Review Paper

# The Impact of AI-based Nursing Documentation on Time Management and Patient Safety: A Systematic Review

**Ghazaleh Afkhami Teimouri[1], Arash Amadeh Taheri[1], Yasamin Moeinipour[2], Malihe Adineh Fathabadi[1], Fatemeh Bagheri[1*]**

*1. Department of Nursing and Midwifery, MMS.C. Islamic Azad University, Mashhad, Iran.*
*2. Department of Cardiothoracic Surgery, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.*

Use your device to scan and read the article online

**Citation** Afkhami Teimouri Gh, Amadeh Taheri A, Moeinipou Y, Adineh Fathabadi M, Bagheri F. The Impact of AI-based Nursing Documentation on Time Management and Patient Safety: A Systematic Review. Journal of Research & Health. 2025; 15(Special Issue: Artificial Intelligence):641-660. http://dx.doi.org/10.32598/JRH.15.specialissue.2821.1

http://dx.doi.org/10.32598/JRH.15.specialissue.2821.1

## ABSTRACT

**Background:** The administrative burden of nursing documentation is a primary contributor to clinician burnout. Artificial intelligence (AI), particularly emerging generative and ambient technologies, offers a potential solution, yet the evidence regarding its dual impact on efficiency and safety remains fragmented. This systematic review aimed to synthesize the evidence on the effects of AI-driven interventions in nursing documentation on time management and patient safety outcomes.

**Methods:** A systematic review was conducted following PRISMA 2020 guidelines and was registered with PROSPERO (CRD420251089257). A comprehensive search was performed in PubMed, Scopus, Web of Science, and Embase with no date or language restrictions. Primary research studies evaluating any AI intervention in nursing documentation for its effect on time or safety were included. Due to significant heterogeneity, a narrative synthesis was performed.

**Results:** From an initial 2,052 records, 18 studies met the inclusion criteria. The included studies were methodologically diverse, comprising randomized trials, quasi-experimental, and qualitative designs, with most assessed at a moderate risk of bias. The findings indicated that generative and ambient AI tools can significantly reduce documentation time and improve efficiency. The impact on patient safety varied. Some AI tools directly prevented adverse events (e.g. medication errors) or identified safety risks more effectively, while others improved safety indirectly by enhancing documentation quality. Critically, several studies highlighted the emergence of new risks, such as AI-generated inaccuracies ("hallucinations") and a lack of clinical nuance, underscoring the necessity of human oversight.

**Conclusion:** AI-driven documentation systems significantly enhance clinical efficiency by reducing documentation time and cognitive workload, thereby improving workflow and allowing greater focus on patient care. However, their reliability for autonomous use remains limited, underscoring the need for human oversight to maintain clinical accuracy and safety. Persistent challenges, including data heterogeneity, interoperability gaps, and ethical concerns, must be addressed through standardized frameworks, advanced natural language processing (NLP) development, and transparent validation. Future large-scale, multi-center studies should evaluate the sustained effects of AI-assisted documentation on efficiency, clinician well-being, and patient outcomes to enable safe, trustworthy, and equitable integration into clinical practice.

**Keywords:** Artificial intelligence (AI), Nursing documentation, Documentation burden, Patient safety, Systematic review

**Article info:**
**Received:** 03 Oct 2025
**Accepted:** 08 Nov 2025
**Publish:** 31 Dec 2025

*** Corresponding Author:**
*Fatemeh Bagheri, Assistant Professor.*
***Address:*** *Department of nursing and midwifery, MMS.C. Islamic Azad University, Mashhad, Iran.*
***Phone:*** *+98 (915) 1056519*
***E-mail:*** *bagherifatemeh59@yahoo.com*

## Introduction

The documentation of uniform nursing practices is characterized by prompt, accurate, and consistent documentation of evaluation, diagnosis, treatment strategies, interventions, and outcomes throughout the nursing process. Empirical evidence suggests that, in addition to continuity of care, there is a direct correlation between the caliber of such documentation and the overall quality and safety of care [1]. From a quality and safety point of view, the introduction of organized electronic frameworks increases the accuracy of the content and usability of the data, which is correlated with a reduction in document inaccuracies and a positive progress in patient safety indicators [2, 3]. The nursing dossier serves as a fundamental framework for patient-oriented, safe, and effective care, and dedicating resources to standardization, improving quality, and further utilizing its data significantly enhances both clinical and administrative outcomes [4].

Despite its foundational role, clinical documentation in practice has become a "hidden burden" for nurses. With the implementation of electronic health records (EHRs), the volume and complexity of documentation tasks have significantly increased, consuming a substantial portion of a nurse's shift. Time allocation studies reveal that nurses spend between 25% and 35.3% of their work time on documentation [5, 6], time that could otherwise be dedicated to direct patient care, education, and support. In the health informatics literature, this phenomenon has been framed as the "documentation burden." This added load is also linked to adverse psychological and professional outcomes. Research has demonstrated a correlation between high documentation volume and inefficient or unusable EHRs with factors, such as emotional exhaustion and depersonalization, which contribute to burnout [7, 8]. Furthermore, documentation processes are inherently susceptible to human errors from incomplete or incorrect medication records to omissions and delays in documenting vital signs that can lead to flawed clinical decision-making and adverse events [3].

In response to the dual challenges of the "documentation time burden" and associated "safety risks," artificial intelligence (AI) has emerged as a transformative solution. Three key branches, automatic speech recognition (ASR), natural language processing (NLP), and machine learning (ML), can optimize the documentation loop from "clinical dialogue" to "structured note," restoring time for direct patient interaction through intelligent automation. In the NLP domain, systematic and integrative reviews have shown that extracting critical components from unstructured nursing notes is reliable for decision support, trend identification, and predicting high-risk events, like patient falls [9, 10]. Moreover, combining emergency department nursing texts with structured data in ML models has facilitated the early detection of critical conditions, such as sepsis, and enhanced the efficacy of early warning systems [11]. At the same time, field experiences with ASR serve as a reminder that automated documentation without review may be prone to transcriptional and conceptual errors. Therefore, a "human-in-the-loop" approach and quality assurance procedures are vital for ensuring the safety and accuracy of records [12, 13]. Collectively, this body of evidence suggests that the targeted application of AI can significantly enhance documentation efficiency while also contributing to patient safety by reducing burnout and improving data quality, provided it is designed responsibly and its clinical outcomes are continuously evaluated [14-18].

In recent years, a growing number of studies have investigated the application of these technologies in clinical settings. Despite promising results, the existing evidence in this field is highly scattered, heterogeneous, and in the early stages of maturity. Many of these studies are small-scale, single-center, and employ quasi-experimental designs, which carry a high potential for bias and limit the generalizability of their findings. There is also considerable heterogeneity in the types of technology used, patient populations, clinical settings, and outcome measures evaluated. This diversity makes it difficult to synthesize the evidence and draw definitive conclusions about the overall effectiveness of these interventions.

Consequently, this systematic review aimed to address this evidence gap and the heterogeneity of existing results by synthesizing the effect of AI-driven interventions in nursing documentation on two key outcomes: time management and patient safety. This study will provide practical, evidence-based recommendations for administrators, policymakers, clinical nurses, and technology developers, an action that is of strategic importance given the increasing investments by healthcare systems in digital health and artificial intelligence.

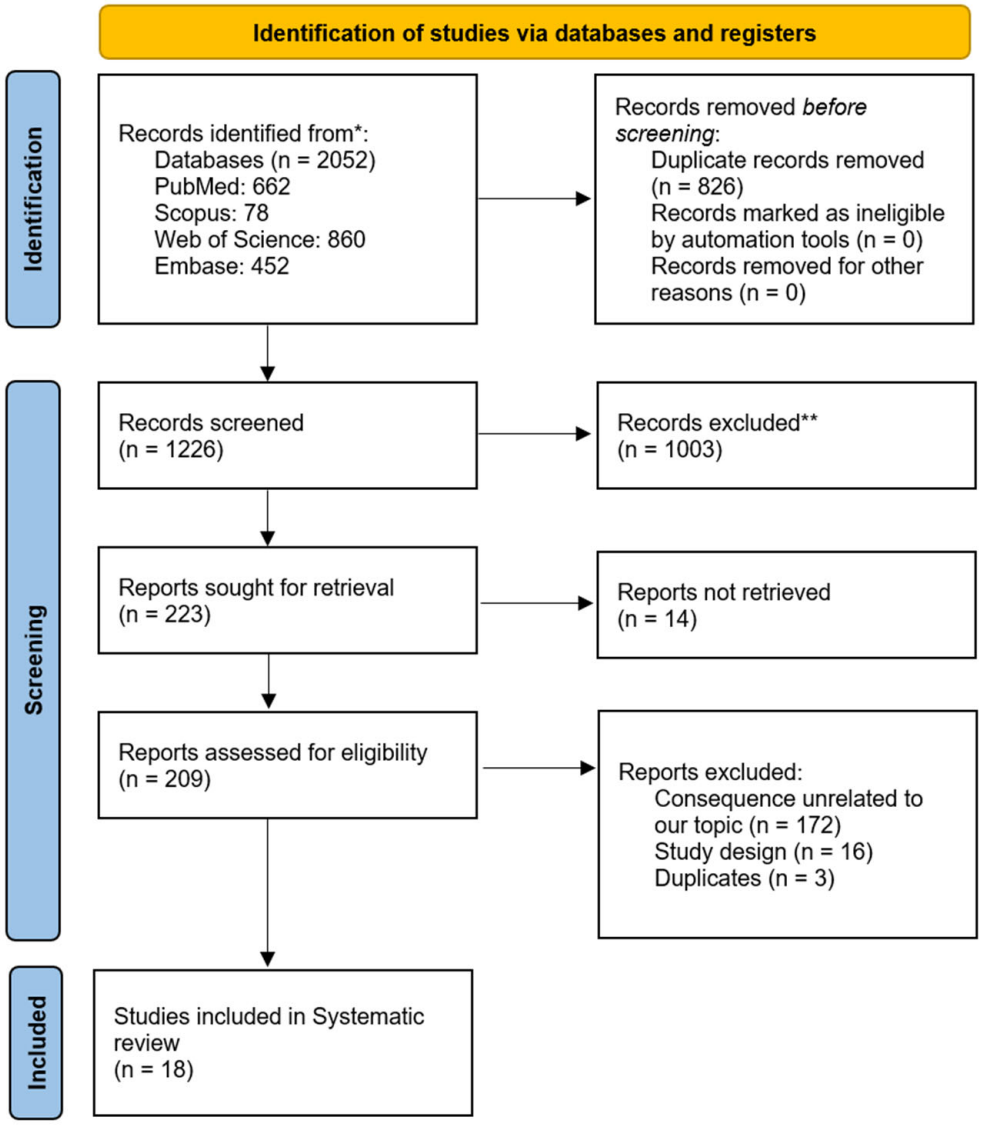


**Figure 1.** Flow diagram for selection of studies

## Methods

### Protocol and registration

This systematic review was conducted and reported in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines (Appendix 1) [19].

### Eligibility criteria

Studies were included based on the population, intervention, comparison, outcomes, and study design (PICOS) framework:

Population (P): Practicing nurses in any clinical setting. Intervention (I): Use of any AI-based tool (e.g. NLP, ML, and ASR) to support or automate nursing documentation. Comparison (C): Standard documentation workflows or pre-intervention conditions. Outcomes (O): Measures related to time management (e.g. documentation time, workload) and/or patient safety (e.g. documentation errors, quality metrics). Study Designs (S): Primary research studies, including randomized controlled trials (RCTs), non-randomized interventional studies, cohort studies, case-control studies, and cross-sectional studies.

Publications were excluded if they were reviews, editorials, conference abstracts, case reports, or non-empirical papers. Studies describing only the technical

development of an AI tool without clinical evaluation were also excluded.

### Information sources and search strategy

A systematic search was conducted in PubMed, Scopus, Web of Science, and Embase, with no date or language restrictions applied. To ensure comprehensive coverage, the search was supplemented by screening results from Google Scholar and by forward/backward citation tracking using Research Rabbit.

### Study selection and data management

After removing duplicates using EndNote software, version 21, two authors independently screened all titles and abstracts on the Rayyan platform. The full texts of potentially eligible articles were subsequently reviewed for final inclusion. Disagreements at either stage were resolved through consensus discussion or, when necessary, adjudication by a third reviewer.

### Data extraction and risk of bias assessment

Two reviewers independently extracted key data from included studies using a standardized form. Extracted fields included study characteristics, population, intervention details, and quantitative or qualitative outcomes relevant to the research questions. The risk of bias for each study was independently assessed by two reviewers using the appropriate Joanna Briggs institute (JBI) critical appraisal checklist for the study's design [20].

### Data synthesis

A quantitative meta-analysis was precluded by the significant heterogeneity observed across AI interventions, clinical settings, and outcome measures. Consequently, a narrative synthesis was employed to systematically integrate and interpret the findings from the included studies.

## Results

### Study selection

The initial database search yielded 2,052 records. After 826 duplicates were removed, the titles and abstracts of the remaining 1,226 records were screened. From this cohort, 1,003 records were excluded for not meeting the inclusion criteria. The full texts of the remaining 223 reports were searched for retrieval, of which 14 reports could not be found. Subsequently, the full texts of 209 reports were assessed for eligibility, and from this, 191 reports were excluded for reasons detailed in the PRISMA flow diagram (Figure 1). Ultimately, 18 studies met the eligibility criteria and were included in this narrative synthesis [16, 21-35].

### Characteristics of included studies

The 18 included studies, published between 2009 and 2025, represented a methodologically and geographically diverse body of evidence, as detailed in Table 1. The studies were conducted across several countries, including the United States, South Korea, the United Kingdom, Germany, China, and Taiwan. Research designs were diverse, reflecting the nascent state of the field, and included quasi-experimental studies, practical trials, observational studies, qualitative studies, and proof-of-concept.

As shown in Table 1, study populations were heterogeneous, ranging from small, focused groups of nurses (e.g. n=11 in King, 2023 [23]) to large-scale analyses involving tens of thousands of patient records [22, 29]. The AI interventions examined were also diverse, spanning a spectrum from established decision support systems to cutting-edge generative AI, with the specific tool type for each study detailed in Table 2.

### Risk of bias in the included studies

The methodological quality of the included studies was assessed to determine the risk of bias. Of the 18 studies included in this systematic review, 16 were critically appraised using the appropriate JBI critical appraisal checklist [20] corresponding to their respective study designs. For the remaining two studies, a suitable JBI checklist could not be identified; Consequently, their findings were reviewed with greater sensitivity and caution regarding their methodological limitations.

A summary of the risk of bias assessment for these studies, categorized by design, is presented in Figure 2 for cross-sectional, Figure 3 for quasi-experimental, and Figure 4 for qualitative studies.

A clear divergence in methodological rigor was observed across the different study designs. The two analytical cross-sectional studies [27, 28] were both assessed as having a low risk of bias across all applicable domains, indicating a high degree of methodological quality.

**Table 1.** Summary of the included studies on AI-based nursing documentation

| Authors (y) | Study Type | Location | Intervention Setting | Participants | Population | Outcome Measures | Key Finding(s) |
|---|---|---|---|---|---|---|---|
| Ju et al. (2025) [21] | Quasi-Experimental (Pre-Post) | South Korea | Online environment (virtual nursing simulation) with hypothetical EMR | 40 nurses | Nurses with at least 6 months of clinical experience | Time to complete documentation; Documentation accuracy and comprehensiveness | Generative AI reduced documentation time by 61%, but requires human validation for accuracy. |
| Levin et al. (2021) [22] | Quasi-Experimental (Controlled ITS) | USA | 600-bed academic hospital (4 inpatient units) in Baltimore | 12,470 patients | Adult inpatients in internal medicine, surgical, and telemetry units | LoS; Prediction performance (AUC) | LoS prediction tool reduced hospital stay by over 12 hours. |
| King et al. (2023) [23] | Qualitative (Interview) | USA | Academic medical center, perioperative unit (St. Louis, Missouri) | 11 nurses | (PACU & ACCS Wards) involved in patient transfer process | Situational awareness and facilitation of patient transfer; Barriers to AI use (qualitative) | AI has the potential to increase situational awareness in handoffs, but risks information overload. |
| Chen et al. (2024) [24] | Observational (Pre-Post) | Taiwan | Chi Mei Medical Center (ICU and general wards) | Unknown | Nurses involved in documentation in ICU and general wards | Documentation time; Documentation quality and accuracy | Generative AI reduced documentation time by 67% and structured records. |
| Dos Santos et al. (2024) [25] | Methodological/Proof of Concept | USA | Non-clinical (simulated clinical environment with oncology scenario) | 3 nursing specialists | Evaluation of ChatGPT's ability to generate an oncology care plan | Care plan quality; Accuracy of standard terminology (SNT) | AI generated care plans with 69% SNT accuracy and proposed new interventions. |
| Ozonoff et al. (2022) [26] | Observational System Development (Pre/Post) | USA | Boston Children's Hospital (pediatric specialty hospital) | 60,375 clinical notes (3150 patients) | Inpatient nurses' shift notes (focus on PIVIE events) | Identification of safety events; model classification accuracy and sensitivity | NLP was able to identify 35% of missed safety events. |
| Poon et al. (2025) [27] | Cross-Sectional Survey | USA | National survey of senior leaders in 43 non-profit health systems | 43 health systems | Senior executives (CMIOs, CIOs) reporting on AI adoption status | AI adoption status and success rate; Main barriers to AI adoption | Ambient Notes is the only tool with 100% adoption, but AI immaturity (77%) is the biggest barrier. |
| Duggan et al. (2025) [16] | Quality Improvement (Pre-Post) | USA | Academic health system (17 specialties in outpatient setting) in Philadelphia | 46 clinicians (physicians, NP, PA) | Outpatient clinicians from 17 specialties | Time spent on notes and closing encounters; After-hours work time and cognitive load | Ambient AI reduced after-hours work time by 30% but requires significant editing. |
| Danello et al. (2009) [36] | Retrospective Observational (Case Study) | USA | St. Joseph's/Candler Health System (5-year case study) | Unknown (987 nurses) | Nurses and other caregivers using smart IV pumps | ADEs, ROI | CDS system directly prevented 471 ADEs. |
| von Wedel et al. (2022) [28] | National Multiple Regression Analysis | Germany | National analysis of 383 German hospitals (using QSR data) | 383 hospitals (nursing and medical staff) | Perceived value of HIT and EHR by clinical users | Clinical outcomes (O/E Ratio); Patient satisfaction and perceived value by user | HIT adoption alone has no effect; user perceived value is the main factor for improving clinical outcomes and patient satisfaction. |

| Authors (y) | Study Type | Location | Intervention Setting | Participants | Population | Outcome Measures | Key Finding(s) |
|---|---|---|---|---|---|---|---|
| Cho et al. (2021) [29] | Quasi-Experimental (CITS) | South Korea | Academic hospital (12 internal-surgical units) | 42,476 admissions (204 nurses) | Adult inpatients in internal-surgical units (fall prevention) | Patient fall rate; Injurious fall rate and LoS | ML tool led to an immediate 29.7% reduction in fall rate. |
| Baxter et al. (2024) [30] | Retrospective Qualitative (Perspective) | USA | Academic health system (UCSD Health, California) | 50 negative patient messages | LLM response to negative patient EHR messages | Communication appropriateness; Response length and risk of escalating communication (qualitative) | LLMs lack empathy and risk escalating communication, and human oversight is crucial. |
| Cho et al. (2023 A) [33] | Pragmatic Clinical Trial (Pre-Post) | South Korea | Tertiary academic hospital (6 nursing units) | 40,839 patients | Inpatients (over 18), focusing on reducing injurious falls | Injurious fall rate; Fall rate and age-subgroup outcomes | CDS led to a significant 33.8% reduction in injurious fall rate. |
| Balloch et al. (2024) [32] | Simulation Study (Crossover) | UK | Simulated clinic (Great Ormond Street Children's Hospital) | 8 experienced clinicians | Outpatient patient consultation (to assess documentation quality) | Documentation quality (SAIL Score); Consultation length and perceived task load | Ambient AI more than doubled documentation quality and reduced consultation time by 26.3%. |
| Cho et al. (2023 B) [31] | Multi-center Longitudinal Observational | South Korea | 4 multi-center hospitals (with different EHRs) | 103,723 patients | Implementation of an AI tool for fall prevention and semantic interoperability | Patient fall rate (longitudinal); Nursing activities and model performance | Semantic interoperability with SNTs was successful, but no significant longitudinal reduction in fall rate was observed. |
| Johnson et al. (2024) [37] | Methodological / Proof of Concept | USA | Non-clinical (simulated clinical environment - perinatal focus) | 7 Nursing researchers | Evaluation of ChatGPT's ability to generate a perinatal care plan | Adequacy of care plan; SNT accuracy and care prioritization | Generative AI proved to reduce the cognitive load of care plan generation and produced plans with 66.7% accurate SNT. |
| Yang et al. (2024) [35] | Prospective Observational | China | Tertiary hospital (outpatient) in Shanghai | 306 patients | Comparison of AI triage system accuracy and time with triage nurses | Triage accuracy; Triage time and Recall | The AI system reduced triage time by 4.22 seconds but was less accurate than nurses. |
| Johnson et al. (2024) [34] | Mixed-Methods Case Study | UK | Teaching hospital (control site) and hospital with a command center (CC Site) | 36 Staff (interview & observation) | Hospital operational management, patient flow, and safety at a system-wide level | Patient flow and patient safety; Data quality and staff perspectives (qualitative) | Small quantitative impact was proven; however, CC qualitatively facilitated operational management during a crisis. |

JRH

Abbreviations: AI: Artificial intelligence; AUC: Area under the curve; AUROC: Area under the receiver operating characteristic; CDS: Clinical decision support; CITS: Controlled interrupted time series; CC: Command center; CMIO: Chief medical information officer; CIO: Chief information officer; EHR: Electronic health record; EMR: Electronic medical record; HIT: Health information technology; ICU: Intensive care unit; ITS: Interrupted time series; LLM: Large language model; LoS: Length of stay; ML: Machine learning; NP: Nurse practitioner; PA: Physician assistant; PACU: Post-anesthesia care unit; PIVIE: Peripheral intravenous infiltration or extravasation; ROI: Return on investment; SNT: Standardized nursing terminology; USA: United States of America; UK: United Kingdom.

**Table 2.** Types and functional features of AI-driven interventions evaluated in the included studies

| Authors (y) | AI Tool Type | Tool Function (Support/Automation) | Comparator (Control Group) | Clinical Domain | Implementation | Reported OR, RR, P, CI, AUC |
|---|---|---|---|---|---|---|
| Ju et al. (2025) [21] | Generative AI (LLM-ChatGPT-3.5) | Automation (generation of coded nursing diagnoses, outcomes, and interventions) | Standard EHR documentation (Manual text entry) | Nursing care plan documentation | Cross-sectional intervention (1-day simulation) | P<0.001 |
| Levin et al. (2021) [22] | ML | Support (predicting patient discharge time (LoS) to improve early discharges) | Control hospital/pre-intervention period | Patient flow management and discharge timing | 32 months (from Jan 2017) | AUC=0.7-0.8 P>0.05 |
| King et al. (2023) [23] | NLP | Support (identifying key points in notes to improve patient handoff reporting) | Standard Handoff (SBAR) | Perioperative nursing handoff | Cross-sectional (interviews) | N/A |
| Chen et al. (2024) [24] | Generative AI (LLM-ChatGPT) | Automation (generation of nursing note drafts (AI-Generated Content)) | Manual documentation in EHR | General and ICU ward documentation | 3-month implementation phase | N/A |
| Dos Santos et al. (2024) [25] | Generative AI (ChatGPT-4) | Support (generation of coded oncology care plan suggestions with SNTs) | Nurse-generated standard care plan | Oncology care plan documentation | Cross-sectional (Testing on one single scenario) | N/A |
| Ozonoff et al. (2022) [26] | NLP | Support (screening nursing notes for safety event detection (PIVIE events)) | Traditional safety event surveillance system | Patient safety (adverse event detection) | 2-year retrospective analysis | CI: 12.1–14.1 |
| Poon et al. (2025) [27] | Predictive & Generative AI (System Report) | Support & Automation (systematic adoption of AI tools at the organizational level) | N/A (survey study) | Health system AI strategy and operations | Cross-sectional (executive survey) | N/A |
| Duggan et al. (2025) [16] | Ambient AI (Ambient Scribe - Nuance DAX) | Automation (automatic conversation summarization to create note drafts) | Standard manual or dictated documentation | Outpatient documentation | 3-month pre/post intervention | P<0.001 |
| Danello et al. (2009) [36] | CDS system | Support (dose and rate alerts, soft/hard limits drug library) | Prior-generation smart IV pumps | Patient safety (medication error prevention) | 5-year retrospective analysis | N/A |
| von Wedel et al. (2022) [28] | HIT/EHR | Support (use of digital systems in documentation and processes) | Hospitals with lower digitization levels | Perceived value of digital systems (HIT/EHR) | Nationwide cross-sectional study | P=0.01 |
| Cho et al. (2021) [29] | Predictive Analytics (ML) | Support (Clinical Decision Support System - CDS for fall risk) | Control units not receiving the CDS | Patient safety (fall prevention) | 12-month study period | P=0.039 |
| Baxter et al. (2024) [30] | Generative AI (LLM) | Automation (generating draft responses to negative patient messages) | Hypothetical standard clinical responses | Patient-provider communication (EHR messages) | Cross-sectional (retrospective message analysis) | N/A |
| Cho et al. (2023 A) [33] | ML | Support (fall risk alert and targeted intervention recommendations) | Baseline phase before CDS intervention | Patient safety (reduction of injurious falls) | 5-month pre/post intervention | N/A |
| Balloch et al. (2024) [32] | Ambient AI (GPT-4) | Automation (automatic summarization of audio to clinical note and letter) | Standard EHR documentation (Manual typing) | Outpatient consultation documentation | Cross-sectional (1-day simulation) | P=0.03 |

| Authors (y) | AI Tool Type | Tool Function (Support/Automation) | Comparator (Control Group) | Clinical Domain | Implementation | Reported OR, RR, P, CI, AUC |
|---|---|---|---|---|---|---|
| Cho et al. (2023 B) [31] | Machine Learning (Bayesian Network) | Support (predicting fall risk and recommending care plans) | Different EMRs (using SNTs for interoperability) | Patient safety (fall prevention) | 4-year longitudinal analysis | N/A |
| Johnson et al. (2024) [37] | Generative AI (ChatGPT-4) | Support (generation of comprehensive perinatal nursing care plan suggestions) | Nurse-generated standard care plan | Perinatal nursing care plan documentation | Cross-sectional (testing on one single scenario) | N/A |
| Yang et al. (2024) [35] | Natural Language Processing (NLP - BERT) | Automation (recommending outpatient departments based on patient's chief complaint) | Experienced Triage Nurses | Outpatient triage and referral | 7-day prospective observation | P<0.001 |
| Johnson et al. (2024) [34] | AI/Decision Support in Command Centre (CC) | Support (data aggregation and display for improved situational awareness) | Control hospital (without Command Centre) | Hospital operations management and patient flow | 2.5-year Interrupted Time Series Analysis | N/A |

Abbreviations: AI: Artificial intelligence; AUC: Area under the curve; AUROC: Area under the receiver operating characteristic; BERT: Bidirectional encoder representations from transformers; CC: Command centre; CDS: Clinical decision support; CI: Confidence interval; EHR: Electronic health record; EMR: Electronic medical record; HIT: Health information technology; ICU: Intensive care unit; ITS: Interrupted time series; LLM: Large language model; LoS: Length of stay; ML: Machine learning; N/A: Not applicable; NLP: Natural language processing; OR: Odds ratio; PIVIE: Peripheral intravenous infiltration or extravasation; RR: Relative risk; SNT: Standardized nursing terminology; USA: United States of America; UK: United Kingdom.

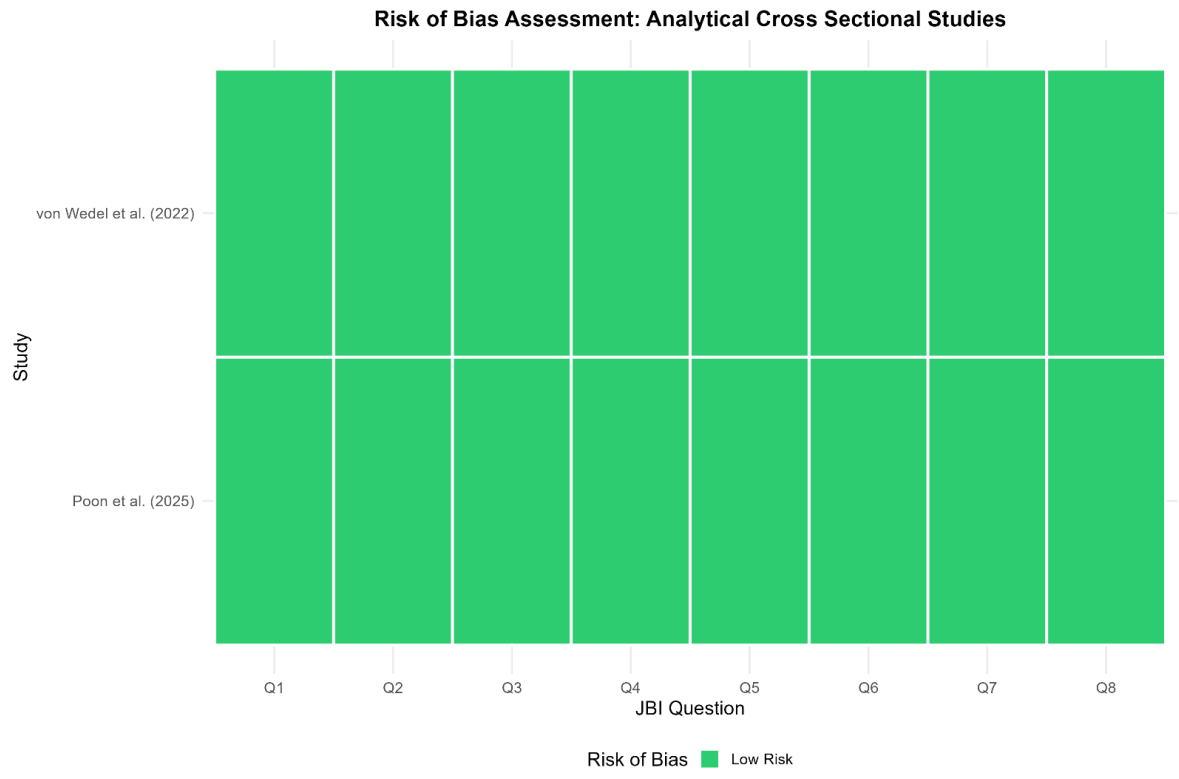


**Figure 2.** Risk of bias for cross-sectional studies

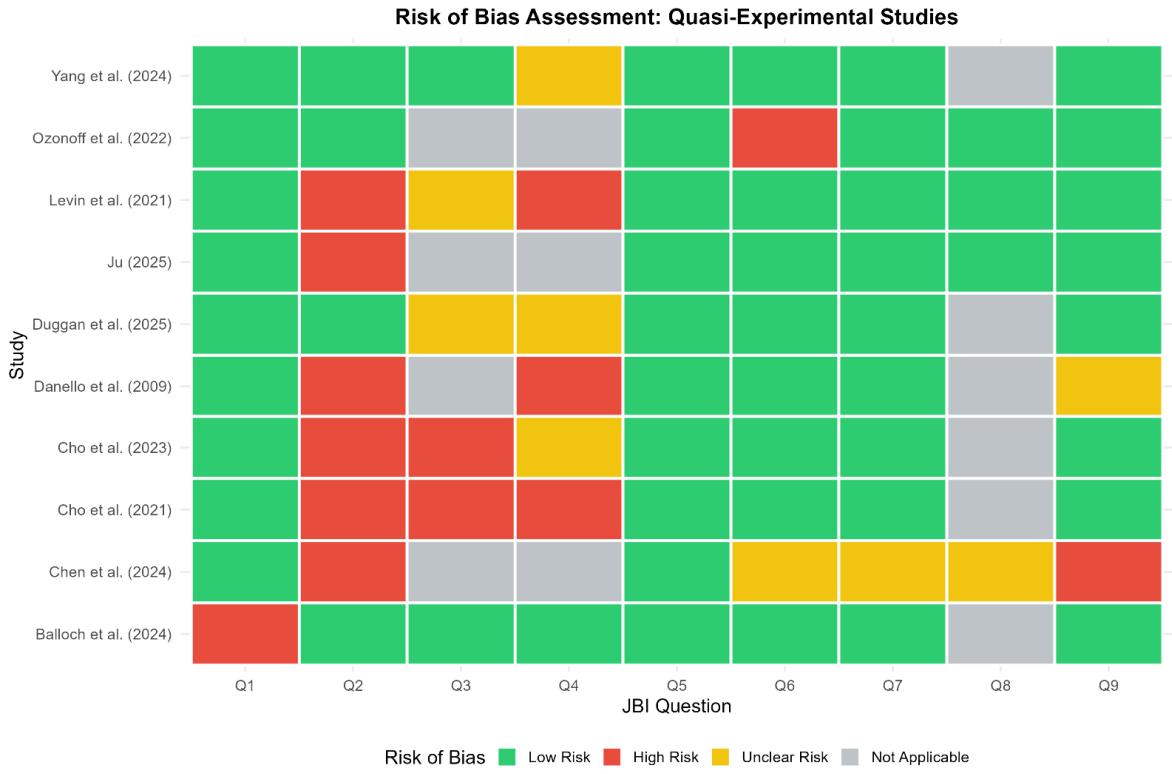


**Figure 3.** Risk of bias for quasi-experimental studies

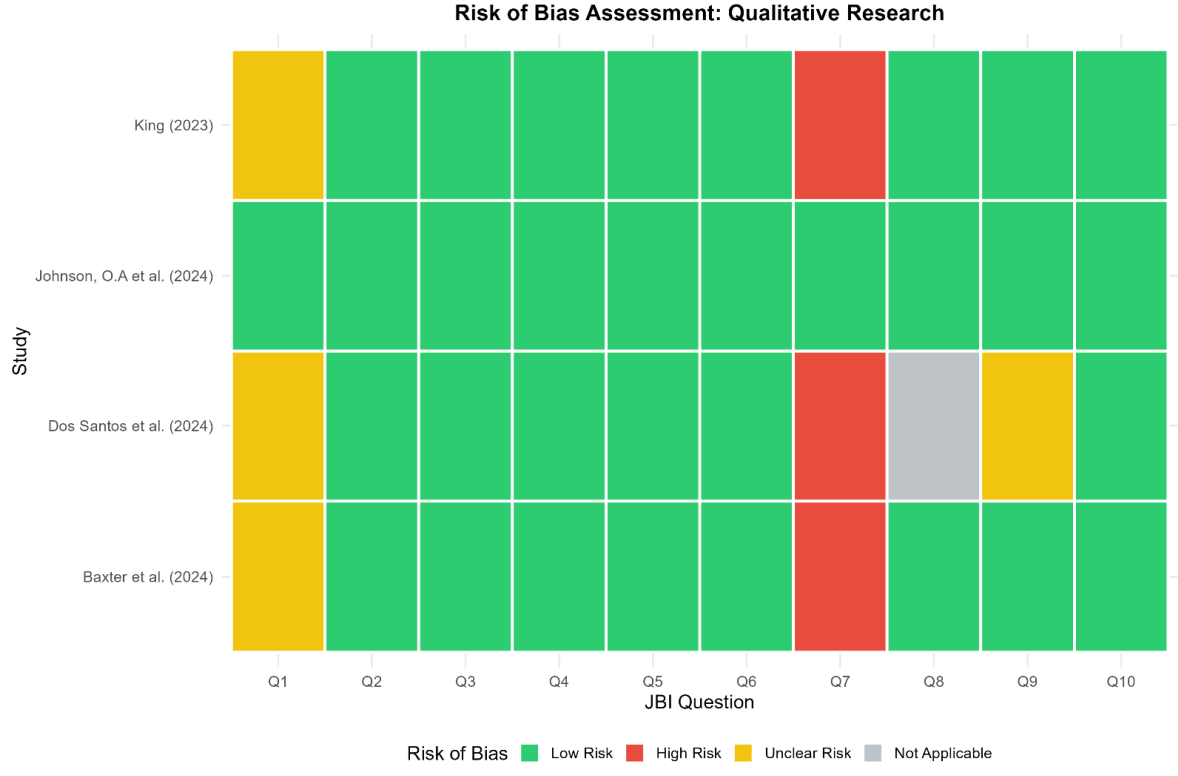


**Figure 4.** Risk of bias for qualitative studies

**Table 3.** Impact of AI-based documentation tools on time management and clinical efficiency

| Authors (y) | Outcome Measure | Change (%)/P | Result(s) (Qualitative Findings and Clinician Feedback) |
|---|---|---|---|
| Ju et al. (2025) [21] | Documentation Time (seconds) | -61/<0.001 | AI reduces the cognitive and physical load of documentation; AI is not fully reliable. |
| Chen et al. (2024) [24] | Documentation Time (minutes) | -66.7/N/A | AI reduces documentation time and enhances documentation quality. |
| Duggan et al. (2025) [16] | Time Spent on Notes (minutes) | -20.4/<0.001 | AI Reduces mental burden. |
| Duggan et al. (2025) [16] | After-Hours Work Time (minutes/day) | -30.0/0.02 | Clinicians have more time for patient care and can reduce burnout. |
| Balloch et al. (2024) [32] | Total Consultation Time (minutes) | -26.3/0.03 | AI reduces perceived task load. |
| Yang et al. (2024) [35] | Triage Time (seconds) | -29.5/<0.001 | AI triage is faster but less accurate than manual triage. |
| Levin et al. (2021) [22] | Length of Stay (LoS) | N/A/<0.002 | The intervention improves patient flow and reduces hospital LoS in medicine and telemetry units. |
| Cho et al. (2021) [29] | Mean LoS (days) | -1.29 days/<0.001 | The AI tool led to a 16% decrease in LoS. |
| Danello et al. (2009) [36] | ROI over 5 years | N/A/N/A | CDS led to a positive ROI over 5 years. |
| Baxter et al. (2024) [30] | Length of LLM Response (words) | N/A/N/A | LLM responses were consistently longer. |
| Johnson et al. (2024) [37] | Time for Documentation | Reduces cognitive load; easy to use and a helpful tool | AI can reduce cognitive load and is a helpful tool for documentation. |
| Johnson et al. (2024) [34] | LoS / Patient Flow | Not Significant (P>0.05) | The CC site qualitatively facilitated operational management during a crisis. |

Abbreviations: AI: Artificial intelligence; CC: Command centre; CDS: Clinical decision support; LoS: Length of stay; LLM: Large language model; N/A: Not applicable; ROI: Return on investment.

The four qualitative studies demonstrated a generally low-to-moderate risk of bias. While strengths were noted, some domains presented potential for bias, with an unclear risk noted in Question 1 (Q1) and a high risk in Q7 for several studies [23, 25, 30].

The ten quasi-experimental studies represented the most methodologically heterogeneous group and contained the most significant potential for bias. High risks of bias were frequently identified in domains concerning participant allocation and control of confounding variables (notably Q2, Q3, and Q4). Furthermore, an unclear risk was common across domains related to follow-up and outcome measurement (Q6, Q7, and Q9). These limitations suggest that the evidence from this substantial portion of the included literature should be interpreted with considerable caution.

In summary, while the overall body of evidence provides valuable insights, the variable methodological quality, particularly within the quasi-experimental designs, underscores the need for careful consideration of the potential for bias when interpreting the findings of this review.

## Synthesis of findings

The findings from the 18 included studies were synthesized into four main thematic areas: 1) impact on time management and efficiency, 2) impact on documentation quality, 3) impact on direct patient safety outcomes, and 4) the critical role of the human-in-the-loop and emergent risks.

## Impact on time management and efficiency

One of the most consistent findings across the included literature was the potential for AI to significantly reduce the time spent on documentation and improve clinical efficiency. As synthesized in Table 3, studies utilizing generative AI or ambient scribe technology reported substantial time savings. For instance, Ju et al.

**Table 4.** Impact of AI-based interventions on direct patient safety events

| Authors (y) | Outcome Measure | AI Findings (Results/Metrics) | Hallucinations/Human-in-the-Loop | Clinical Implications/Staff Feedback |
|---|---|---|---|---|
| Danello et al. (2009) [36] | ADEs (medication events) | CDS system prevented 471 adverse events. | Human-in-the-loop: The system provided alerts for nurses and physicians to review and take action. | AVIDERS, a smart pump system, is a useful tool for preventing ADEs. |
| Ozonoff et al. (2022) [26] | PIVIEs (safety events) | NLP model identified 44% of safety events. | N/A | The AI system can identify previously missed safety events. |
| Cho et al. (2023 A) [33] | Fall Rate (per 1000 days) | The CDS decreased falls by 0.45 per 1000 days. | N/A | CDS implementation significantly reduced injurious falls (P=0.02) but had no significant effect on overall fall rate (P=0.18). |
| Cho et al. (2021) [29] | Fall Rate (per 1000 days) | The AI model led to a 29.73% reduction in falls. | N/A | The AI fall prediction tool led to a 29.73% reduction in falls. |
| Cho et al. (2023 B) [31] | Semantic interoperability | N/A | N/A | The AI-powered CDS was successfully implemented across hospitals with high predictive accuracy (AUROC=0.81–0.96); although fall rates did not decrease significantly (P=0.16), nursing intervention rates increased substantially. |
| Balloch et al. (2024) [32] | SAIL Score - Notes Quality | Documentation quality was improved by 100%. | AI-generated text contained no hallucinations in the two cases evaluated. | Reduced perceived task load. |
| Ju et al. (2025) [21] | Documentation time (seconds) (5 notes) | N/A | AI had a low hallucination rate () and a high error rate (). | AI reduces the cognitive and physical load of documentation, but it is not fully reliable and requires human validation. |
| Duggan et al. (2025) [16] | Note Length (characters) | N/A | AI-generated notes had a 20.6% rate of unvalidated (phone call) information. | Note Bloat: AI-generated notes were longer and had a higher rate of errors, which increased cognitive load for nurses. |
| Dos Santos et al. (2024) [25] | SNT accuracy (69%) | AI had an SNT accuracy of 69%. | N/A | AI is a promising tool for creating care plans. |
| Johnson et al. (2024) [37] | SNT Accuracy (66.7%) | AI had an SNT accuracy of 66.7%. | Human-in-the-loop: LLM plans need to be reviewed by nurses for accuracy. | AI reduces the cognitive burden of care plan generation. |
| Baxter et al. (2024) [30] | Lack of empathy/ escalation | N/A | The AI lacked empathy and could escalate situations. | AI lacks empathy and could escalate situations. |
| King et al. (2023) [23] | Situational Awareness | N/A | AI can increase situational awareness but risks information overload and cognitive burden. | AI could facilitate handoffs. |
| Chen et al. (2024) [24] | Documentation Quality/Time | N/A | The AI model had no hallucinations. | AI reduces documentation time and improves quality. |
| Yang et al. (2024) [35] | Triage accuracy | Accuracy: 91.53% | The AI system made a mistake in 38% of cases, primarily related to the patient's chief complaint. | The AI system is faster but less accurate than nurses. |
| Johnson et al. (2024) [34] | Data quality/patient safety | N/A | N/A | A CC qualitatively facilitated operational management during a crisis. |
| von Wedel et al. (2022) [28] | EHR user value | N/A | N/A | Perceived value by the user is a stronger predictor of clinical outcomes than HIT adoption. |
| Poon et al. (2025) [27] | Risk stratification | 38% of risk stratification was automated. | N/A | AI is not yet mature, and there are many barriers to adoption. |

Abbreviations: ADE: Adverse drug event; AI: Artificial intelligence; AUROC: Area under the receiver operating characteristic; CC: Command centre; CDS: Clinical decision support; EHR: Electronic health record; HIT: Health information technology; LLM: Large language model; N/A: Not applicable; NLP: Natural language processing; PIVIE: Peripheral intravenous infiltration or Extravasation; SNT: Standardized nursing terminology; SAIL: Structure, assessment, integration, and logic (documentation quality metric).

demonstrated an approximate 61% reduction in documentation time (from 467 to 183 seconds, P<0.001), while Chen et al. reduced this time from 15 to approximately 5 minutes per patient. Similarly, Duggan et al. observed a 20.4% decrease in time spent in notes and a 30% reduction in after-hours work (P=0.02), and Balloch et al. found that AI-assisted consultations were 26.3% shorter (P=0.03) [16, 21, 24, 32].

These benefits extended beyond note-writing to system-level efficiency. As shown in Table 3, predictive AI tools were associated with significant reductions in patient length of stay [22, 29] and expedited clinical processes, like outpatient triage [35]. However, this time-saving effect was not absolute. Qualitative findings from several studies [16, 30] revealed that the time saved in initial content generation was often reallocated to the critical task of reviewing and editing the AI output to ensure clinical accuracy and safety [16, 22, 29, 30, 35].

### Impact on documentation quality

A second major theme was the capacity of AI to enhance the quality and completeness of clinical documentation. Evidence for this is summarized in Table 4. A simulation study by Balloch et al. provided compelling quantitative data, with 100% of AI-generated clinical notes receiving a "good/very good" score on the validated SAIL instrument, compared to only 43% of manually written notes (P=0.004) [32].

Proof-of-concept studies, also detailed in Table 4, demonstrated this potential as well. AI-generated care plans were found to be not only comparable to those created by expert nurses, but in some cases, enhanced them by providing additional relevant interventions [25, 34]. This was supported by findings of increased overall note length and completeness [16]. However, this improvement in structural quality was coupled with concerns about content accuracy. Both Dos Santos et al. and Johnson et al. found that AI correctly used standard nursing terminologies in only about two-thirds of instances. Qualitative feedback from clinicians consistently emphasized the need for significant editing to correct factual errors and refine clinical nuances [25, 37].

### Impact on direct patient safety events

The direct impact of AI on patient safety events, detailed in Table 4, was complex and context-dependent. Decision support systems targeting specific risks showed clear benefits. The study by Danello et al. on smart IV pumps reported the prevention of at least 471 adverse drug events over five years. Similarly, an NLP-based tool developed by Ozonoff et al. was able to identify 44% of safety events that were missed by the hospital's existing reporting systems [26, 36].

Conversely, evidence regarding patient fall prevention, a key nursing-sensitive indicator, was contradictory. An early controlled study showed an immediate 29.7% reduction in the overall fall rate after AI implementation [29], and a subsequent pragmatic trial reported a significant reduction in injurious falls (P=0.0171) even if the overall fall rate did not change [33]. However, a large, multi-center implementation study found no significant long-term change in fall rates [31]. Furthermore, a large study of an AI-powered hospital command center found no discernible impact on macro-level safety indicators, like mortality or readmission rates [34], suggesting that the effectiveness of AI is highly contingent on the clinical context and the specific safety outcome being measured.

### The role of the human user and emerging risks

A critical theme that emerged from the synthesis is that AI in nursing documentation functions as a powerful tool, not an autonomous agent. This highlights both the irreplaceable role of the clinician and the new risks introduced by the technology itself. The study by Baxter et al. [30] provided a cautionary illustration of these risks. When a generative AI was used to draft responses to patient messages, it not only lacked empathy but dangerously suggested a patient file a formal complaint against their physician, revealing the potential to harm the therapeutic relationship and create medico-legal issues. As noted across multiple studies [24, 32], this risk of generating false or inappropriate content, often termed "hallucination," underscores the non-negotiable need for human clinical judgment to review, edit, and validate all AI-generated outputs before they enter the official patient record.

Finally, the effectiveness of any AI tool was intrinsically linked to its integration into the clinical workflow and its acceptance by end-users. A large national study in Germany by von Wedel et al. concluded that positive clinical outcomes did not correlate with the mere adoption of technology, but rather with its "user-perceived value." This finding underscores the paramount importance of user-centered design and ensuring that AI tools support, rather than disrupt, the complex cognitive work of nurses [28].

## Discussion

This systematic review synthesized evidence from 18 included studies on the impact of diverse AI technologies on nursing documentation, focusing on the dual outcomes of time management and patient safety. Our findings revealed a rapidly innovating field with promising initial results, yet one simultaneously marked by significant methodological heterogeneity and the emergence of critical socio-technical challenges. The principal findings of this review can be organized into four core themes:

### Efficiency and time management: A paradigm shift from creation to curation

AI interventions, particularly generative and ambient scribe technologies, consistently demonstrate the potential to alleviate the clinical administrative burden. Our synthesis quantitatively confirms that AI can substantially reduce time spent on documentation, with simulation studies reporting reductions exceeding 60% [21]. Furthermore, in real-world outpatient settings, time spent in notes decreased by 20.4%, and significant relief was found in "after-hours work time" (Pajama Time), which saw a reduction of 30.0% [16]. These findings strongly support the primary hypothesis that AI can liberate nursing time for direct patient interaction and theoretically mitigate clinician burnout.

However, a crucial finding is that this time-saving is not absolute. The "raw output" generated by AI consistently requires meticulous human review, editing, and clinical validation. Qualitative feedback indicates that time reallocated to editing inaccuracies, correcting factual errors, and adding clinical nuance can sometimes equate to or even outweigh the time initially saved in content generation [16]. This mandates a reframing of the net impact on time: AI drives a strategic shift in clinical workload from content creation to content curation.

### Patient safety: Bifurcated impact and the emergence of new risks

The impact of AI on patient safety is complex and multifactorial, lacking uniform positivity. Our analysis identified three distinct pathways of influence:

Direct safety improvement (hard outcomes): Certain AI tools designed to target specific, measurable risks demonstrated clear, quantifiable safety gains. This includes decision support systems (CDS), like smart IV pumps, which objectively averted ≥471 preventable adverse drug events (ADEs) over five years [36]. Similarly, NLP successfully identified 35% of safety events, potentially identifiable and preventable inpatient safety events (PIVIEs) missed by conventional hospital reporting systems [26].

Indirect safety improvement via documentation quality: Studies involving generative AI often linked potential safety improvements to enhancements in documentation quality. AI tools demonstrated an ability to generate notes that were more structured, comprehensive [24], and adhered better to standardized nursing terminologies (SNTs), albeit imperfectly [25, 34]. The implicit mechanism is that better data quality supports clearer communication and safer care.

Emergence of novel safety risks: Critically, our review highlights that AI introduces new vectors for patient harm. The risk of AI "hallucinations" (factual inaccuracies) was noted even in controlled environments [21, 32]. More concerningly, LLMs demonstrated potential for severe socio-technical harm, generating responses to patient messages that lacked empathy and carried the risk of escalation, even suggesting filing formal complaints [30]. This complexity strongly reinforces the non-negotiable need for robust "human-in-the-loop" protocols to manage the intersection of AI capabilities and clinical responsibility.

### Methodological landscape: The critical need for rigor

The evidence base supporting AI in nursing documentation remains nascent. The included studies were predominantly limited to quasi-experimental pre-post designs or simulation/qualitative approaches. While valuable for initial evaluation and hypothesis generation [23], this heterogeneity prevented a quantitative meta-analysis on efficiency outcomes and limits the ability to draw firm causal conclusions.

A key limitation is the inconsistency in findings regarding the strongest clinical outcome, fall prevention. While some studies reported immediate success [29]. Cho et al. reported a 29.7% immediate fall reduction, a larger longitudinal implementation failed to detect a sustained, significant decrease in the overall fall rate [31]. This lack of definitive evidence underscores a critical need for more methodologically rigorous designs, particularly well-executed controlled interrupted time series (CITS) and RCTs, to confidently attribute

observed gains to the intervention rather than confounding factors.

### Implementation success: The socio-technical imperative

Our findings suggest that the successful integration of AI is not purely a technical exercise but a socio-technical challenge. The national analysis by von Wedel et al. was particularly insightful, concluding that clinical benefits were correlated not merely with technology adoption, but strongly with the "user-perceived value" of the installed systems (β=−0.138) [28].

This reinforces the importance of human factors. Qualitative data highlighted that poor system design can lead to alarm fatigue [23] or create a negative organizational culture where front-line staff feel "monitored" [34]. Therefore, effective AI integration must prioritize user-centered design and maintain the centrality of human clinical judgment and professional autonomy. The transition of the nurse's role from content creator to critical curator demands comprehensive training and a governance framework that empowers clinicians to support, not override, AI-generated recommendations.

## Conclusion

The collective evidence demonstrates that AI-based documentation tools markedly improve time efficiency and workflow optimization across multiple clinical contexts. Most studies reported substantial reductions in documentation time, ranging from 20 % to over 60 %, as well as meaningful decreases in after-hours work and clinician cognitive burden. These time savings translate into improved patient flow, shorter length of stay, and enhanced opportunities for direct patient interaction. However, qualitative findings consistently emphasize that while AI systems significantly reduce the physical and mental load of documentation, they are not yet fully reliable for unsupervised clinical use. Therefore, the optimal integration of such systems requires human oversight and post-AI editing to ensure clinical accuracy, contextual relevance, and patient safety. In summary, AI holds clear potential to transform documentation efficiency and mitigate burnout, but sustained clinician engagement and rigorous validation are essential to achieve trustworthy, safe, and equitable implementation in healthcare settings.

### Limitations

This systematic review is subject to several methodological and practical limitations that should be considered when interpreting the findings. First, the primary constraint was the methodological immaturity and inherent heterogeneity across the included studies. The predominance of quasi-experimental (pre-post) designs, single-center trials, and simulation-based studies limits the ability to draw firm causal conclusions regarding the net impact of AI on patient care. Specifically, the high variability in AI technologies, patient populations, and inconsistent reporting of quantitative outcomes precluded a formal statistical meta-analysis of time-related endpoints.

Second, the generalizability (external validity) of the findings is constrained. Many of the studies evaluating efficiency gains were conducted at a single site or focused on niche populations, often relying on institution-specific EHR configurations. Furthermore, the reliance on simulation and proof-of-concept designs means that the full spectrum of operational and socio-technical challenges encountered in real-world clinical deployment may be underestimated.

Third, the review's scope was limited to formally published literature, meaning that relevant "grey literature" (e.g. technical reports, internal health system evaluations, or conference proceedings) that might contain valuable early evidence or operational insights could have been inadvertently missed.

Finally, the rapid pace of AI development, particularly within generative and ambient technologies, means that the evidence base is constantly and quickly evolving. The findings represent a snapshot in time and should be continuously evaluated against emerging technological capabilities and new implementation data.

### Challenges and future works

Despite the accelerating integration of AI into nursing documentation, several challenges remain unresolved. Data heterogeneity, limited interoperability across EHR systems, and inconsistent use of SNTs continue to constrain model performance and generalizability. Future research should prioritize the development of unified data frameworks and advanced natural language processing pipelines capable of handling multilingual and context-dependent nursing language. Another critical challenge is ensuring clinical validity and trustworthiness: AI-generated documentation must be continu-

ously validated through rigorous human-in-the-loop evaluation and transparent auditing to prevent misinformation and safeguard patient safety. The ethical and legal dimensions of AI use, particularly data privacy, algorithmic bias, and accountability, demand clear governance policies and interdisciplinary oversight. Finally, large, multi-center pragmatic trials are needed to evaluate the long-term impact of AI-assisted documentation on efficiency, patient outcomes, and clinician well-being, with special emphasis on user-centered design and real-world implementation science. Only through addressing these challenges can AI tools move from promising prototypes to reliable clinical companions in nursing practice.

## Ethical Considerations

### Compliance with ethical guidelines

This systematic review was conducted in accordance with established academic and ethical standards. As the work synthesized data solely from publicly available published studies and did not involve new data collection or patient/clinical interventions, Institutional Review Board (IRB) approval and participant consent were not required. The review protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) (Code: CRD420251089257). Reporting followed the PRISMA 2020 Statement to ensure clarity, reproducibility, and full disclosure of the review process.

### Funding

The authors received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Authors' contributions

Conceptualization, investigation and writing of the original draft: Arash Amadeh Taheri and Ghazaleh Afkhami Teimouri; Methodology: Yasamin Moeinipour and Arash Amadeh Taheri; Data collection: Yasamin Moeinipour, Arash Amadeh Taheri and Ghazaleh Afkhami Teimouri; Review, and editing: Yasamin Moeinipour and Fatemeh Bagheri; Supervision: Malihe Adineh Fathabadi; Project Administration: Arash Amadeh Taheri.

### Conflict of interest

The authors declared no conflict of interest.

### Acknowledgments

The authors would like to thank all individuals who contributed to this study through informal discussions or technical support but did not meet the criteria for authorship.

## References

[1] De Groot K, De Veer AJE, Munster AM, Francke AL, Paans W. Nursing documentation and its relationship with perceived nursing workload: A mixed-methods study among community nurses. BMC Nursing. 2022; 21(1):34. [DOI:10.1186/s12912-022-00811-7] [PMID]

[2] Ebbers T, Kool RB, Smeele LE, Dirven R, Den Besten CA, Karssemakers LHE, et al. The Impact of Structured and Standardized Documentation on Documentation Quality; a Multicenter, Retrospective Study. Journal of Medical Systems. 2022; 46(7):46. [DOI:10.1007/s10916-022-01837-9] [PMID]

[3] Wahyuni ED, Nursalam N, Dewi YS, Arifin H, Benjamin LS. Electronic nursing documentation for patient safety, quality of nursing care, and documentation: A systematic review. The Journal of the Pakistan Medical Association. 2024; 74(9):1669-77. [DOI:10.47391/JPMA.9996] [PMID]

[4] Shi Q, Wotherspoon R, Morphet J. Nursing informatics and patient safety outcomes in critical care settings: A systematic review. BMC Nursing. 2025; 24(1):546. [DOI:10.1186/s12912-025-03195-6] [PMID]

[5] Bakhoum N, Gerhart C, Schremp E, Jeffrey AD, Anders S, France D, et al. A Time and Motion Analysis of Nursing Workload and Electronic Health Record Use in the Emergency Department. Journal of Emergency Nursing. 2021; 47(5):733-41. [DOI:10.1016/j.jen.2021.03.007] [PMID]

[6] Moore EC, Tolley CL, Bates DW, Slight SP. A systematic review of the impact of health information technology on nurses' time. Journal of The American Medical Informatics Association. 2020; 27(5):798-807. [DOI:10.1093/jamia/ocz231] [PMID]

[7] Gesner E, Dykes PC, Zhang L, Gazarian P. Documentation Burden in Nursing and Its Role in Clinician Burnout Syndrome. Applied Clinical Informatics. 2022; 13(05):983-90. [DOI:10.1055/s-0042-1757157] [PMID]

[8] Nguyen OT, Shah S, Gartland AJ, Parekh A, Turner K, Feldman SS, et al. Factors associated with nurse well-being in relation to electronic health record use: A systematic review. Journal of the American Medical Informatics Association. 2021; 28(6):1288-97. [DOI:10.1093/jamia/ocaa289] [PMID]

[9] Mitha S, Schwartz J, Hobensack M, Cato K, Woo K, Smaldone A, et al. Natural Language Processing of Nursing Notes: An Integrative Review. CIN: Computers, Informatics, Nursing. 2023; 41(6):377-84. [DOI:10.1097/CIN.0000000000000967] [PMID]

[10] Trinh VQN, Zhang S, Kovoor J, Gupta A, Chan WO, Gilbert T, et al. The use of natural language processing in detecting and predicting falls within the healthcare setting: A systematic review. International Journal for Quality in Health Care. 2023; 35(4):mzad077. [DOI:10.1093/intqhc/mzad077] [PMID]

[11] Brann F, Sterling NW, Frisch SO, Schrager JD. Sepsis prediction at emergency department triage using natural language processing: Retrospective cohort study. Journal of Medical Internet Research Artificial Intelligence. 2024; 3:e49784. [DOI:10.2196/49784] [PMID]

[12] Peivandi S, Ahmadian L, Farokhzadian J, Jahani Y. Evaluation and comparison of errors on nursing notes created by online and offline speech recognition technology and handwritten: An interventional study. BMC Medical Informatics and Decision Making. 2022; 22(1):96. [DOI:10.1186/s12911-022-01835-4] [PMID]

[13] Hou SY, Wu YL, Chen KC, Chang TA, Hsu YM, Chuang SJ, et al. Code-switching automatic speech recognition for nursing record documentation: System development and evaluation. JMIR Nursing. 2022; 5(1):e37562. [DOI:10.2196/37562] [PMID] []

[14] Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. NEJM Catalyst. 2024; 5(3). [DOI:10.1056/CAT.23.0404]

[15] Wendt SJ, Dinh CT, Sutcliffe M, Jones K, Scanlan JM, Smitherman JS. Deploying ambient clinical intelligence to improve care: A research article assessing the impact of nuance DAX on documentation burden and burnout. Future Healthcare Journal. 2025; 12(3):100450. [DOI:10.1016/j.fhj.2025.100450] [PMID]

[16] Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell JT, Sheinberg M, et al. Clinician Experiences With Ambient Scribe Technology to Assist With Documentation Burden and Efficiency. JAMA Network Open. 2025; 8(2):e2460637. [DOI:10.1001/jamanetworkopen.2024.60637] [PMID]

[17] Liu TL, Hetherington TC, Dharod A, Carroll T, Bundy R, Nguyen H, et al. Does AI-Powered Clinical Documentation Enhance Clinician Efficiency? A Longitudinal Study. New England Journal of Medicine – Artificial Intelligence. 2024; 1(12). [DOI:10.1056/AIoa2400659]

[18] Haberle T, Cleveland C, Snow GL, Barber C, Stookey N, Thornock C, et al. The impact of nuance DAX ambient listening AI documentation: A cohort study. Journal of The American Medical Informatics Association. 2024; 31(4):975-9. [DOI:10.1093/jamia/ocae022] [PMID]

[19] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021; 372:n71. [DOI:10.1136/bmj.n71] [PMID]

[20] Lockwood C, Porrit K, Munn Z, Rittenmeyer L, Salmond S, Bjerrum M, et al. Systematic reviews of qualitative evidence. Jbi Manual for Evidence Synthesis. 2020; 46658:1-4. [Link]

[21] Ju H, Park M, Jeong H, Lee Y, Kim H, Seong M, et al. Generative AI-Based Nursing Diagnosis and Documentation Recommendation Using Virtual Patient Electronic Nursing Record Data. Healthcare Informatics Research. 2025; 31(2):156-65. [DOI:10.4258/hir.2025.31.2.156] [PMID]

[22] Levin S, Barnes S, Toerper M, Debraine A, DeAngelo A, Hamrock E, et al. Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. BMJ Innovations. 2021; 7(2). [DOI:10.1136/bmjinnov-2020-000420]

[23] King CR, Shambe A, Abraham J. Potential uses of AI for perioperative nursing handoffs: A qualitative study. JAMIA Open. 2023; 6(1):ooad015. [DOI:10.1093/jamiaopen/ooad015] [PMID]

[24] Chen CJ, Liao CT, Tung YC, Liu CF. Enhancing healthcare efficiency: Integrating ChatGPT in nursing documentation. Studies in Health Technology and Informatics. 2024; 316:851-2. [Link]

[25] Dos Santos FC, Johnson LG, Madandola OO, Priola KJB, Yao Y, Macieira TGR, et al. An example of leveraging AI for documentation: ChatGPT-generated nursing care plan for an older adult with lung cancer. Journal of the American Medical Informatics Association. 2024; 31(9):2089-96. [DOI:10.1093/jamia/ocae116] [PMID]

[26] Ozonoff A, Milliren CE, Fournier K, Welcher J, Landschaft A, Samnaliev M, et al. Electronic surveillance of patient safety events using natural language processing. Health Informatics Journal. 2022; 28(4):14604582221132429. [DOI:10.1177/14604582221132429] [PMID]

[27] Poon EG, Lemak CH, Rojas JC, Guptill J, Classen D. Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges. Journal of The American Medical Informatics Association. 2025; 32(7):1093-100. [DOI:10.1093/jamia/ocaf065] [PMID]

[28] von Wedel P, Hagist C, Liebe JD, Esdar M, Hübner U, Pross C. Effects of Hospital Digitization on Clinical Outcomes and Patient Satisfaction: Nationwide Multiple Regression Analysis Across German Hospitals. Journal of Medical Internet Research. 2022; 24(11):e40124. [DOI:10.2196/40124] [PMID]

[29] Cho I, Jin IS, Park H, Dykes PC. Clinical Impact of an Analytic Tool for Predicting the Fall Risk in Inpatients: Controlled Interrupted Time Series. JMIR Medical Informatics. 2021; 9(11):e26456. [DOI:10.2196/26456] [PMID]

[30] Baxter SL, Longhurst CA, Millen M, Sitapati AM, Tai-Seale M. Generative artificial intelligence responses to patient messages in the electronic health record: Early lessons learned. JAMIA Open. 2024; 7(2):ooae028. [DOI:10.1093/jamiaopen/ooae028] [PMID]

[31] Cho I, Cho J, Hong JH, Choe WS, Shin H. Utilizing standardized nursing terminologies in implementing an AI-powered fall-prevention tool to improve patient outcomes: A multihospital study. Journal of The American Medical Informatics Association. 2023; 30(11):1826-36. [DOI:10.1093/jamia/ocad145] [PMID]

[32] Balloch J, Sridharan S, Oldham G, Wray J, Gough P, Robinson R, et al. Use of an ambient artificial intelligence tool to improve quality of clinical documentation. Future Healthcare Journal. 2024; 11(3):100157. [DOI:10.1016/j.fhj.2024.100157] [PMID]

[33] Cho I, Kim M, Song MR, Dykes PC. Evaluation of an approach to clinical decision support for preventing inpatient falls: A pragmatic trial. JAMIA Open. 2023; 6(2):ooad019. [DOI:10.1093/jamiaopen/ooad019] [PMID]

[34] Johnson OA, McCrorie C, McInerney C, Mebrahtu TF, Granger J, Sheikh N, et al. Implementing an artificial intelligence command centre in the NHS: A mixed-methods study. Health and Social Care Delivery Research. 2024; 12(41):1-108. [DOI:10.3310/TATM3277]

[35] Yang L, Pang J, Zuo S, Xu J, Jin W, Zuo F, et al. Evolution of the "Internet Plus Health Care" Mode Enabled by Artificial Intelligence: Development and Application of an Outpatient Triage System. Journal of Medical Internet Research. 2024; 26:e51711. [DOI:10.2196/51711] [PMID]

[36] Danello SH, Maddox RR, Schaack GJ. Intravenous Infusion Safety Technology: Return on Investment. Hospital Pharmacy. 2009; 44(8):680-8. [DOI:10.1310/hpj4408-680]

[37] Johnson LG, Madandola OO, Dos Santos FC, Priola KJB, Yao Y, Macieira TGR, et al. Creating perinatal nursing care plans using ChatGPT: A pathway to improve nursing care plans and reduce documentation burden. The Journal of Perinatal & Neonatal Nursing. 2025; 39(1):10-9. [Link]

## Appendix

**Table 1.** PRISMA 2020 checklist

| Section and Topic | Item | Checklist item | Location Where Item is Reported |
|---|---|---|---|
| | | **Title** | |
| Title | 1 | Identify the report as a systematic review. | Page 1 |
| | | **Abstract** | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Page 1 |
| | | **Introduction** | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Page 2 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 2 |
| | | **Methods** | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Page 3 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Page 3 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Page 3 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Page 3 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 3 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Page 3 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Page 3 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Page 3 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Page 3 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 3 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Page 3 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Page 3 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 3 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Page 3 |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Page 3 |

| Section and Topic | Item | Checklist item | Location Where Item is Reported |
|---|---|---|---|
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Page 3 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | N/A |
| **Results** | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Page 4 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Page 4 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Page 4 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Page 8 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Page 5-7 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Page 8 |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Page 9 |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Page 9 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Page 9 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Page 9 |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | N/A |
| **Discussion** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Page 12 |
| | 23b | Discuss any limitations of the evidence included in the review. | Page 12 |
| | 23c | Discuss any limitations of the review processes used. | Page 12 |
| | 23d | Discuss implications of the results for practice, policy, and future research. | Page 12 |
| **Other Information** | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | Page 3 |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | N/A |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | N/A |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | N/A |
| Competing interests | 26 | Declare any competing interests of review authors. | N/A |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | N/A |

This Page Intentionally Left Blank