey dashed line represents the interquartile range (Q1 to Q3), indicating the middle 50% span of publication activity. Coloured bubbles denote the median publication year for each term. The bubble size indicates the total occurrence frequency, while the colour gradient (yellow to bright green) explicitly encodes recency, allowing for a visual distinction between sustained foundational terms (e.g. 'neural network') and emerging frontier topics (e.g. 'chatgpt').

The progression from traditional quantitative image-based deep learning architectures to natural language processing (NLP) tools and large language models (LLMs) represents one of the most consequential methodological transitions in clinical AI research. Earlier AI applications in healthcare focused predominantly on structured data (laboratory results, billing codes, and medical images used in radiology and pathology) for risk prediction, diagnostic support, and operational optimisation [27]. However, LLMs represent a paradigm shift because of their unique capacity to analyse unstructured text data, a significant advancement given the abundance of clinically relevant information found in unstructured text formats such as clinical notes, discharge summaries, and prescribing records [27]. This transition is not merely technical but reflects a fundamental reorientation of patient safety research priorities from preventing diagnostic errors through image classification to addressing the broader systemic vulnerabilities embedded in clinical documentation, communication, and decision-making workflows.

The distinction between these two generations of clinical AI is architecturally important. CNN-based systems excel at pattern recognition within bounded, well-defined image classification tasks, whereas large language models (LLMs) (grounded in transformer neural network architectures) exhibit remarkable proficiency in processing, understanding, and generating human language across open-ended, contextually complex clinical scenarios [27]. This architectural difference translates directly into a difference in the patient safety domains that each technology can address: CNNs are optimally suited to prevent misdiagnoses arising from image interpretation errors, whereas LLMs are positioned to address a far broader category of errors arising from documentation failures, prescribing ambiguities, communication breakdowns, and cognitive overload among clinicians [4].

The literature increasingly posits that LLMs can address these vulnerabilities by automating documentation tasks. Current regional research reflects optimism that reducing the cognitive burden on clinicians can yield structured records to support safer prescribing. However, it is critical to note that while bibliometric trends show a massive surge in academic interest in LLM-driven adverse drug event (ADE) extraction, these models largely remain in the experimental or silent-trial phases. The actual safety impact, clinical effectiveness, and reduction of preventable harm cannot be established through publication volume alone and must await validation through prospective clinical outcome studies. A significant application of LLMs in improving clinical workflows is the substantial reduction in the documentation burden that has long affected doctors and nurses, a problem that persisted even after the transition from paper to electronic health records [28]. Major EHR vendors have begun embedding generative AI features in production environments, with early site reports indicating reduced documentation time; however, large-scale independent evaluations of clinical text generation quality, safety, and downstream outcomes are limited [29]. According to expert consensus from a Delphi study on LLMs in healthcare, documentation tasks (including automatic clinical coding, summarisation of clinical documents, automatic structuring of clinical narratives, and medical charting assistance) are among the most agreed-upon use cases for LLM-based systems [4].

The susceptibility of clinical documentation to human cognitive biases and administrative fatigue is well-established. Clinicians operating under conditions of high cognitive load and time pressure are disproportionately prone to documentation errors, omissions, and ambiguities that can propagate downstream into prescribing decisions and care coordination failures [29]. Large language models (LLMs) address this vulnerability by automating or semi-automating documentation tasks, thereby reducing the cognitive burden on clinicians and creating structured, auditable records that support safer prescribing and handover processes [30]. In the domain of adverse drug events (ADEs) (a major category of preventable patient harm), LLMs, including transformer-based architectures such as BERT and GPT series, have demonstrated promising performance in automating ADE extraction from clinical data; studies have suggested that LLM-driven methods can improve the early identification and classification of ADEs and assist with medication management decisions [31]. Integrating LLM output into clinical workflows may help clinicians prevent harm by alerting them to emerging safety signals or prescribing risks, representing a direct application of generative AI to one of the most consequential patient safety challenges in clinical practice.

The rapid adoption of generative AI tools in Southeast Asia (SEA) is consistent with the broader global trajectory but carries region-specific implications shaped by the area's heterogeneous healthcare infrastructure and workforce constraints. Generative AI tools, such as ChatGPT-4o and Gemini, are rapidly influencing public health research and communication, with their capacity to assist with drafting, summarising, and translating content offering significant potential, particularly in multilingual and resource-limited settings [6], characteristics that directly describe the linguistic diversity and resource constraints of SEA healthcare systems. Notably, generative AI developments in Asia are accelerating, with SEA-based startups such as SayHeart (Malaysia and Singapore) launching algorithms that translate medical jargon, health reports, and complex imaging into accessible content[32]. This demonstrates that the region is not merely a passive recipient of globally developed tools but is an active contributor to generative AI innovation in healthcare.

The potential of LLMs to optimise digital healthcare workflows is particularly relevant for SEA, where workforce shortages, administrative burdens, and geographic barriers to specialist access create strong incentives for automation [7], [28]. AI participants in SEA healthcare systems identified reducing human error, optimising triaging, and improving human resource management as key priorities for AI adoption, with Singapore demonstrating that AI can support healthcare operations by streamlining medical records, optimising resources, and reducing operational costs [7]. The integration of LLMs into cancer screening programs in ASEAN (including the application of GPT-4 to provide contextual knowledge on appropriate colonoscopy intervals) further illustrates the region's active engagement with generative AI tools for clinical decision support beyond the imaging domain [21]. The current trajectory of the literature suggests that future patient safety research in SEA is increasingly exploring generative AI as a potential core infrastructure for clinical documentation, prescribing safety, and workflow automation. If successfully validated, this could extend the region's AI patient safety agenda well beyond diagnostic imaging; however, translating this academic focus into clinical reality fundamentally requires robust, prospective implementation evidence.

Based on the bibliometric evidence and the identified thematic shifts within the Southeast Asian context, several distinct implications emerge: (1) Implications for Clinical Practice: The high publication productivity regarding AI in SEA does not equate to widespread clinical deployment. Healthcare administrators and practitioners must approach the integration of generative AI with cautious optimism. Before local hospital systems adopt these tools for triaging or documentation, practitioners must demand rigorous, localised clinical validation to ensure that these models do not introduce new, unintended medical errors (e.g. AI hallucinations in clinical notes). (2) Implications for Policy and Regulation: The rapid emergence of LLMs in regional literature underscores a critical regulatory gap. Policymakers must urgently develop unified ASEAN-level regulatory frameworks for AI in healthcare. These policies must prioritise algorithmic transparency, data privacy compliance, and human-in-the-loop (HITL) system designs to ensure that AI remains a fail-safe rather than a liability. (3) Implications for Future Research: This bibliometric mapping reveals that the theoretical groundwork for AI in patient safety has been thoroughly established in SEA. Consequently, future research funding and academic focus must transition from retrospective model-training studies to implementation science. Priority should be given to prospective randomised clinical trials that directly measure the real-world impact of AI tools on patient mortality, morbidity, and workflow efficiency in resource-constrained regional hospitals.

While this study provides a comprehensive overview of the artificial intelligence and patient safety research landscape in Southeast Asia, several inherent bibliometric limitations must be rigorously acknowledged. First, the reliance exclusively on the Scopus database introduces indexing disparities; while extensive, it may underrepresent regional literature published in non-indexed, local SEA journals or national repositories (e.g., the ASEAN Citation Index). This is compounded by an English-language dominance bias, which inherently omits native-language studies that might offer critical insights into localized implementation challenges. Second, the bibliometric methodology is susceptible to publication bias, wherein successful AI models or positive conceptual frameworks are disproportionately published compared to failed implementations. Third, bibliometric methodologies are inherently susceptible to search-string sensitivity bias. While our Boolean query was constructed to be as comprehensive as possible across both computational methodologies and patient safety domains, the rapid and continuous evolution of artificial intelligence terminology means that some relevant studies utilizing novel, highly niche, or newly coined algorithmic nomenclature may have been inadvertently excluded from the dataset. Furthermore, metrics such as citation counts are vulnerable to citation inflation and represent academic network influence rather than clinical utility. Crucially, as a measure of academic output, this study cannot confirm the actual clinical effectiveness, economic impact, or real-world safety outcomes of the proposed AI systems. Consequently, to translate these academic trends into clinical reality, future research must move beyond bibliometric mapping and utilize systematic clinical reviews and prospective trials to rigorously evaluate the true efficacy of these tools within diverse regional hospital settings

4. CONCLUSION

This study maps the exponential growth of AI patient safety research in Southeast Asia. A strategic regional divergence is evident: while high-income hubs focus on mature diagnostic algorithms, research in emerging economies increasingly explores large language models to conceptualize solutions for administrative vulnerabilities, such as clinical documentation errors. To safely manage this academic and technological shift, ASEAN-level regulatory frameworks emphasizing algorithmic transparency are essential. Ultimately, transitioning these tools from theoretical models to clinical reality presents profound implementation challenges. Rather than relying solely on retrospective modeling, future research must cautiously bridge this gap through carefully designed prospective studies to assess the true feasibility, safety, and impact of AI on patient outcomes within resource-diverse workflows.

REFERENCES

[1] D. Ambarwati and A. Dewi, "The implementation of patient safety goals for patients' satisfaction in the hemodialysis unit," *Jurnal Administrasi Kesehatan Indonesia*, vol. 12, no. 1, pp. 25–36, 2024. https://doi.org/10.20473/jaki.v12i1.2024.25-36.

[2] D. Purbohadi, S. Afriani, N. Rachmanio, and A. Dewi, "Developing medical virtual teaching assistant based on speech recognition technology," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 04, pp. 107–120, 2021. https://doi.org/10.3991/ijoe.v17i04.21343.

[3] T. Iffath, P. Athkuri, M. Nadeem, and P. Goyal, "Redefining radiology - artificial intelligence integration in medical imaging," *La Prensa Medica Argentina*, vol. 110, no. 1, 2024. https://doi.org/10.47275/0032-745x-410.

[4] K. Denecke, R. May, LlmhealthGroup, and O. Rivera Romero, "Potential of large language models in health care: Delphi study," *J Med Internet Res*, vol. 26, p. e52399, 2024. https://doi.org/10.2196/52399.

[5] A. J. e. a. Jinia, "Artificial intelligence-based incident analysis and learning system to enhance patient safety and improve treatment quality," *NPJ Digit Med*, vol. 9, no. 1, p. 254, 2026. https://doi.org/10.1038/s41746-026-02390-2.

[6] K. K. Yoga Ratnam, "Generative artificial intelligence in public health research and scientific communication: A narrative review of real applications and future directions," *Digit Health*, vol. 11, p. 20552076251362070, 2025. https://doi.org/10.1177/20552076251362070.

[7] M. F. Wibowo, A. Pyle, E. Lim, J. W. Ohde, N. Liu, and J. Karlstrom, "Insights into the current and future state of ai adoption within health systems in southeast asia: Cross-sectional qualitative study," *J Med Internet Res*, vol. 27, p. e71591, 2025. https://doi.org/10.2196/71591.

[8] M. e. a. Almanfaluthi, *Menyusun Penelitian Kesehatan dan Kedokteran Seri: Konsep Dasar Penelitian dan Statistik*. Sleman, Indonesia: Deepublish, 2025.

[9] C. e. a. Liu, "Research hotspot and trend analysis in the diagnosis of inflammatory bowel disease: A machine learning bibliometric analysis from 2012 to 2021," *Front Immunol*, vol. 13, p. 972079, 2022. https://doi.org/10.3389/fimmu.2022.972079.

[10] R. Alkhammash, "Bibliometric, network, and thematic mapping analyses of metaphor and discourse in covid-19 publications from 2020 to 2022," *Front Psychol*, vol. 13, p. 1062943, 2022. https://doi.org/10.3389/fpsyg.2022.1062943.

[11] A. Zaidi, S.-S. M. Ajibade, M. Musa, and F. V. Bekun, "New insights into the research landscape on the application of artificial intelligence in sustainable smart cities: A bibliometric mapping and network analysis approach," *International Journal of Energy Economics and Policy*, vol. 13, no. 4, pp. 287–299, 2023.

[12] RCoreTeam, "R: A language and environment for statistical computing," 2022. https://www.R-project.org/.

[13] Y. A. Fahim, I. W. Hasani, S. Kabba, and W. M. Ragab, "Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives," *Eur J Med Res*, vol. 30, no. 1, p. 848, 2025. https://doi.org/10.1186/s40001-025-03196-w.

[14] G. e. a. Tocu, "The potential of artificial intelligence in the diagnosis and prognosis of sepsis: A narrative review," *Diagnostics (Basel)*, vol. 15, no. 17, p. 2169, 2025. https://doi.org/10.3390/diagnostics15172169.

[15] S. e. a. Shen, "Mapping the landscape of machine learning in chronic disease management: A comprehensive bibliometric study," *Digit Health*, vol. 11, p. 20552076251361614, 2025. https://doi.org/10.1177/20552076251361614.

[16] M. Kiran, Y. Xie, N. Anjum, G. Ball, B. Pierscionek, and D. Russell, "Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis," *Front Digit Health*, vol. 7, p. 1557467, 2025. https://doi.org/10.3389/fdgth.2025.1557467.

[17] C. Boudry, H. Al Hajj, L. Arnould, and F. Mouriaux, "Analysis of international publication trends in artificial intelligence in ophthalmology," *Graefes Arch Clin Exp Ophthalmol*, vol. 260, no. 5, pp. 1779–1788, 2022. https://doi.org/10.1007/s00417-021-05511-7.

[18] J. Zhao, Y. Lu, S. Zhu, K. Li, Q. Jiang, and W. Yang, "Systematic bibliometric and visualized analysis of research hotspots and trends on the application of artificial intelligence in ophthalmic disease diagnosis," *Front Pharmacol*, vol. 13, p. 930520, 2022. https://doi.org/10.3389/fphar.2022.930520.

[19] S. Chen, L. Huang, X. Li, Q. Feng, H. Lu, and J. Mu, "Hotspots and trends of artificial intelligence in the field of cataracts: A bibliometric analysis," *Int Ophthalmol*, vol. 44, no. 1, p. 258, 2024. https://doi.org/10.1007/s10792-024-03207-5.

[20] M. A. e. a. Urina-Triana, "Machine learning and ai approaches for analyzing diabetic and hypertensive retinopathy in ocular images: A literature review," *IEEE Access*, vol. 12, pp. 54590–54607, 2024. https://doi.org/10.1109/ACCESS.2024.3378277.

[21] H. M. Tun, H. A. Rahman, L. Naing, and O. A. Malik, "Artificial intelligence utilization in cancer screening program across asean: A scoping review," *BMC Cancer*, vol. 25, no. 1, p. 703, 2025. https://doi.org/10.1186/s12885-025-14026-x.

[22] S. Fatimah, "Smart healthcare revolution: Artificial intelligence-based futuristic healthcare innovation as a form of optimization health facilities in indonesia," vol. 29, pp. 65–72, 2025. https://doi.org/10.4028/p-izr1ni.

[23] A. J. Keith, "Governance of artificial intelligence in southeast asia," *Global Policy*, vol. 15, no. 5, pp. 937–954, 2024. https://doi.org/10.1111/1758-5899.13458.

[24] X. Peng, R. Zhu, K. Zhang, J. Ji, C. Lv, and F. Feng, "Global collaboration and innovation in malignant cerebral edema research: A bibliometric perspective," *Front Neurol*, vol. 16, p. 1624101, 2025. https://doi.org/10.3389/fneur.2025.1624101.

[25] Y. Li, X. Yi, J. Fu, Y. Yang, C. Duan, and J. Wang, "Reducing misdiagnosis in ai-driven medical diagnostics: A multidimensional framework for technical, ethical, and policy solutions," *Front Med (Lausanne)*, vol. 12, p. 1594450, 2025. https://doi.org/10.3389/fmed.2025.1594450.

[26] D. Giansanti, A. Lastrucci, A. Pirrera, S. Villani, E. Carico, and E. Giarnieri, "Ai in cervical cancer cytology diagnostics: A narrative review of cutting-edge studies," *Bioengineering (Basel)*, vol. 12, no. 7, p. 769, 2025. https://doi.org/10.3390/bioengineering12070769.

[27] K. H. Jung, "Large language models in medicine: Clinical applications, technical challenges, and ethical considerations," *Healthc Inform Res*, vol. 31, no. 2, pp. 114–124, 2025. https://doi.org/10.4258/hir.2025.31.2.114.

[28] P. Yu, H. Xu, X. Hu, and C. Deng, "Leveraging generative ai and large language models: A comprehensive roadmap for healthcare integration," *Healthcare (Basel)*, vol. 11, no. 20, p. 2776, 2023. https://doi.org/10.3390/healthcare11202776.

[29] Y. e. a. Artsi, "Challenges of implementing llms in clinical practice: Perspectives," *J Clin Med*, vol. 14, no. 17, p. 6169, 2025. https://doi.org/10.3390/jcm14176169.

[30] C. Lin and C. F. Kuo, "Roles and potential of large language models in healthcare: A comprehensive review," *Biomed J*, vol. 48, no. 5, p. 100868, 2025. https://doi.org/10.1016/j.bj.2025.100868.

[31] M. M. Zitu, D. Owen, A. Manne, P. Wei, and L. Li, "Large language models for adverse drug events: A clinical perspective," *J Clin Med*, vol. 14, no. 15, p. 5490, 2025. https://doi.org/10.3390/jcm14155490.

[32] P. Zhang and M. N. Kamel Boulos, "Generative ai in medicine and healthcare: Promises, opportunities and challenges," *Future Internet*, vol. 15, no. 9, p. 286, 2023. https://doi.org/10.3390/fi15090286.