Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Human-Centered Artificial Intelligence in Mental Health Nursing: Clinical Augmentation, Therapeutic Relationships, and Responsible Governance

Kaianny Cardoso Pereira[1]
[1]Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: Katycard@icloud.com
ORCID: https://orcid.org/0009-0004-6842-8027

Abstract

Artificial intelligence (AI) is entering mental health nursing through prediction, natural language processing, digital phenotyping, conversational agents, and generative documentation. This integrative narrative review examines how these functions can support care while preserving therapeutic relationships, recovery, professional accountability, and patient rights. A targeted literature and regulatory update was completed on October 26, 2026; the revised bibliography comprises 35 sources, including six contemporary nursing or nursing-education reviews. Sources were synthesized thematically, without pooled effects or a claim of exhaustive systematic retrieval. Recent reviews establish substantial conceptual overlap and identify persistent limitations in external validity, person-centred outcomes, sustained engagement, and implementation evidence. Accordingly, this article contributes an author-developed nurse-led implementation and assurance framework that connects relational nursing theory, sociotechnical safety, evidence maturity, intended-use risk, and Brazilian governance. Operational requirements include calibration, prevalence-sensitive predictive values, prospective silent validation, measurable escalation, competent human review, abstention, version control, incident grading, and controlled revalidation. Documentation benefits must be assessed through total verification workload and protected therapeutic time. The risk matrix is a conceptual proposal, not a validated clinical instrument or binding legal classification. The framework has not been prospectively evaluated. AI adoption is defensible only when local evidence, service capacity, equity, privacy, and patient choice support the intended use; technical performance alone cannot establish clinical benefit.

Keywords: artificial intelligence; mental health nursing; therapeutic relationship; clinical decision support; governance

1. Introduction

Mental health conditions contribute substantially to disability, social exclusion, interrupted education and employment, and pressure on health systems. The World Health

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Organization estimates that nearly one billion people were living with a mental disorder in 2019, while access to effective, rights-based care remains uneven within and between countries. These gaps make scalable digital tools attractive, but scarcity also creates a danger: technology may be treated as a substitute for adequate staffing and community services rather than as one component of a stronger system (World Health Organization [WHO], 2022).

Mental health nurses occupy a distinctive position in this debate. They observe changes across shifts, administer and evaluate treatment, coordinate multidisciplinary plans, support families, promote recovery, and create the interpersonal conditions in which patients can speak safely about distress. Because much of this knowledge is narrative, contextual, and relational, it cannot be reduced to a laboratory value or a single prediction. AI may help nurses interpret information, yet its usefulness depends on whether it respects the epistemic value of proximity, dialogue, and longitudinal knowledge (Peplau, 1991; Yıldız, 2025).

Current AI systems encompass different technologies and should not be discussed as one intervention. Machine-learning models classify or predict outcomes from structured data; natural language processing extracts information from notes or speech; computer vision and sensors analyze behavioral or physiological signals; conversational agents deliver scripted or generated dialogue; and large language models produce flexible text from prompts. Each architecture has different evidence, failure modes, data requirements, and regulatory implications (Li et al., 2024; WHO, 2025).

Enthusiasm often rests on demonstrations rather than clinical outcomes. A model may distinguish cases in a retrospective dataset yet fail when documentation changes, a new population is served, or staff respond differently to its alerts. Likewise, a chatbot may be acceptable during a short trial without producing durable improvement or safe escalation. Evaluation must therefore move beyond technical accuracy to patient outcomes, workload, equity, therapeutic alliance, and unintended consequences (Vasey et al., 2022; WHO, 2021).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

This article asks which AI functions can be bounded, validated, governed, and integrated into mental health nursing without compromising recovery-oriented relational care. Its contribution is an author-developed implementation and assurance framework combining interpersonal nursing theory, sociotechnical safety, evidence maturity, intended-use risk, and Brazilian contextualization. It does not claim to be the first review of human-centred AI in this field. The framework converts established principles into reviewable implementation requirements and remains a conceptual proposal requiring empirical evaluation.

Recent integrative, scoping, umbrella, and bibliometric reviews already map this field and its ethical tensions (Bond et al., 2026; Milasan & Scott-Purdy, 2025; Odame et al., 2026; Özdemir, 2026; Si & Kwon, 2026). Their findings are compared explicitly in Section 5.1. The present synthesis uses that evidence to specify what institutions should demonstrate before an intended use is deployed, expanded, updated, or withdrawn.

2. Literature Review

Peplau's interpersonal relations framework places the nurse–patient relationship within the therapeutic process rather than treating communication as an accessory to technical care. Orientation, working, and resolution unfold through trust, clarification, participation, and professional use of self. AI can prepare information or identify patterns, but it does not bear moral responsibility, participate in embodied care, or understand a person's life in the human sense. The framework therefore positions technology as an instrument within nursing practice, not as a relational replacement (Peplau, 1991).

Recovery-oriented care adds a second criterion. Success is not limited to symptom reduction or prediction of service use; it includes hope, agency, meaningful roles, connection, and the right to influence one's own care. A technically accurate system may still be harmful if it labels a person through a permanent risk profile, narrows choices, or encourages coercive intervention. AI outcomes must therefore include autonomy and lived experience as well as operational efficiency (WHO, 2021, 2022).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

A sociotechnical perspective rejects the idea that an algorithm acts alone. Outcomes emerge from data definitions, interfaces, staffing, institutional rules, patient behavior, incentives, and local resources. A warning that is accurate but arrives at the wrong time can be ignored; a flawed summary can become more influential because it is embedded in the record; an efficiency tool can increase work if nurses must correct it. Safety analysis must map the entire work system (Sittig & Singh, 2010).

Care ethics is especially relevant because mental health services involve vulnerability, asymmetry of power, stigma, and at times constrained choices. Ethical quality cannot be achieved by a consent checkbox or a global accuracy score. It requires attention to who is observed, whose data are absent, who receives an alert, what response follows, and whether the patient can understand and contest an inference (WHO, 2021).

These perspectives produce a practical test: AI should strengthen nurses' capacity to recognize individual needs and respond proportionately. If it instead fragments attention, standardizes conversation, or shifts responsibility to an opaque vendor, it conflicts with the clinical and ethical purposes of nursing even when its benchmark performance appears impressive (Li et al., 2024; Yıldız, 2025).

3. Methodology

3.1 Review design and purpose

This study uses an integrative narrative review with thematic synthesis because evidence on artificial intelligence in mental health nursing spans clinical and observational research, technical validation, systematic and scoping reviews, nursing theory, ethical analysis, reporting guidance, law, and institutional governance. The review does not claim the exhaustive retrieval or pooled estimation of a registered systematic review. Its purpose is to compare levels of evidence, identify boundary conditions, and connect technical performance to clinical utility, therapeutic relationships, nursing workflow, patient rights, and organizational accountability (Grant & Booth, 2009; Whittemore & Knafl, 2005).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

3.2 Information sources and final search date

The submitted version reported a final search date of September 6, 2026 and described PubMed/MEDLINE, CINAHL-oriented discovery, Scopus-linked publisher platforms, Google Scholar, IEEE Xplore, and institutional repositories. Complete original database exports, database-specific strings, retrieval counts, and the exact CINAHL access platform were not retained in the materials available for revision. These historical searches cannot be independently reconstructed from the bibliography. They are therefore reported as an original discovery account, not a reproducible systematic database search. A targeted update was conducted on October 26, 2026 using public web discovery followed by inspection of PubMed/PMC records, publisher pages, Crossref metadata, and official Brazilian and WHO sources. Subscription CINAHL, Scopus, and IEEE searches were not rerun. Appendix A records the executed update queries and their scope.

3.3 Search strategy

The original search concepts combined artificial intelligence, machine learning, large language models, natural language processing, digital phenotyping, and conversational agents with mental health nursing, psychiatric care, documentation, education, therapeutic relationships, surveillance, bias, privacy, and governance. Their original database syntax and filters are unavailable. The revision used named-study verification and focused concept queries rather than claiming a retrospective database protocol. No search-engine date or language filter was applied to the update; selection prioritized English scientific literature and Portuguese primary regulatory material available by October 26, 2026. Older theoretical and methodological sources were retained deliberately. Exact executed update queries are recorded in Appendix A; they are not represented as CINAHL or Scopus strings.

3.4 Eligibility and exclusion criteria

Eligible sources made a direct contribution to at least one review domain through empirical evidence, systematic or scoping synthesis, validated reporting guidance, nursing

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

theory, legal authority, or transparent institutional recommendations. Priority was given to peer-reviewed research and authoritative primary documents. Sources were excluded when they consisted of promotional vendor claims, anonymous comparisons, unverifiable methods, duplicate propositions, Wikipedia, or demonstrations that did not distinguish technical feasibility from clinical benefit. Institutional documents were used for law, standards, and governance rather than treated as independent proof of effectiveness.

3.5 Selection process and corpus composition

The revised bibliography contains 35 unique cited sources: 25 analytical sources and 10 normative or contextual documents. The analytical group includes empirical and review literature, nursing theory, and methodological or clinical evaluation papers. The contextual group comprises two Brazilian laws, three WHO reports, the NIST framework, one University of Melbourne project webpage, two ANPD webpages, and one parliamentary legislative record. This explicit classification replaces the earlier 19/5 split. The update added 11 sources: five mental-health nursing or mental-health reviews, one nursing-education review, two prediction-evaluation papers, and three Brazilian institutional records. The six reviews are secondary evidence; their underlying studies were not reclassified as individually included primary studies. Titles and DOI identifiers were reconciled manually, with institutional records matched by issuing body and URL; overlapping studies within reviews were not counted as independent confirmations. Historical retrieval and exclusion totals are unavailable and are not reconstructed.

Bibliographic recency was recalculated over all 35 references. Twenty-three have dated publications or dated webpage updates in 2020–2026 (65.7%); the remaining 12 are earlier publications or undated material. Using the narrower calendar-year window 2022–2026, 20/35 references qualify (57.1%). Retrieval dates alone do not make an undated document recent. Foundational theory and prediction methods are retained for conceptual necessity rather than to increase recency.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

3.6 Data extraction and thematic coding

The revision organized sources by publication type, AI modality, clinical function, intended user, autonomy, setting, validation maturity, outcomes, and limitations. Comparisons emphasized assessment and monitoring; patient-facing support; documentation; workforce and education; therapeutic relationships; equity and privacy; and governance. Source access varied between full text, abstracts, bibliographic records, and official webpages; quantitative details were used only when available in the inspected source. No complete contemporaneous extraction dataset from the original search was available for audit. The updated synthesis and Table 2 therefore provide a transparent conceptual comparison rather than a claim of a newly completed systematic extraction exercise.

3.7 Critical appraisal and synthesis rules

Critical reading considered study design, representativeness, external or prospective validation, comparator quality, outcome relevance, calibration, subgroup uncertainty, and connection to actionable care. No standardized risk-of-bias instrument was applied independently across every source, and no numerical quality score is claimed. Reviews were interpreted as secondary syntheses with possible overlap in primary studies; bibliometrics mapped research activity rather than effectiveness. Guidance was used normatively, not as proof of clinical benefit. Where direct psychiatric-nursing evidence was unavailable, recommendations are identified as author-proposed safeguards or cautious extrapolations (Liu et al., 2020; Vasey et al., 2022).

3.8 Single-author procedure and limitations

The submitted review reports a single-author procedure. This revision retains that account; no second independent reviewer, inter-rater statistic, or retrospective Cohen's kappa is claimed. The update was assisted by AI-supported discovery and editing, with the access limitations described above. Transparent search dates and queries improve auditability but cannot recover missing original exports or eliminate selection and confirmation bias. The

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

narrative design, heterogeneous interventions, proprietary updates, English-language emphasis, limited prospective nursing-specific evidence, and sparse Brazilian empirical evaluation constrain the conclusions. No PRISMA-compliant systematic review, meta-analysis, or validated clinical classification is claimed.

4. Results

4.1 AI Modalities and Their Clinical Meaning

Predictive machine learning learns statistical associations between recorded variables and a target such as readmission, symptom trajectory, or use of restrictive intervention. Performance is commonly expressed with discrimination measures such as area under the receiver operating characteristic curve. Yet a high area under the curve does not communicate positive predictive value, calibration, consequences of false results, or whether an intervention changes outcomes. Nurses need measures tied to actual decisions and prevalence in the deployment setting (Liu et al., 2020).

For a specified outcome, threshold, and follow-up horizon, positive predictive value is PPV = TP / (TP + FP), and negative predictive value is NPV = TN / (TN + FN), where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives. These values depend on prevalence and should be estimated in the intended population. As a hypothetical illustration, 90% sensitivity and 90% specificity at 1% prevalence produce approximately 9 true alerts and 99 false alerts per 1,000 assessments, giving PPV ≈ 8.3%. This calculation illustrates the base-rate problem; it is not a result from an included trial.

Calibration assessment should include a flexible calibration plot, calibration intercept (ideal 0), and slope (ideal 1), with uncertainty and sufficient events. Where appropriate, the Brier score summarizes overall probabilistic error, but is not a pure calibration measure and should not replace calibration plots (Van Calster et al., 2019). These are recommended reporting requirements, not measurements newly estimated in this review.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Decision-curve analysis can examine net benefit at clinically justified threshold probabilities: $NB(p_t) = TP/N - (FP/N) \times p_t/(1 - p_t)$, where N is the number assessed and p_t expresses the threshold trade-off between missed events and unnecessary intervention. Candidate models should be compared with relevant assess-all, assess-none, and current-care strategies across a justified threshold range (Vickers & Elkin, 2006). A favorable curve does not demonstrate improved care unless the proposed intervention is available, proportionate, and effective.

Natural language processing can extract symptoms, social context, treatment response, and care processes from narrative records. This is relevant because nursing notes contain observations not represented in diagnosis codes. The same flexibility creates risks: documentation varies by institution, cultural expression is context-dependent, and historical notes may encode biased interpretations. Extracted variables should preserve provenance and allow review of the source text rather than appear as unquestionable facts (Li et al., 2024).

Digital phenotyping uses smartphones, wearables, or environmental sensors to infer behavior from mobility, sleep, communication, and physiological patterns. Continuous data may reveal change between appointments, but inferred meaning is rarely unique. Reduced mobility may reflect illness, work, weather, disability, or a deliberate privacy choice. Reliable systems must establish a personal baseline, tolerate missingness, and avoid converting ordinary variability into pathology (Torous et al., 2021).

Individual baseline assessment should precede interpretation of sensor-derived change. Missingness should be examined as potentially informative: device costs, disability, illness, privacy choices, and technical failures can change data availability and predicted risk. Evaluation should report missing-data patterns, subgroup coverage, and sensitivity to alternative missingness assumptions rather than treating absent observations as reassurance or automatically as deterioration (Torous et al., 2021).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Conversational agents range from rule-based scripts to generative systems. Structured agents can deliver psychoeducation and exercises consistently, while generative models offer more flexible dialogue but can produce unsupported or unsafe statements. Meta-analyses report promising short-term effects for some outcomes, yet interventions and populations are heterogeneous, follow-up is often limited, and attrition can be substantial (Gaffney et al., 2019; Li et al., 2023).

Evidence from a scripted, fixed-content agent does not establish the safety of a generative agent whose dialogue varies with prompts, model releases, and retrieved content. Studies should specify architecture, content constraints, autonomy, and update policy before benefits or harms are transferred between these categories (Li et al., 2023; WHO, 2025).

Generative AI can summarize notes, draft educational material, simulate encounters, and answer questions. Fluency is not evidence of truth: models may invent sources, omit decisive context, or change responses across runs. Retrieval from an approved knowledge base can improve grounding but cannot eliminate retrieval errors or faulty synthesis. High-stakes outputs therefore require traceable sources, constrained tasks, and professional verification (WHO, 2025).

4.2 Assessment, Documentation, and Decision Support

AI-assisted assessment may organize questionnaire responses, highlight missing items, and display longitudinal trends. Its appropriate role is to support a broader nursing assessment rather than turn screening into diagnosis. Standardized scores are sensitive to context, language, physical illness, and the circumstances of completion. Automated interpretation should state its limits and preserve space for the patient's account (Li et al., 2024; Wimbarti et al., 2024).

Analysis of electronic records can identify patterns that precede deterioration or unplanned service use. Such models may help prioritize review when caseloads are large. However, prediction is useful only when nurses have a proportionate action available, such as contact,

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

reassessment, or multidisciplinary review. Alerting without resources may merely document unmet need and increase moral distress (Torous et al., 2021; Vasey et al., 2022).

Natural language processing can reduce manual chart review by extracting temporal information from notes. A safe interface should display the relevant source passage, author, date, and confidence, allowing the nurse to determine whether an observation remains current. Generated summaries must distinguish reported speech, prior history, and present assessment; otherwise, uncertainty can harden into a misleading record (Sittig & Singh, 2010).

Ambient or generative documentation may reduce clerical time, but it can also introduce omissions, attribution errors, and polished statements that were never established during the encounter. The nurse remains responsible for the final record. Organizations should measure correction time, note quality, patient experience, and whether saved time actually returns to therapeutic care rather than being absorbed by higher throughput (WHO, 2025).

Draft generation with mandatory clinician approval must be distinguished from autonomous final documentation. The latter can propagate errors without an accountable verification step and belongs to a higher-risk intended use. Evaluation should record total task time, verification time, substantive corrections, harmful omissions, attribution errors, downstream corrections, and cognitive workload. Protected therapeutic time should be measured separately through observed direct-care time and patient-reported continuity; a faster note is not sufficient evidence of a relational benefit (WHO, 2025).

Clinical decision support should present options and reasons rather than a single opaque directive. For medication monitoring, care planning, or referral, the interface should show relevant patient data, applicable guidance, uncertainty, and contraindicating information. Nurses need a straightforward means to disagree and record why. Override rates should be studied for system improvement, not used to discipline clinicians (WHO, 2021).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

4.3 Patient-Facing Conversational and Digital Support

Conversational agents can offer psychoeducation, structured exercises, symptom check-ins, and navigation outside clinic hours. Their scalability may reduce practical barriers for people who prefer asynchronous contact or live far from services. Evidence from systematic reviews suggests potential benefit, particularly for selected short-term outcomes, but the effect is not uniform and should not be generalized to every condition or level of need (Gaffney et al., 2019; Li et al., 2023).

Engagement is a central limitation. Initial curiosity can produce high early use, while participation declines when dialogue becomes repetitive, burdensome, or poorly matched to need. Attrition is not merely a statistical inconvenience: it may concentrate benefits among digitally confident users and obscure the experience of those who disengage. Trials should report retention, reasons for withdrawal, accessibility, and longer follow-up (Jabir et al., 2024).

Claims of empathy require caution. A system can produce language that users perceive as supportive, but simulated empathy is not equivalent to accountable therapeutic presence. Anthropomorphic design may encourage disclosure beyond what users understand about data storage or may create unrealistic expectations of monitoring. Interfaces should clearly identify the system, its limits, whether a human reviews messages, and how to obtain professional help (Sanjeewa et al., 2024; WHO, 2025).

A stepped-care model offers a defensible boundary. Low-intensity, structured support may complement prevention and follow-up, while complex, worsening, or urgent presentations require timely human assessment. Escalation must be operational rather than rhetorical: the service needs defined thresholds, available staff, response times, documentation, and testing of missed and excessive escalations (WHO, 2025).

Escalation must be tested as a service process. Prespecified measures should include sensitivity for situations requiring human assessment, false-escalation and missed-escalation

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

rates with explicit denominators, time from trigger to clinical contact, and actual availability of qualified responders. Rates should include confidence intervals and subgroup results. Services should define urgent fallback routes, downtime coverage, and stop criteria when capacity is exceeded. These are proposed assurance measures; this review does not claim validated universal escalation thresholds (Odame et al., 2026; WHO, 2025).

Mental health nurses can govern patient-facing tools by reviewing content, defining language, monitoring conversations within consented workflows, and connecting digital observations to individualized plans. Their involvement helps prevent a generic wellness product from being presented as treatment and ensures that technology complements continuity rather than creating a disconnected parallel service (Nashwan et al., 2023).

4.4 Therapeutic Relationships, Trust, and Humanization

The therapeutic relationship depends on attention, consistency, confidentiality, and the patient's experience of being recognized as a person. AI may support these conditions when it reduces repetitive administration or helps a nurse remember preferences and prior concerns. It undermines them when screens dominate encounters, generated scripts standardize responses, or patients feel continuously evaluated by hidden models (Peplau, 1991; Yıldız, 2025).

Trust should not be treated as user acceptance alone. Appropriate trust is calibrated: users understand what a system does well, where it fails, and who is accountable. Excessive trust leads to automation bias, while blanket distrust wastes useful support. Training and interface design should communicate uncertainty and make verification easy instead of relying on disclaimers that users rarely read (Goddard et al., 2012; WHO, 2021).

AI can reshape professional attention even without replacing a conversation. Risk scores may anchor the initial impression, documentation prompts may privilege symptoms that are easy to code, and dashboards may compete with subtle interpersonal cues. Reflective practice

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

should examine how technology changes what nurses notice and which narratives become institutionally visible (Sittig & Singh, 2010).

Patients should be able to choose a human route when feasible and should know when AI contributes to assessment, communication, or documentation. Refusal should not reduce access or be interpreted automatically as nonadherence. These safeguards are particularly important where services involve involuntary elements or where patients reasonably fear stigma and secondary use of sensitive data (WHO, 2021).

The most meaningful efficiency outcome is protected therapeutic time. Procurement claims often count messages generated or minutes per note, but mental health nursing should test whether nurses spend more time in direct, purposeful contact and whether patients experience greater continuity. Productivity without relational benefit is an incomplete measure of value (Li et al., 2024).

Longitudinal evaluation should distinguish engagement, satisfaction, and therapeutic alliance. Report retention by follow-up interval, reasons for disengagement, accessibility barriers, and outcomes among nonusers where feasible. Refusal should not be recoded as failure to comply or used to diminish access. A person-centred assessment should include dignity, trust, shared decision-making, and the possibility that a patient reasonably prefers human contact (Bond et al., 2026; Jabir et al., 2024).

4.5 Safety, Risk Prediction, and Restrictive Practices

Predictive systems are sometimes proposed to identify acute deterioration, aggression, readmission, or need for intensive review. These outcomes are infrequent and context-dependent, so apparently strong discrimination can coexist with many false alerts. Positive predictive value, calibration, subgroup performance, workload, and the consequences of action must be reported alongside headline accuracy (Liu et al., 2020).

A risk score should trigger supportive assessment rather than function as proof about future behavior. When historical labels and institutional responses are used as targets, the

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

model may reproduce earlier patterns of surveillance or coercion. The ethical objective is prevention through timely, least-restrictive care, not earlier restriction based solely on probability (WHO, 2021, 2022).

Sensor-assisted observation may reduce interruptions and help staff notice physiological or behavioral change, but continuous monitoring can make inpatient care feel carceral. Necessity, proportionality, retention limits, access controls, independent oversight, and patient involvement should be established before deployment. Safety cannot be invoked as an unlimited justification for surveillance (University of Melbourne, n.d.).

Proportionality review should document necessity, the least intrusive feasible alternative, likely safety benefit, and privacy, autonomy, and surveillance burdens. These dimensions cannot be reduced to accuracy or a single numerical subtraction. Involuntary treatment creates particular power asymmetries; consent does not remove the need to justify monitoring or prevent secondary uses. The Melbourne source is an institutional project webpage, not a peer-reviewed effectiveness evaluation (University of Melbourne, n.d.).

False negatives can delay assessment, whereas false positives can increase anxiety, unnecessary observation, and staff desensitization. Thresholds therefore represent value judgments as well as statistical choices. Patients, nurses, clinicians, ethics specialists, and rights advocates should participate in defining acceptable trade-offs, and decisions should be revisited after real-world monitoring (WHO, 2021).

For high-risk situations, AI must never be the sole assessor or decision maker. Systems require immediate human review, redundant communication routes, downtime procedures, and clear accountability. The evidence base should demonstrate not only retrospective prediction but prospective improvement in a complete care pathway (Vasey et al., 2022; WHO, 2025).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

4.6 Equity, Bias, Privacy, and Data Governance

Mental health datasets are not neutral samples of human experience. They reflect who reaches care, how professionals document behavior, which communities are surveilled, and what services are available. A model trained on service use may learn access patterns rather than need. External validation and subgroup analysis are essential, but quantitative parity alone cannot correct inequitable target definitions (Obermeyer et al., 2019; WHO, 2021).

Subgroup evaluation should report sample size, event count, uncertainty intervals, missingness, and calibration as well as discrimination and error rates. Small groups may yield imprecise or unstable estimates; absence of a statistically detectable difference is not evidence of fairness. The target itself must be examined: predicting past restriction, access, or service use may reproduce institutional behavior rather than clinical need (Obermeyer et al., 2019).

Language models and speech systems may perform differently across dialects, literacy levels, disability, and cultural expressions of distress. Translation can improve access while also erasing nuance. Evaluation should use locally produced data and patient-defined measures rather than assume that English-language performance transfers to Brazilian Portuguese or Indigenous and migrant communities (WHO, 2021).

Mental health information is exceptionally sensitive because disclosure can affect relationships, employment, insurance, and social standing. Data minimization requires collecting only what the intended use needs. Institutions should define retention, access, encryption, vendor use, model training, deletion, and breach response, and should prevent function creep into employment screening or unrelated surveillance (Brazil, 2018; WHO, 2021).

Consent must be understandable, specific, and revisitable. Yet consent cannot legitimize an unsafe or unnecessary system, particularly where power is unequal. Patients need explanations of what data are used, what inference is produced, who sees it, how it affects

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

care, and how to challenge an error. Alternative care should remain available whenever practicable (WHO, 2021).

Algorithmic accountability requires more than identifying a vendor. Health services remain responsible for clinical integration, monitoring, and redress. Procurement contracts should guarantee audit access, change notification, data portability, security duties, and incident cooperation. Patients and nurses need a visible route to report harm and obtain correction (National Institute of Standards and Technology [NIST], 2023).

4.7 Nursing Workforce, Well-Being, and Professional Boundaries

AI is frequently presented as a response to workforce shortages and burnout. Scheduling tools, documentation support, and workload forecasting may help when they address a defined operational problem. They can also intensify work by increasing surveillance, adding alerts, or converting predicted capacity into higher targets. Workforce benefit must be measured directly rather than inferred from automation (Nashwan et al., 2023).

Predictive analytics about staff absence, turnover, or mental health create serious boundary concerns. Employment decisions based on inferred vulnerability may stigmatize workers and discourage help-seeking. Organizations should improve working conditions and provide voluntary, confidential support rather than use individual risk scoring for hiring, promotion, scheduling, or discipline (WHO, 2021).

Generative systems may reduce documentation time, but correction burden is uneven. Novices may fail to detect plausible errors, and experienced nurses may spend considerable effort verifying text. Studies should measure total work, cognitive load, interruptions, and downstream corrections. A system that moves work from typing to vigilance is not automatically efficient (Goddard et al., 2012).

Professional boundaries require clear assignment of responsibility. AI cannot hold licensure, explain a decision to a regulatory body, or repair a damaged therapeutic relationship. Nurses should not be positioned as the final safety layer for products they cannot

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

inspect or influence. Governance must provide authority, training, staffing, and the ability to suspend unsafe use (WHO, 2025).

The appropriate workforce goal is augmentation: eliminating avoidable clerical friction, improving information retrieval, and supporting coordination while preserving skilled roles. Technology investment should be compared with staffing, supervision, workflow redesign, and simpler digital tools. Sometimes the safer and more effective intervention is organizational rather than algorithmic (Sittig & Singh, 2010).

Contemporary nursing-focused syntheses identify professional readiness and changing roles as implementation concerns, but do not establish that AI reduces staffing needs or burnout (Milasan & Scott-Purdy, 2025; Si & Kwon, 2026). Bond et al. (2026) situate these changes within relational care and organizational power. Workforce evaluation should therefore measure workload, supervision demands, autonomy, moral distress, and protected patient-contact time rather than infer gains from the number of automated tasks. Özdemir (2026) maps growing research attention; bibliometric prominence is not evidence of workforce benefit.

4.8 Education, Simulation, and AI Literacy

AI can generate case variations, simulate patient interviews, and provide formative feedback. In mental health nursing, simulation can expose learners to different communication challenges without presenting a generated response as the single correct script. Faculty should design activities in which students identify omissions, bias, and unsafe assumptions, then compare output with evidence and therapeutic principles (Nashwan et al., 2023).

Jiang et al. (2026) synthesized 19 nursing-simulation studies involving 1,253 participants. Findings suggested benefits in knowledge and selected confidence or communication outcomes, while complex skill outcomes were inconsistent and long-term transfer remained uncertain. This evidence concerns nursing education broadly and cannot establish effectiveness in psychiatric nursing practice. Psychiatric simulation should consequently

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

retain educator-led debriefing, standardized-patient or clinical comparators, independent assessment, and follow-up of retained communication and clinical reasoning skills. Automated feedback should be checked for stereotypes, unsafe guidance, and omission of nonverbal or contextual information.

Virtual patients can support repeated practice, but educational realism must not rely on stereotypes. Training data and prompts may reduce diagnoses to exaggerated behaviors or culturally narrow speech. Cases should be reviewed by educators, clinicians, and people with lived experience, and should portray strengths, context, and recovery as well as symptoms (WHO, 2021).

AI literacy includes more than prompting. Nurses need to understand intended use, data provenance, prevalence, calibration, false results, distribution shift, automation bias, privacy, security, and incident reporting. They must also know when a deterministic rule or ordinary clinical consultation is preferable to generative output (NIST, 2023; WHO, 2025).

Assessment policies should distinguish permitted assistance from independent competence. Students must demonstrate their own ability to communicate, reason clinically, evaluate evidence, and document accurately. AI can be used as an object of critique and a supervised tool, but undisclosed generation can weaken learning and make authorship unclear (WHO, 2025).

Continuing education is necessary because systems change after deployment. Training should be tied to the local configuration and include representative failure cases, escalation, downtime, patient explanation, and correction procedures. Competence should be reassessed after major updates rather than presumed permanent (Vasey et al., 2022).

4.9 Community Care, Telehealth, and the Brazilian Context

Brazil's community-based mental health system makes continuity, territorial care, and social participation central concerns. AI applications in primary care, Psychosocial Care Centers, and telehealth should reinforce referral coordination and longitudinal nursing work

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

rather than create isolated digital front doors. Local evidence remains limited, so implementation should begin with carefully evaluated pilots (Brazil, 2001; WHO, 2022).

The Unified Health System's commitments to universality, comprehensiveness, and equity provide criteria for digital adoption. A tool that reaches connected urban users while excluding rural, low-income, disabled, or low-literacy populations can widen inequality despite increasing total contacts. Programs need non-digital routes, accessibility, connectivity support, and monitoring of who is not represented (WHO, 2021).

The Brazilian General Data Protection Law establishes principles including purpose, adequacy, necessity, transparency, security, prevention, and nondiscrimination. Compliance requires more than a privacy notice: services must map data flows, legal bases, operators, retention, international transfers, and rights. Sensitive health data should not be sent casually to public generative systems (Brazil, 2018).

As checked on October 26, 2026, the official Chamber of Deputies record lists Bill 2338/2023 as awaiting the rapporteur's opinion in a special committee. It is treated here as proposed legislation, not an enacted comprehensive AI statute (Câmara dos Deputados, 2026). Existing duties under the LGPD and applicable health and professional rules continue to require a documented legal analysis of the particular deployment. Article 20 concerns review of decisions based solely on automated personal-data processing; this paper's stronger requirement for clinical human review is also an ethical and safety proposal, not a claim that every provision of the proposed AI bill is already binding (Brazil, 2018).

The ANPD regulatory agenda, updated on October 20, 2026, identifies ongoing work on AI, legal bases, data-subject rights, governance, and interpretation of LGPD Article 20. Its sandbox page, updated on October 16, 2026, lists a testing methodology and three partial monitoring reports (Agência Nacional de Proteção de Dados [ANPD], 2026a, 2026b). Regulatory experimentation should not be interpreted as clinical endorsement. Nursing services should document data flows, minimization, purpose and retention limits, access

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

permissions, portability, contestation and correction procedures, and the legal basis applicable to sensitive health information. Portuguese-language performance and local care capacity remain separate validation obligations.

Local validation must address Brazilian Portuguese, regional vocabulary, service organization, professional scope, medication availability, and referral pathways. Imported models may generate recommendations that assume insurance structures or emergency resources that do not exist locally. Knowledge bases and escalation messages should be jurisdiction-specific and maintained by accountable institutions (WHO, 2025).

Nurses, service users, families, community representatives, informatics teams, legal specialists, and managers should jointly design pilots. Participation is not symbolic consultation; stakeholders should influence the problem definition, acceptable data, evaluation outcomes, and stop criteria. Public-sector procurement should require transparency and avoid irreversible vendor dependence (NIST, 2023; WHO, 2021).

4.10 A Nurse-Led Implementation and Assurance Framework

Implementation should begin with a clinical problem, not a product. Teams should specify the user, setting, input, output, decision affected, expected benefit, foreseeable harm, and prohibited uses. Narrowly defined functions—such as drafting a reviewed follow-up summary—are easier to test and govern than broad promises to manage mental health (NIST, 2023).

Participatory workflow mapping should identify where information originates, who verifies it, what action follows, and what occurs during failure. The map must include patients and frontline nurses because formal procedures often differ from actual work. If an alert has no resourced response, or a summary cannot be corrected, the design is incomplete (Sittig & Singh, 2010).

Validation should use representative local cases separated from development data. Outcomes should include accuracy, harmful omission, calibration, false-alert burden,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

subgroup performance, source traceability, reproducibility, time, workload, patient experience, and effect on therapeutic contact. Thresholds and stop criteria should be declared before testing (Liu et al., 2020; Vasey et al., 2022).

Progression should follow explicit gates: retrospective validation, external validation, prospective silent validation, controlled impact evaluation, and monitored deployment. Silent validation means collecting prospective outputs without allowing them to guide care; clinical impact testing exposes the care pathway to the intervention and requires its own safeguards. At each gate, the nurse-led team should record the evidence, residual uncertainties, service capacity, and authorized scope. No gate is passed merely because a model meets an AUROC target (Vasey et al., 2022).

Human oversight must be meaningful. Reviewers need sufficient time, appropriate expertise, access to source data, and authority to reject output. High-risk or uncertain cases require escalation, while failed retrieval or missing information should produce abstention rather than confident completion. Interfaces should make disagreement easy and record correction for learning (WHO, 2025).

Operational oversight requires named competent reviewers, adequate protected time, access to source data and uncertainty, authority to reject recommendations, a documented override route, and resourced escalation. High-risk uncertainty should permit abstention, such as requesting human review when information is insufficient. Monitor both the frequency and appropriateness of abstention, including missed unsafe cases and unequal refusal across groups. A nominal human signature without time or authority does not meet this requirement (WHO, 2025).

Deployment should monitor drift, complaints, overrides, correction burden, incidents, disparities, alert volume, and therapeutic contact. Each output should be traceable to model, prompt, retrieval-corpus and index, and interface versions, together with its date and relevant configuration. Logs themselves require minimization and access controls. Material changes

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

include altered model weights or provider release, system prompt, retrieval content or indexing, clinical threshold, population, workflow, or interface behavior that may affect decisions. Such changes require risk review, regression testing, and revalidation proportionate to their impact before expansion; major changes require renewed governance approval. Maintain an approved rollback version and downtime care process (NIST, 2023; WHO, 2025).

Incident management should distinguish near misses, minor workflow errors, clinically relevant errors, and serious patient-safety events, recording actual and potential harm, frequency, and affected groups. These are proposed local learning categories, not statutory reporting definitions. A named nursing safety lead should triage reports; serious events or predefined excess harm should trigger immediate containment and qualified review, with privacy/legal notification handled under applicable requirements. Corrective action, patient redress, and criteria for restart should be documented (NIST, 2023).

Table 1 summarizes the author-developed intended-use categories and their minimum assurance requirements; it should be interpreted together with the operational safeguards described above.

Table 1 Author developed conceptual governance matrix for mental health nursing AI

Use category	Illustrative intended use	Minimum assurance
Lower risk	Reformat approved educational content; administrative drafting.	Constrained input; nurse review; privacy controls; version trace.
Moderate risk	Extract trends; structured check-ins, scheduling, or psychoeducation.	Local validation; provenance; accessibility; human route; error monitoring.
High risk	Predictions changing urgency, observation, medication review, or referral.	Calibration and subgroup uncertainty; prospective pathway evidence; immediate qualified review; tested escalation.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Use category	Illustrative intended use	Minimum assurance
Generally unacceptable without exceptional evidence	Autonomous diagnosis; AI-only involuntary-care decisions; unsupervised high-risk triage; employment profiling.	Prohibition by default in this conceptual proposal. AI alone must not determine involuntary care or unsupervised high-risk triage; exceptional review does not replace evidence or applicable law.

Note. Author-developed conceptual matrix derived from the narrative synthesis. Categories are provisional and depend on intended use, autonomy, setting, and downstream consequences. They are neither externally validated clinical classes nor binding regulatory classifications. No autonomous diagnosis or restriction of liberty is justified by this table.

4.11 Regulation, Economic Evaluation, and Evidence Maturity

Regulatory classification depends on intended use rather than the general label "AI." A tool that schedules visits or reformats approved text differs from software that recommends a clinical action. Developers and institutions should determine whether the application functions as a medical device, clinical decision-support component, documentation aid, or administrative system, and should document the reasoning. Changing the model, data source, or workflow may also change the risk profile and trigger renewed review (WHO, 2025).

Generative systems complicate conventional approval because outputs are probabilistic and commercial models may change frequently. A static premarket test cannot establish permanent safety when model weights, prompts, retrieval sources, or guardrails evolve. Assurance therefore requires a life-cycle approach that connects initial validation, controlled change, post-deployment surveillance, incident investigation, and the ability to roll back a release (NIST, 2023; WHO, 2025).

Evidence maturity should be described in stages. Technical feasibility shows that a model can perform a task under controlled conditions; external validation asks whether performance transfers; prospective silent testing examines the local workflow without affecting care; an impact study tests clinical and organizational outcomes; and surveillance assesses sustained

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

performance. Skipping these stages converts uncertainty into exposure for patients and staff (Vasey et al., 2022).

Benchmark comparisons often report an average score without clarifying whether errors are clinically important. In mental health nursing, a harmless wording difference and a missed need for human assessment cannot be counted equally. Evaluation panels should predefine critical errors, use multiple qualified reviewers, report disagreement, and include people with lived experience when judging acceptability, dignity, and comprehensibility (Liu et al., 2020; WHO, 2021).

Economic evaluation must consider the entire care pathway. Acquisition price is only one component; integration, cybersecurity, data preparation, training, verification time, model updates, patient support, accessibility, monitoring, and incident response all consume resources. Benefits should include health and relational outcomes as well as labor time. A tool that appears inexpensive may be poor value if it generates avoidable reviews or fragments continuity (Sittig & Singh, 2010).

Total cost of ownership should be assessed over an explicitly stated local time horizon, including licensing, implementation, integration, cybersecurity, local validation, training, verification, monitoring, incident response, vendor transition, and accessibility. Report whose budget and time are counted, uncertainty in utilization, and sensitivity to changing subscription and staffing assumptions. Compare AI with additional staffing, clinical supervision, workflow redesign, rule-based tools, and improved electronic-record functionality. Cost estimates are implementation requirements proposed here; no economic effectiveness estimate is derived from this narrative corpus.

Opportunity cost is equally important. Funds devoted to AI are unavailable for nurses, supervision, community outreach, interpreters, or established digital infrastructure. Decision makers should compare the proposed system with feasible alternatives, including workflow

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

simplification and rule-based tools. AI is justified when it adds measurable value to a defined problem, not merely because it represents innovation (NIST, 2023).

Distributional analysis asks who receives benefits and who carries burdens. Remote monitoring may help patients with reliable devices while transferring data costs and privacy risks to users. Automated triage may accelerate care for typical presentations while delaying people whose language or circumstances are poorly represented. Economic and clinical evaluations should therefore report outcomes across relevant groups and incorporate accessibility costs (WHO, 2021).

Distributional analysis should examine whether apparent average efficiency transfers burden to low-literacy users, minority-language groups, rural communities, people with disabilities, or patients who decline AI. Include device and connectivity costs, human alternative pathways, unpaid support work, and differences in waiting time. Compare the allocation of benefits with the allocation of privacy and surveillance risks (WHO, 2021).

Procurement should require evidence appropriate to the proposed use, not demonstrations selected by the vendor. Contracts should identify system versions, validated settings, service levels, data processing, subcontractors, update procedures, audit rights, exit arrangements, and liability. Institutions should retain their records and be able to continue care when a vendor changes terms or withdraws a product (NIST, 2023).

Independent oversight is particularly important when a system influences observation, urgency, or access. Multidisciplinary governance should include mental health nurses, clinicians, informaticians, privacy and security specialists, management, and service-user representation. Conflicts of interest should be disclosed, minutes retained, and decisions linked to evidence and explicit risk acceptance (WHO, 2021, 2025).

Mature adoption is therefore deliberate rather than slow for its own sake. It supports small, reversible pilots; measures outcomes before expansion; communicates uncertainty; and stops when benefits do not outweigh burden. This approach protects innovation by making

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

successful uses credible and prevents weak applications from consuming trust that the broader field will need (Vasey et al., 2022).

5. Discussion

The evidence supports neither technological determinism nor categorical rejection. AI can contribute to mental health nursing when the task is bounded, sources and data are appropriate, and professionals can verify and act. It is least defensible when a fluent interface conceals uncertainty, replaces access to care, or directs restrictive decisions without contextual assessment (Li et al., 2024; WHO, 2025).

A recurring weakness is the gap between predictive performance and clinical value. Retrospective studies answer whether a model can recover patterns in recorded data, not whether it improves a patient's life. Prospective evaluation must test the entire pathway, including staff response, available resources, unintended behavior, and outcomes important to service users (Liu et al., 2020; Vasey et al., 2022).

Relational care is not the residual work left after automation. It is a mode of clinical knowledge through which meaning, preference, change, and trust become visible. AI should be assessed partly by whether it protects this capacity. Time savings that lead only to more throughput may not improve quality, while a modest tool that strengthens continuity may have greater value (Peplau, 1991).

The distribution of benefit and risk requires explicit attention. People poorly represented in data may receive less accurate predictions, while those already subject to surveillance may receive more restrictive responses. Equity evaluation should investigate data generation, target definitions, intervention availability, and patient experience, rather than relying solely on equal average error rates (Obermeyer et al., 2019; WHO, 2021).

Mental health nurses should have governance authority because they understand continuous care, documentation, medication effects, interpersonal dynamics, and

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

implementation realities. Their role is not simply to approve output after procurement; it includes defining the problem, selecting outcomes, creating test cases, setting escalation, educating users, and deciding when technology should not be used (Yıldız, 2025).

5.1 Positioning against contemporary reviews

Milasan and Scott-Purdy (2025) synthesize 78 studies and map applications relevant to mental health nursing; their coverage makes a general applications catalogue insufficient as this article's novelty. Bond et al. (2026) examine person-centred alignment, identifying unresolved questions concerning trust, surveillance, and therapeutic relationships. This article builds on those concerns by specifying intended-use gates and accountable review requirements rather than claiming that relational concerns are newly recognized.

Odame et al. (2026) integrate 27 systematic reviews and describe a more developed evidence base for screening and prediction than for sustained therapeutic intervention, with recurring external-validity and equity concerns. Their synthesis supports caution but does not establish the effectiveness of the framework proposed here. Review-level convergence may also reflect shared primary studies. Özdemir (2026) maps 141 publications and emerging themes; publication volume indicates research activity, not clinical readiness.

Si and Kwon (2026) synthesize 42 sources across psychosocial support, risk management, therapeutic relationships, ethics and governance, and nursing readiness. These domains overlap substantially with the present review. The incremental contribution offered here is the explicit integration of relational nursing theory, sociotechnical evaluation, evidence maturity, operational safeguards, and Brazilian legal and service conditions into a nurse-led assurance sequence. Neither the sequence nor its matrix has been validated, and it should be tested against usual governance arrangements in prospective implementation research.

Table 2 compares the scope and interpretive role of the recent reviews. It is a source-level orientation table, not a formal AMSTAR assessment or a pooled evidence hierarchy. Nursing-

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

education evidence is included with an explicit transferability limit; it does not supply missing psychiatric-care effectiveness data.

Table 2 Contemporary reviews and their role in the revised synthesis

Source	Design and coverage	Role in this article	Interpretive boundary
Milasan and Scott-Purdy (2025)	Integrative review; 78 studies.	Maps nursing-relevant applications.	Application breadth does not establish prospective local benefit.
Bond et al. (2026)	Scoping review of person-centred alignment.	Addresses trust, power, surveillance, and relational outcomes.	Supports relational criteria; does not validate the present framework.
Odame et al. (2026)	Umbrella review; 27 systematic reviews.	Compares applications and recurring implementation constraints.	Secondary reviews may overlap; prediction evidence cannot establish sustained therapy benefit.
Özdemir (2026)	Bibliometric analysis; 141 publications.	Maps growth, thematic structure, and research priorities.	Publication patterns do not measure intervention effectiveness.
Si and Kwon (2026)	Scoping review; 42 sources.	Integrates care, risk, relationships, governance, and readiness.	Broad conceptual overlap requires restrained novelty claims.
Jiang et al. (2026)	Mixed methods review; 19 studies; 1,253 participants.	Adds contemporary nursing simulation evidence.	General nursing education; clinical transfer and psychiatric applicability remain uncertain.

5.2 Research Priorities

Future research should prioritize multicenter prospective studies, diverse languages and settings, and outcomes beyond short-term symptom scores. Trials should measure access, continuity, therapeutic alliance, adverse events, staff workload, cost, equity, and sustained

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

use. Model and prompt versions must be reported so results can be reproduced and interpreted after products change (Vasey et al., 2022).

Comparative studies should test AI-assisted nursing against current best practice and simpler alternatives, not against unaided memory. Component studies can separate the effects of structured screening, retrieval, generation, interface, and human review. Economic analysis should include integration, verification, training, privacy controls, vendor changes, and incident response (Sittig & Singh, 2010).

Participatory research should treat service users and nurses as co-designers with decision power. Qualitative evidence can reveal changes in trust, dignity, disclosure, and work that accuracy metrics miss. Research should also examine nonuse and refusal, since people who reject a tool may experience different access or may identify harms invisible to enthusiastic users (WHO, 2021).

5.3 Limitations

This is a selective integrative narrative review, not an exhaustive systematic search. The expanded 35-source corpus remains modest for the breadth of modalities and settings. Original retrieval counts, complete database exports, the exact CINAHL platform, and a reproducible historical screening log are unavailable. The targeted update improves contemporary coverage but does not remedy those historical gaps. Single-author selection and synthesis, AI-assisted discovery, and absence of independent formal risk-of-bias appraisal increase the possibility of selection, extraction, and interpretation bias.

Interventions, populations, endpoints, and degrees of autonomy are heterogeneous. Prospective psychiatric-nursing impact evidence and Brazilian empirical evidence remain limited, and some arguments draw on broader digital mental health or general nursing education. Recent reviews overlap conceptually and may share primary studies, so their inclusion must not be interpreted as independent replication. Proprietary systems can change faster than publication, and reviewed versions may differ from deployed configurations. The

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

author-developed matrix and assurance framework lack prospective validation and cannot establish effectiveness, legal compliance, or universal risk thresholds. Economic and monitoring recommendations are proposals for evaluation rather than findings from new clinical or cost data.

6. Conclusion

AI can help mental health nurses organize information, detect selected patterns, prepare documentation, deliver structured education, and coordinate follow-up. These functions may increase reach and return time to care, but only when systems are integrated with accountable services and evaluated in the populations that will use them (Li et al., 2024; WHO, 2025).

The principal hazards are not limited to incorrect output. They include biased targets, intrusive observation, false certainty, alert burden, privacy loss, automation bias, unequal access, and displacement of relational work. Because these harms emerge from workflows and institutions as well as algorithms, safety must be governed sociotechnically (NIST, 2023; Sittig & Singh, 2010).

The proposed contribution is a nurse-led assurance sequence linking bounded intended use, representative evidence, relational compatibility, competent human oversight, actionable escalation, equity and privacy safeguards, continuous monitoring, and controlled change. Its conceptual risk matrix is intended to structure local deliberation, not replace legal classification or clinical judgment. Prospective evaluation must determine whether using this framework improves safety, patient experience, workload, and protected therapeutic time beyond existing governance.

The appropriate future is not automated psychiatric nursing. It is human-centered, AI-augmented care in which technology expands the capacity to notice and communicate while professional judgment, therapeutic presence, and patient agency remain decisive. Innovation earns legitimacy only when it strengthens rights-based, recovery-oriented, and equitable mental health care.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Ethical Approval

Ethical approval was not required because this study analyzed publicly available literature and institutional documents and involved no human participants, identifiable personal data, animals, or intervention.

Funding

No external funding was received for the preparation of this article.

Conflicts of Interest

The author declares no conflicts of interest.

AI Assistance and Author Responsibility

Generative AI assisted literature discovery, translation, organization, and revision of this manuscript. Its involvement does not constitute independent peer review, specialist clinical or legal approval, or a second screening assessment. The human author retains responsibility for the submitted text, references, interpretations, and declarations. No original experimental data or inter-rater statistics were generated for this revision.

Author Responsibility Statement

The sole author is responsible for the conception, interpretation, content, reference verification, and final approval of this manuscript and for ensuring that the version submitted to the journal accurately represents the work performed. AI assistance does not transfer authorship or accountability to the system. The methodological access limitations and absence of independent specialist sign-off are disclosed in the manuscript.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

References

Agência Nacional de Proteção de Dados. (2026a, October 20). Agenda regulatória [Regulatory agenda; official webpage]. https://www.gov.br/anpd/pt-br/assuntos/regulacao/agenda-regulatoria-1

Agência Nacional de Proteção de Dados. (2026b, October 16). Sandbox [Official regulatory project webpage]. https://www.gov.br/anpd/pt-br/assuntos/projetos-acoes-iniciativas/sandbox

Bond, C., Watson, A. L., Smith, G. D., Aveyard, H., & Jackson, D. (2026). How does artificial intelligence align with person-centred principles in mental health nursing? a scoping review. International Journal of Mental Health Nursing, 35(3), e70282. https://doi.org/10.1111/inm.70282

Brazil. (2001). Lei No. 10.216, de 6 de abril de 2001. Presidência da República. https://www.planalto.gov.br/ccivil_03/leis/leis_2001/l10216.htm

Brazil. (2018). Lei No. 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais. Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Câmara dos Deputados. (2026). PL 2338/2023 [Bill tracking record; proposed legislation]. Retrieved October 26, 2026, from https://www.camara.leg.br/proposicoesWeb/fichadetramitacao/?idProposicao=2487262

Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Mental Health, 6(10), e14166. https://doi.org/10.2196/14166

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal, 26(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Jabir, A. I., Lin, X., Martinengo, L., Sharp, G., Theng, Y. L., & Tudor Car, L. (2024). Attrition in conversational agent–delivered mental health interventions: systematic review and meta-analysis. Journal of Medical Internet Research, 26, e48168. https://doi.org/10.2196/48168

Jiang, H., Wang, Z., Shen, W., Meng, M., Yang, D., Li, X., & Hao, Y. (2026). Ai-powered simulation for nursing education: mixed methods systematic review. Journal of Medical Internet Research, 28, e95167. https://doi.org/10.2196/95167

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. npj Digital Medicine, 6(1), 236. https://doi.org/10.1038/s41746-023-00979-5

Li, H., Zhu, G., Zhong, Y., Zhang, Z., Li, S., & Liu, J. (2024). Applications of artificial intelligence in psychiatric nursing: a scope review. Studies in Health Technology and Informatics, 315, 74–80. https://doi.org/10.3233/SHTI240109

Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., & The SPIRIT-AI and CONSORT-AI Working Group (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine, 26(9), 1364-1374. https://doi.org/10.1038/s41591-020-1034-x

Milasan, L. H., & Scott-Purdy, D. (2025). The future of artificial intelligence in mental health nursing practice: an integrative review. International Journal of Mental Health Nursing, 34(1), e70003. https://doi.org/10.1111/inm.70003

Nashwan, A. J., Gharib, S., Alhadidi, M., El-Ashry, A. M., Alamgir, A., Al-Hassan, M., Khedr, M. A., Dawood, S., & Abufarsakh, B. (2023). Harnessing artificial intelligence:

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

strategies for mental health nurses in optimizing psychiatric patient care. Issues in Mental Health Nursing, 44(10), 1020-1034. https://doi.org/10.1080/01612840.2023.2263579

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

Odame, J., Obeng-Gyamfi, G., Heo, D., Kusi, A. B., Kwanin, C., Adarkwah, M. A., & Osei, E. A. (2026). An umbrella review of artificial intelligence applications in mental health care. Journal of Psychiatric and Mental Health Nursing, Advance online publication. https://doi.org/10.1111/jpm.70180

Özdemir, F. E. (2026). Artificial intelligence in mental health nursing: a bibliometric analysis of research trends, knowledge structures and emerging themes. Journal of Psychiatric and Mental Health Nursing, Advance online publication. https://doi.org/10.1111/jpm.70196

Peplau, H. E. (1991). Interpersonal relations in nursing: A conceptual frame of reference for psychodynamic nursing. Springer.

Sanjeewa, R., Iyer, R., Apputhurai, P., Wickramasinghe, N., & Meyer, D. (2024). Empathic conversational agent platform designs and their evaluation in the context of mental health: systematic review. JMIR Mental Health, 11, e58974. https://doi.org/10.2196/58974

Si, E., & Kwon, K. H. (2026). Artificial intelligence in mental health care: implications for psychiatric mental health nursing—a scoping review. International Journal of Mental Health Nursing, 35(5), e70361. https://doi.org/10.1111/inm.70361

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Quality and Safety in Health Care, 19(Suppl 3), i68-i74. https://doi.org/10.1136/qshc.2010.042085

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry, 20(3), 318-335. https://doi.org/10.1002/wps.20883

University of Melbourne. (n.d.). Looking into the psychiatric panopticon: Ethical, legal and social issues of automated nursing observations in acute psychiatric settings [Research project webpage; grey literature]. Centre for AI and Digital Ethics. Retrieved October 26, 2026, from https://www.unimelb.edu.au/caide/research/research-projects/looking-into-the-psychiatric-panopticon-ethical,-legal-and-social-issues-of-automated-nursing-observations-in-acute-psychiatric-settings

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: the achilles heel of predictive analytics. BMC Medicine, 17(1), 230. https://doi.org/10.1186/s12916-019-1466-7

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., …, & the DECIDE-AI expert group (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine, 28(5), 924-933. https://doi.org/10.1038/s41591-022-01772-9

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making, 26(6), 565-574. https://doi.org/10.1177/0272989X06295361

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. Journal of Advanced Nursing, 52(5), 546-553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Wimbarti, S., Kairupan, B. H. R., & Tallei, T. E. (2024). Critical review of self-diagnosis of mental health conditions using artificial intelligence. International Journal of Mental Health Nursing, 33(2), 344-358. https://doi.org/10.1111/inm.13303

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200

World Health Organization. (2022). World mental health report: Transforming mental health for all. https://www.who.int/publications/i/item/9789240049338

World Health Organization. (2025). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. https://www.who.int/publications/i/item/9789240084759

Yıldız, E. (2025). Artificial intelligence in mental health nursing: balancing clinical efficiency and the human touch—a quest for a new synthesis. Journal of Psychiatric and Mental Health Nursing, 32(4), 946–952. https://doi.org/10.1111/jpm.13173

Appendix A Targeted update and evidence access

The update was completed on October 26, 2026. The following are executed discovery queries, transcribed as used. They were submitted to public web search; quoted names and titles were then checked against accessible publisher, PubMed/PMC, Crossref, and official institutional records. This is a targeted update, not a rerun of all originally named databases. Search-result totals depend on the discovery service and were not used as screening denominators.

A1 Scientific literature queries

Q1: Milasan Scott-Purdy 2025 artificial intelligence mental health nursing review

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

Q2: Bond 2026 artificial intelligence person centred mental health nursing review

Q3: Odame 2026 umbrella review artificial intelligence mental healthcare

Q4: Özdemir 2026 bibliometric 141 mental health nursing artificial intelligence

Q5: "Si" "Kwon" "Artificial Intelligence in Mental Health Care" 2026

Q6: "nursing" "generative artificial intelligence" "simulation" "systematic review" 2025

Q7: "AI-Powered Simulation for Nursing Education" authors 2026

Q8: "Calibration: the Achilles heel" 2019

A2 Regulatory and bibliographic verification queries

Q9: site.gov.br anpd sandbox inteligencia artificial 2026

Q10: site.gov.br/anpd "agenda" "inteligência artificial" "2025-2026"

Q11: site.camara.leg.br "2338/2023" "Situação"

Q12: "Looking into the psychiatric panopticon" Gooding

Q13: "10.1111/inm.13303" Wimbarti authors

Additional exact-DOI and title lookups checked the retained bibliography. DOI records were compared with the citation identity; publisher and institutional pages resolved missing fields and publication-year discrepancies. The WHO large multimodal guidance is cited as 2025 following its current official publication page, which dates it March 25, 2025. The Gooding source is undated and is identified as grey literature rather than assigned an unsupported 2024 publication year. Nashwan et al. author and pagination metadata were corrected. Si and Kwon are included irrespective of differing online and issue-date labels; the verified DOI is 10.1111/inm.70361.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 18/09/2026
Accepted: 26/10/2026
Vol:01.05
DOI: 10.66104/cb31g085
Pages: 1–39

A3 Selection and access boundaries

Eleven additional unique sources were retained for the revision: the five named mental-health reviews, the Jiang nursing-simulation review, calibration and decision-curve methods, and ANPD agenda, sandbox, and parliamentary records. Candidate discovery counts, rejected-record counts, and the original deduplication totals are not available as an auditable dataset and are not invented. The revised bibliography is an exact count of cited documents, not a claim of a complete screened research universe. DOI and title matching removed bibliographic duplicates; duplicate primary studies inside secondary reviews were not separately extracted.

Full text was inspected where accessible, including Milasan and Scott-Purdy, Bond et al., Odame et al., Jiang et al., and the calibration methods paper. Abstract or publisher/indexed summaries and DOI metadata informed the limited scope statements for Özdemir and Si and Kwon. Regulatory claims were checked directly on official webpages. Metadata verification establishes citation identity; it does not imply full-text quality appraisal. No subscription-platform access, independent screening, or specialist legal/clinical sign-off is claimed.

The reference ledger contains 35 documents: 25 analytical and 10 normative/contextual. Of the dated records, 23/35 (65.7%) fall in 2020–2026, and 20/35 (57.1%) in 2022–2026. The undated Melbourne webpage remains in the denominator but contributes to neither recent numerator. Official English translations are supplied for Portuguese source descriptors; original legal titles are retained for accurate identification.

CC BY