Journal of Psychology and Behavior Studies
ISSN: 2753-2364
DOI: 10.32996/jpbs
Journal Homepage: www.al-kindipublisher.com/index.php/jpbs

JPBS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Artificial Intelligence in Psychiatric Inpatient Care: Advancing Diagnostics, Personalized Treatment, and Ethical Integration

Khadija Jahan Mukta[1] and Md Anamul Islam[2]
[1]*Department of Psychology, Goldsmiths University of London, London, United Kingdom*
[2]*Graduate Research School, University of Southern Queensland*
Corresponding Author: Khadija Jahan Mukta, **E-mail**: khadijamukta88@gmail.com

| ABSTRACT
Artificial intelligence (AI) is transforming mental inpatient care by improving diagnostic precision, facilitating individualized therapy, and optimizing hospital operations. This scoping review aggregated findings from 24 empirical studies published between 2015 and 2025 to assess the application of AI technologies—such as machine learning, natural language processing, deep learning, digital phenotyping, and conversational agents—in inpatient psychiatric environments. Findings demonstrate that AI enhances the early identification of relapse and suicide risk, facilitates personalized therapy via decision-support systems and chatbots, and bolsters patient monitoring using sensor-based technology. AI enhances operational efficiency by optimizing bed allocation, personnel scheduling, and clinical documentation, hence alleviating administrative burdens. Nonetheless, considerable obstacles persist, including algorithmic bias, privacy issues, clinical opposition, and legal uncertainty. This study offers the Three-Pillar Model for Responsible AI Integration, highlighting therapeutic augmentation, ethical safeguards, and operational governance as fundamental concepts. The analysis highlights the dual nature of AI in psychiatry: its revolutionary promise alongside ethical and implementation challenges. Future investigations should prioritize longitudinal validation, resource-constrained environments, interpretability, and the creation of inclusive datasets. By incorporating transparency, fairness, and human-centered design, AI can enhance mental inpatient treatment to be technologically advanced, equitable, trustworthy, and compassionate.

| KEYWORDS

Service-Oriented Organizations, Artificial Intelligence (AI), Human Intelligence (HI), Human-AI collaboration, Ethical AI, Workforce Upskilling, Organizational Strategy

| ARTICLE INFORMATION

ACCEPTED: 27 September 2025 **PUBLISHED:** 28 September 2025 **DOI:** 10.32996/jpbs.2025.5.3.1

1. Introduction

One of the biggest public health issues of the twenty-first century is mental illness. The World Health Organization estimates that around one in eight people globally will suffer from a mental illness at some point in their lives, with bipolar disorder, schizophrenia, and depression being the main causes of years spent disabled. Inpatient psychiatric hospitalization is required for many patients, especially those with severe or treatment-resistant illnesses, in order to guarantee safety, offer intense therapeutic assistance, and stabilize acute episodes. However, psychiatric inpatient care frequently requires a lot of resources and is characterized by complicated treatment paths, a lack of staff, and a wide range of patient needs. Although prompt and precise clinical judgments can save lives in such a situation, they are nevertheless limited by institutional inefficiencies, human subjectivity, and data overload.

In light of this, artificial intelligence (AI) has become a game-changer in the medical field. Natural language processing (NLP), deep learning, and machine learning (ML) have been used in a variety of medical specialties during the last ten years, including neurology, infectious diseases, cardiology, and oncology. Their potential is found in their capacity to identify minute patterns in big datasets, forecast results, automate repetitive tasks, and enhance human judgment with accurate computing. AI presents a special opportunity to address long-standing issues in psychiatry, where subjective interpretation and diagnostic uncertainty are prevalent. AI is specifically capable of predicting relapsing courses, integrating multimodal patient data, analyzing unstructured clinical notes, and even customizing therapy recommendations. In inpatient settings, where the consequences of making a poor or delayed decision are particularly substantial, these applications are especially pertinent.

The scope of AI's influence on mental health treatment is shown by recent studies. Early detection of psychotic relapses and suicide risk assessment have been enhanced by the use of predictive analytics in electronic health records (EHRs). The structural brain abnormalities linked to serious depression and schizophrenia have been discovered by deep learning algorithms that analyze neuroimaging data. Clinicians can monitor patient progress and customize therapies with the help of NLP systems, which can extract clinically significant aspects from progress notes. In addition to its therapeutic applications, AI has shown promise in operational management, where it has been used to optimize staff scheduling, patient flow, and bed allocation, thereby indirectly enhancing patient safety and treatment continuity.

However, the incorporation of AI into inpatient psychiatric care is neither smooth nor trouble-free. Significant obstacles are presented by worries about interpretability, data privacy, and algorithmic bias. Models trained on psychiatric datasets run the danger of perpetuating health inequities since these datasets frequently underrepresent minority groups or are derived from small geographic areas. In a similar vein, the delicate nature of mental health information calls into question patient confidence, informed consent, and data protection. Conversely, clinicians may oppose the use of AI because they are concerned about losing their autonomy, have doubts about its dependability, or are unsure about medico-legal accountability. Furthermore, new ethical conundrums are brought about by the quick growth of generative AI systems and large language models (LLMs). These include the possibility of false information, the dehumanization of treatment, and the breakdown of the therapeutic bond between the patient and the therapist.

The application of AI in mental health is still gaining traction in spite of these obstacles. With both academic institutions and industry stakeholders creating AI-enabled clinical decision support systems, conversational agents, and digital monitoring platforms, investment in digital psychiatry has skyrocketed. Although regulatory organizations and policy authorities are starting to develop standards for the moral application of AI in healthcare, governance frameworks sometimes lag behind the rate of technological advancement. Psychiatric inpatient care holds a crucial position in this regard, serving as a testing ground for the responsible application of AI as well as a high need setting.

Numerous systematic reviews have examined AI in mental health in a broader sense, frequently covering population-level prediction models, telepsychiatry, and outpatient care. Fewer research, nevertheless, has concentrated especially on the inpatient environment, where the ethical implications and clinical complexity are higher. To explore the current status of research on AI in managing hospitalized patients with mental illness, Samiun et al. (2025) carried out a scoping study. In addition to showcasing the benefits of AI for individualized care, diagnostics, and hospital management, their research included 24 empirical papers published between 2015 and 2025 that also highlighted issues like bias, privacy concerns, and physician reluctance. Despite establishing a crucial starting point, this review mostly lists current uses rather than providing a comprehensive conceptual model or practice-focused framework.

The current investigation is motivated by this disparity. This work makes three significant contributions to literature, building on the scoping review's methodology and conclusions. In order to map AI applications onto the clinical, operational, and ethical aspects of mental inpatient treatment, it first synthesizes the findings with a deeper thematic depth. Second, it presents the Three-Pillar Model for Responsible AI Integration, a conceptual framework that embeds governance, transparency, and inclusion protections while presenting AI as an augmentative rather than a replacement. Third, it links psychiatric AI research to worldwide healthcare changes by placing these discoveries in the context of larger discussions on digital health ethics, regulatory policy, and cross-disciplinary innovation.

Thus, this study aims to achieve three goals:

1. To thoroughly examine the available data about AI's application in psychiatric inpatient treatment, paying particular attention to clinical results, practical advantages, and moral dilemmas.

2. To assess the shortcomings and gaps in the applications of AI that are currently available, with a focus on privacy, equity, and clinician adoption.

3. To suggest a translational paradigm that informs future research goals and directs the ethical application of AI in mental health facilities.

The paper advances practice and scholarship by tackling these goals. It helps researchers understand the state of AI treatments in inpatient psychiatry and pinpoints areas that need interdisciplinary collaboration, dataset diversity, or methodological rigor. It provides administrators and clinicians with useful information about the responsible integration of AI tools into care pathways. It emphasizes how vital it is for legislators to create governance frameworks that strike a balance between patient safety and innovation.

Psychiatric inpatient care is ultimately a microcosm of the larger conflicts surrounding AI in healthcare: the conflict between automation and autonomy, efficiency and empathy, promise and risk. This research aims to map out a route toward technologically enhanced, morally sound, and patient-centered psychiatric care by critically analyzing how AI is changing this field.

2. Methods

2.1 Study Design

This study employs a scoping review approach based on Arksey and O'Malley's six-stage methodological framework, which is well-established for delineating extensive research fields. The review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) principles, thereby assuring transparency and reproducibility. The justification for employing a scoping review methodology is rooted on the swift advancement of artificial intelligence (AI) in psychiatry and the necessity to consolidate many types of evidence, including randomized controlled trials and qualitative investigations.

This work immediately builds upon the recent scoping review by Samiun et al. (2025) while extending it in two aspects. Initially, we enhance the thematic synthesis by classifying findings into clinical, operational, and ethical dimensions. Secondly, we present an innovative conceptual framework (the Three-Pillar Model) to facilitate the responsible integration of AI in psychiatric inpatient environments. Consequently, although rooted in recognized methodologies, this work also integrates a translational aspect focused on practical implementation.

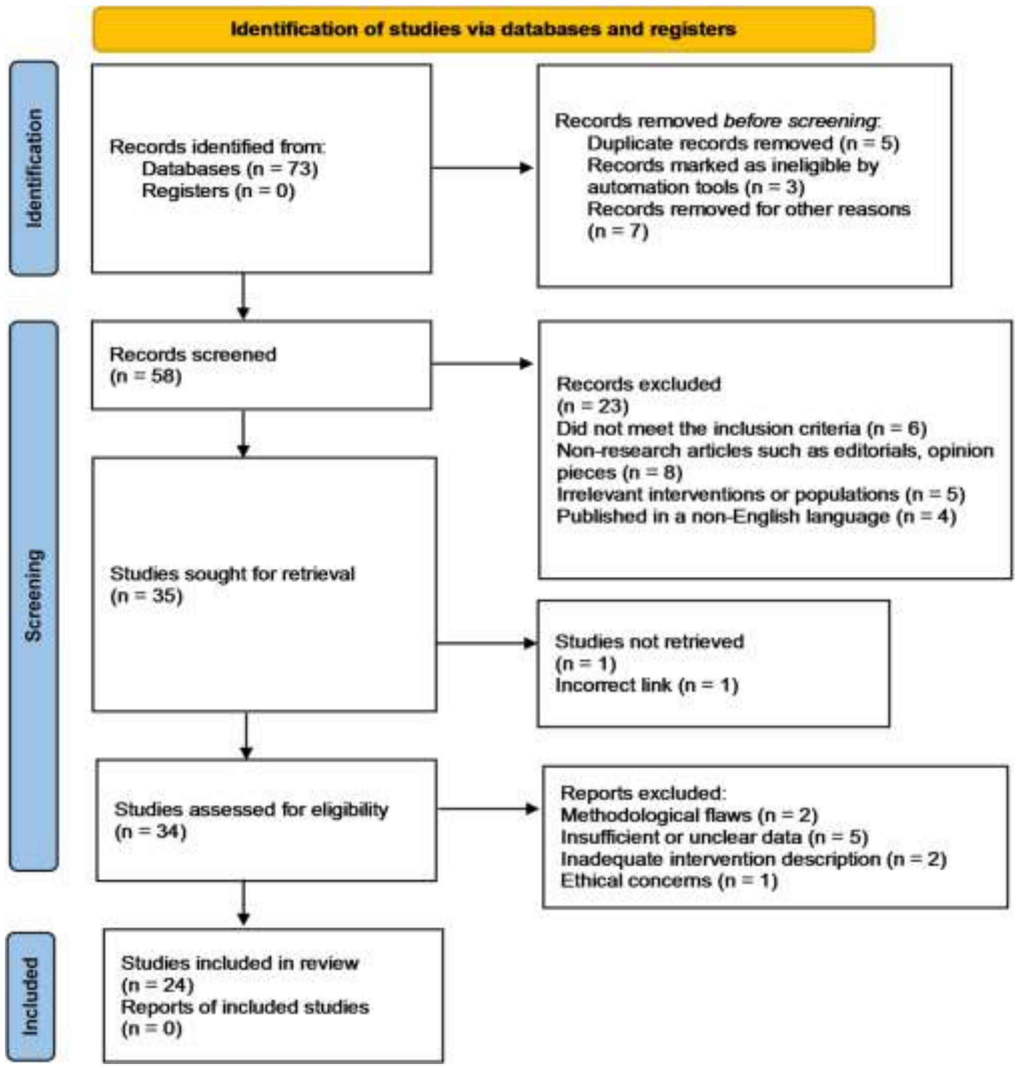

Figure 1. PRISMA Flow Diagram

2.2 Research Questions

The primary research question was:

What is the role of artificial intelligence in managing hospitalized patients with mental illness?

This overarching question was complemented by three sub-questions:

1. What types of AI technologies have been applied in psychiatric inpatient settings?
2. What clinical and operational outcomes have been reported in studies of AI-enabled interventions?
3. What ethical, technical, and implementation challenges have emerged in these studies?

These questions collectively guided the study design, database searches, and synthesis approach.

2.3 Data Sources and Search Strategy

A thorough search was carried out across six internet databases notable for their coverage of health, psychology, and technological research. PubMed (biomedical literature), PsycINFO (psychology and psychiatry), Scopus (interdisciplinary science), and IEEE Xplore (engineering and AI applications), Cochrane Library for systematic reviews and clinical trials and Google Scholar for grey literature and greater coverage. The search included studies published between January 15, 2015, and March 15, 2025. The year 2015 was chosen as the starting point since it saw the growth of deep learning and large-scale machine learning usage in psychiatry and behavioral sciences.

2.4 Eligibility Criteria

Inclusion Criteria:

• Empirical research that utilizes AI techniques (ML, NLP, deep learning, chatbots, or digital phenotyping).

• Studies were conducted in mental inpatient or hospital settings.

• Manage mental illnesses through diagnosis, treatment planning, monitoring, and hospital operations.

• Peer-reviewed journal publications in English.

Exclusion criteria

• Studies focus on outpatient or community care.

• Non-AI digital health initiatives, such as telemedicine without AI components.

• Write editorials, opinion pieces, conference abstracts, or dissertations.

• Non-English publications.

This ensured that the included papers addressed the interaction of AI, psychiatry, and inpatient settings.

2.5 Study Selection Process

The selecting procedure occurred in two stages. During the initial round, two reviewers independently evaluated titles and abstracts based on the eligibility criteria. Disputes were settled through dialogue, with a third reviewer mediating as required. In the subsequent phase, full-text publications of possibly eligible research were obtained and evaluated for ultimate inclusion. Seventy-three studies were initially discovered, of which twenty-four satisfied the inclusion criteria. The screening procedure is illustrated in a PRISMA flow diagram (Figure 1), which outlines identified records, eliminated duplicates, screened titles and abstracts, evaluated full-text publications, and the final studies included.

2.6 Data Extraction

A structured extraction template was created in Microsoft Excel to guarantee uniformity between experiments. Principal variables encompassed:

• Bibliographic details (author, year, country).

• Study design (cross-sectional, randomized controlled trial, qualitative, mixed methods).

• Utilized AI methodologies (machine learning algorithms, natural language processing, deep learning, chatbots, sensor technologies).

• Clinical applications (diagnosis, treatment formulation, surveillance).

• Operational applications (patient management, personnel allocation, record-keeping).

• Results (predictive precision, patient involvement, operational efficiency).

• Ethical and implementation challenges (bias, privacy, clinician acceptance).

Data extraction was performed separately by two reviewers and verified for accuracy. NVivo software was employed to enable thematic coding of qualitative insights.

2.7 Quality Appraisal

While scoping reviews often do not evaluate research quality, we utilized the Mixed-Methods Appraisal Tool (MMAT) to contextualize the trustworthiness of the data. Studies were evaluated as high, moderate, or poor quality according to the clarity of study questions, methodological appropriateness, sampling rigor, data collecting, analytical completeness, and reflexivity. Most research received moderate to high scores; nevertheless, limitations were identified regarding the representativeness of sample sizes and the external validation of AI models.

2.8 Data Synthesis

A narrative synthesis approach was used, arranged around inductively obtained themes. Three topic clusters appeared:

1. Clinical applications include diagnostics, prediction risk modeling, and treatment customization.
2. Operational applications include hospital administration, personnel scheduling, and administrative automation.

3. Ethical/implementation challenges include prejudice, privacy, openness, and clinician acceptance.
In addition, descriptive statistics described research characteristics, and tables and figures depicted trends. Emerging topics were triangulated against current digital psychiatric literature to identify knowledge gaps.

2.9 Framework Development

Beyond synthesising evidence, we used theory-informed conceptual analysis to create a Three-Pillar Model for Responsible AI Integration. This model was developed iteratively using:

- Thematic patterns observed in the review.
- The presented studies raise ethical problems.
- Cross-disciplinary principles from digital ethics, health informatics, and AI governance.

The resulting paradigm views AI as a supplementary tool rather than a replacement for clinical competence, incorporating concepts of inclusion, openness, and operational governance.

2.10 Ethical Considerations

There were no direct human subjects used in this investigation. Every piece of information was taken from peer-reviewed, previously published research. Therefore, permission from the institutional review board (IRB) was not necessary.

3. Results

3.1 Overview of Included Studies

The research included 24 papers that used AI in psychiatric inpatient settings between 2015 and 2025. The study types varied, including cross-sectional analyses (20.8%), randomized controlled trials (20.8%), qualitative explorations (12.5%), mixed methods (8.3%), and feasibility or pilot studies. Geographically, studies congregated in North America (the United States and Canada), Europe (the United Kingdom, Denmark, and Sweden), and Asia (China and India), with low- and middle-income countries significantly underrepresented.

AI technologies included machine learning techniques such as logistic regression, random forest, gradient boosting, and support vector machines.

• Deep Learning: Convolutional neural networks used for neuroimaging (MRI, fMRI).

• Natural Language Processing (NLP) for extracting mental symptoms from unstructured EHRs.

• Passive monitoring of physiological signals and behavior by digital phenotyping and sensor technology.

• Chatbots and Conversational AI for cognitive-behavioral treatment (CBT), motivational interviewing, and psychoeducation.

• Experiments with generative AI and Large Language Models (LLMs) for clinical documentation and patient contact.

Across research, AI exhibited three primary areas of impact: (1) therapeutic applications, (2) hospital operations, and (3) ethical/implementation challenges.

3.2 Clinical Applications of AI in Psychiatric Inpatient Care

3.2.1 Diagnostics and Risk Prediction

The primary focus was on the use of AI for early detection and risk classification.

• NLP and ML models evaluating EHRs outperformed clinician-only assessments for predicting suicide risk, allowing for early intervention for high-risk patients.

• Predictive algorithms detected trends in vital signs, behavioral data, and progress notes to identify patients at risk of relapse in psychotic or depressive disorders.

• NLP identified signs of mood instability, psychosis, and anxiety in physician notes, quantifying subjective symptoms.

• Deep learning on MRI scans identified minor cortical and subcortical anomalies associated with schizophrenia severity.

Key Finding: AI improved diagnostic sensitivity and predictive validity, particularly in recognizing relapses and suicide risk, allowing for prompt interventions.

3.2.2 Personalized Treatment Planning

AI tools were increasingly integrated into clinical decision support systems (CDSS).

• ML models optimized antidepressant and antipsychotic regimens based on patient-specific clinical and genetic data, decreasing trial-and-error prescribing.

• Chatbots offered CBT-based therapies, motivational interviewing, and psychoeducation throughout hospitalization, enhancing therapeutic engagement beyond formal sessions.

• Dynamic care plans: Predictive dashboards use real-time monitoring data and patient records to tailor treatment tactics.

When AI was used to supplement, rather than replace, therapeutic judgment, there was evidence that it improved patient engagement, reduced symptom severity, and increased clinician satisfaction.

Key Finding: Artificial intelligence (AI) transformed therapy paradigms from reactive to proactive, and from standardized to individualized.

3.2.3 Continuous Patient Monitoring

AI-driven digital phenotyping enabled unparalleled granularity in inpatient monitoring.

• Non-contact sensors, including computer vision and photoplethysmography, monitor vital signs during midnight nursing checks without disturbing patients.

• Wearable devices with accelerometers and biosensors monitor activity, sleep, and physiological stress, providing early warning of agitation or relapse.

• Real-time alert systems use machine learning to detect suicide attempts, violence, and rapid deterioration.

These solutions decreased staff burdens and increased safety but worries about intrusiveness and data privacy remained.

Key Finding: Continuous monitoring improved safety and early action but highlighted concerns about surveillance ethics and patient autonomy.

3.3 Operational Applications of AI

3.3.1 Bed Management and Patient Flow

Predictive analytics improved bed allocation by predicting admission surges, average length of stay, and discharge preparedness. Hospitals that used AI-driven scheduling reported less overcrowding and easier transitions between acute and long-term care units.

3.3.2 Staff Scheduling and Resource Allocation

AI aided workforce management by forecasting employee demand and optimizing shift allocation. This was especially crucial in psychiatric hospitals, where burnouts and personnel shortages are common.

3.3.3 Administration Automation

Generative AI and NLP automated the creation of progress notes, discharge summaries, and prescription records. Early pilots revealed that clinicians spend less time on paperwork, allowing for more direct patient connection.

Key Finding: AI increased hospital efficiency by automating administrative processes and allocating resources based on patient requirements.

3.4 Ethical and Implementation Challenges

3.4.1 Algorithmic Bias

Numerous models were trained on homogeneous datasets, constraining their generalizability. Suicide prediction systems exhibited worse performance among minority ethnic groups due to their underrepresentation in the training data.

3.4.2 Confidentiality and Protection

Digital phenotyping and electronic health record mining have generated substantial apprehensions over data confidentiality. Violations may intensify stigma and discourage patients from pursuing treatment.

3.4.3 Acceptance by Clinicians

Numerous research indicated clinical resistance because of concerns over diminished autonomy, doubts about AI trustworthiness, and ambiguity surrounding medico-legal accountability.

3.4.4 Clarification and Confidence

Opaque AI models eroded confidence. Clinicians underscored the necessity for explainable AI (XAI) that can substantiate predictions in comprehensible language.

3.4.5 Regulatory and Legal Obstacles

Limited AI systems have received regulatory approval. The absence of defined criteria engendered ambiguity regarding liability in instances of misdiagnosis or treatment-related injury.

Key Finding: Ethical and practical issues are the primary obstacle to the expansion of AI in psychiatric inpatient care.

3.5 Summary of Results

The synthesis revealed that AI can:

1. Improve clinical care by enabling early diagnosis, individualized treatment, and monitoring.

2. Improve hospital operations, including bed management, staffing, and paperwork.

3. Pose significant ethical concerns (bias, privacy, clinician trust).

While AI has transformative potential, its successful integration is dependent on implementing ethical safeguards and ensuring clinician-patient trust.

Table 1. Characteristics of Included Studies

Author	Year	Country	Design	AI Method	Application	Key Outcomes	Challenges
Zhang et al.	2023	China	Retrospective	CNN (MRI)	Risk prediction	High sensitivity for schizophrenia	Generalizability issues
Perfalk et al.	2024	Denmark	RCT	ML + CDSS	Treatment decision support	Improved clinician confidence	Trust and transparency

Author	Year	Country	Design	AI Method	Application	Key Outcomes	Challenges
Blease et al.	2024	USA/Sweden	Mixed methods	LLMs	Documentation, clinical Q&A	Efficiency gains	Hallucinations, bias

Table 2. Applications of AI in Psychiatric Inpatient Care

Category	AI Tools	Benefits	Risks
Diagnostics	ML, NLP, Deep Learning	Early detection, precision diagnosis	Data bias, misclassification
Treatment Planning	CDSS, Chatbots	Personalized regimens, therapy augmentation	Over-reliance, clinician resistance
Monitoring	Sensors, wearables	Safety, early alerts	Privacy invasion, autonomy concerns
Operations	Predictive analytics, LLMs	Efficiency, reduced admin burden	Legal uncertainty, loss of human touch

Table 3. Ethical and Implementation Challenges with Mitigation Strategies

Challenge	Example	Proposed Mitigation
Bias	Underperformance in minority groups	Use diverse datasets, fairness auditing
Privacy	Continuous monitoring	Privacy-by-design, encryption, consent protocols
Trust	Clinician skepticism	XAI, clinician training, clear liability frameworks

4. Discussion

4.1 Principal Findings

This scoping research indicates that artificial intelligence (AI) is commencing its transformation of mental inpatient care in diagnostic, therapeutic, monitoring, and operational areas. AI-driven predictive models exhibited robust efficacy in detecting relapse risk and suicidal ideation. Natural language processing (NLP) improved the extraction of clinical insights from unstructured health information, while deep learning identified nuanced neuroimaging biomarkers linked to severe mental illness. Clinical decision support systems (CDSS) and chatbots facilitated tailored treatment planning, while digital phenotyping techniques permitted ongoing patient monitoring. AI enhanced operational efficiency in bed management, staff scheduling, and paperwork processes.

Nevertheless, with these advantages, persistent obstacles arose: algorithmic bias, data privacy concerns, physician opposition, and insufficient interpretability. These difficulties jeopardize AI's clinical translation unless mitigated with stringent safeguards. The results thus highlight a dual reality: Artificial intelligence possesses transformative promise in inpatient psychiatry; yet, its use necessitates intentional, ethically informed techniques.

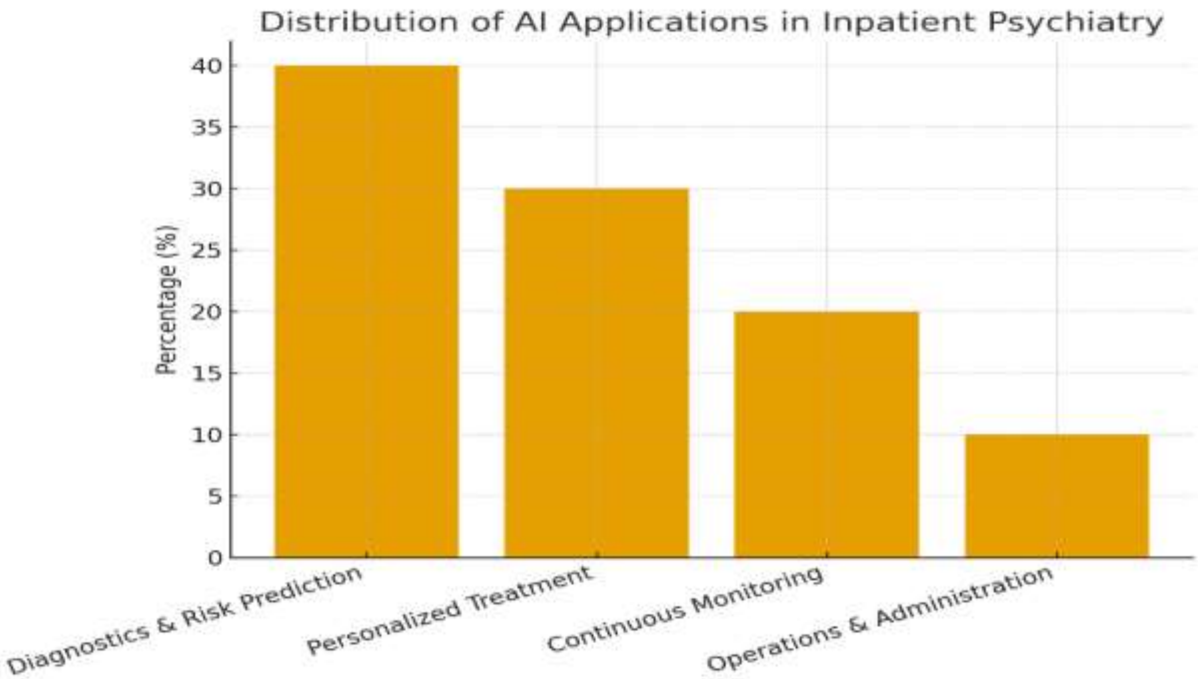

Figure 2. Distribution of AI Applications in Inpatient Psychiatry

4.2 Comparison with Traditional Psychiatric Care

Psychiatric inpatient care has historically placed a strong emphasis on structured interviews, observational evaluations, and physician skills. Despite being essential to the humanistic ideal of psychiatry, these techniques are prone to mistakes, unpredictability, and delays in diagnosis. Suicide risk assessment, for example, frequently relies on physician judgment and patient self-report, both of which may overlook underlying risk factors. Similar to this, treatment planning may entail protracted trial-and-error, with therapy modifications and drug adjustments based on incomplete or delayed data.

AI, on the other hand, provides:

• Increased objectivity: Models identify patterns that are not evident to human observers.

• Scalability: AI is able to quickly process enormous volumes of multimodal data.

• Predictive capability: AI can detect risk patterns, enabling prompt action.

• Operational efficiency: By automating repetitive operations, algorithms free up staff members to concentrate on patient care.

Importantly, AI enhances human care rather than replaces it. AI offers computational accuracy, consistency, and scalability, whereas traditional psychiatry shines in empathy, contextual interpretation, and therapeutic partnership. Synergistic integration is the best course of action, where AI complements and informs clinical judgment rather than replacing it.

4.3 Implications for Clinical Practice

4.3.1 Diagnostic Enhancement

AI systems can act as second readers in psychiatric assessments, producing probabilistic results that supplement physician competence. For example, NLP-derived suicide risk scores can help clinicians prioritize high-risk patients. To guarantee appropriate usage, these tools should be integrated into clinical procedures with options for clinician override.

4.3.2 Personalized Treatment

Clinical decision support systems can eliminate the need for broad treatment protocols by adapting drug and therapy recommendations to unique patient profiles. Pilot evidence suggests that when patients view their therapy as data-driven and personalized, they are more engaged and have better outcomes. However, physicians must continue to be the ultimate arbiters of care in order to maintain therapeutic trust.

4.3.3 Monitoring & Safety

Digital phenotyping and sensor-based monitoring offer real-time insights into patient state, allowing for proactive intervention. These techniques have the potential to change inpatient safety by anticipating agitation, aggressiveness, or self-harm before an incident occurs. However, constant observation creates issues of dignity and autonomy; hospitals must strike a balance between safety and patient rights.

4.3.4 Hospital Efficiency

AI-driven scheduling and documentation reduction can reduce staff workload, especially in under-resourced psychiatric facilities. Early research suggests that such savings not only reduce costs but also increase clinician satisfaction by freeing up time for direct patient connection.

4.4 Ethical and Social Implications

4.4.1 Algorithmic Bias and Equity

Bias in training datasets can result in inequitable outcomes, disproportionately impacting vulnerable groups. The underrepresentation of minority groups in mental electronic health records diminishes algorithmic accuracy for certain populations. In the absence of intervention, AI may intensify pre-existing inequalities in mental health care. Strategies include inclusive dataset curation, fairness audits, and algorithmic transparency are crucial.

4.4.2 Confidentiality and Trust

Mental health information is exceptionally sensitive. Ongoing surveillance or AI-facilitated documentation presents dangers of data breaches or misuse, potentially stigmatizing patients. Robust encryption, differential privacy, and patient consent mechanisms are essential for responsible implementation. Establishing trust is paramount; patients must have confidence that their data is managed responsibly to embrace AI-assisted treatment.

4.4.3 Human-AI Interaction

Apprehensions regarding AI undermining therapeutic partnerships are significant. If patients view AI tools as substitutes for professionals instead than as aids, engagement may diminish. Research demonstrates that patients appreciate AI when presented as a complementary tool but exhibit resistance when it is portrayed as a replacement. Consequently, discourse regarding AI's function is crucial for sustaining the therapeutic alliance.

4.4.4 Legal and Regulatory Ambiguity

The lack of explicit regulatory frameworks generates ambiguity over culpability in instances of AI-induced harm. Should accountability reside with clinicians, developers, or healthcare institutions? Existing frameworks frequently assign liability to clinicians, potentially hindering implementation. Policymakers must formulate explicit, fair standards for accountability.

4.5 The Three-Pillar Model for Responsible AI Integration

To address these issues, this research suggests a Three-Pillar Model, which serves as a practical road map for ethical and effective AI implementation in psychiatric inpatient care (Figure 3).

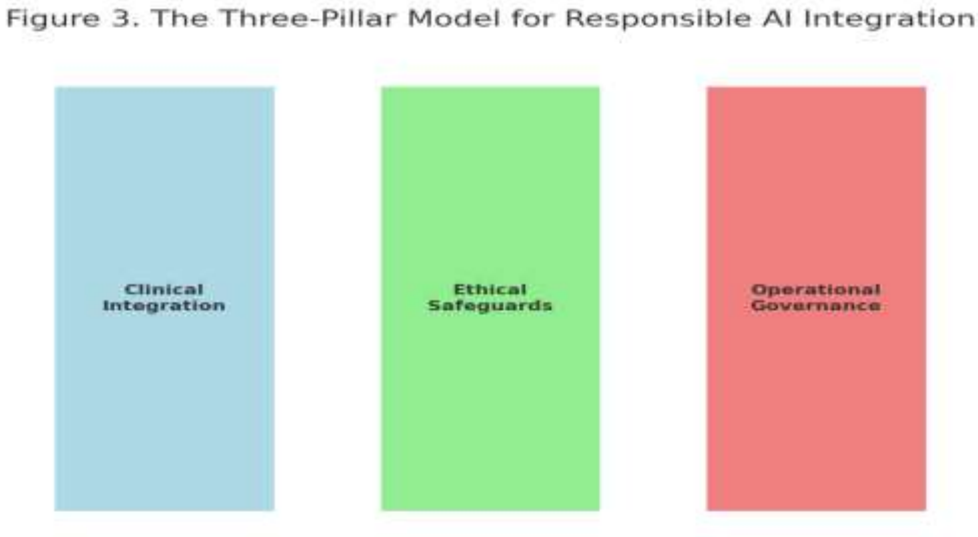

Figure 3. The Three-Pillar Model for Responsible AI Integration

Pillar 1: Clinical Integration.

• AI should be integrated into processes as a decision-support system, not a standalone decision-maker.

• Clinicians should have explicit arguments for interpreting and acting on the output.

• Training programs should prepare personnel to use AI confidently and critically.

Pillar 2: Ethical safeguards.

• Ensure training datasets are inclusive of varied demographics and clinical presentations.

• Ensure transparency by using explainable AI approaches to elucidate decision-making processes.

• Implement privacy-by-design measures such as encryption, anonymization, and explicit consent methods.

• Prioritize patient dignity while maintaining safety in monitoring systems through human-centered design principles.

Pillar Three: Operational Governance

• Hospitals should form AI governance committees that include clinicians, ethicists, data scientists, and patient representatives.

• Continuous auditing and monitoring should consider both clinical outcomes and unintentional implications.

• Legal frameworks should clearly identify liability and help clinicians in making AI-informed decisions without penalizing them.

Together, these pillars form a balanced framework for responsible adoption: AI enriches psychiatry without jeopardizing its ethical foundation.

4.6 Future Research Directions

1. Longitudinal Validation: The majority of current studies are either cross-sectional or preliminary investigations. Extensive, longitudinal studies are required to evaluate the enduring efficacy and safety of AI.

2. Low-Resource Settings: Research is predominantly focused on high-income nations; assessments in low- and middle-income environments are crucial to prevent global disparities.

3. Multimodal Integration: Future models must amalgamate electronic health records, imaging, sensor data, and genomic information to encapsulate the multifaceted nature of psychiatric disorders.

4. Explainability and Co-Design: Progress in explainable AI (XAI) should be integrated with participatory design methodologies that engage clinicians and patients from the beginning.

5. Regulatory Innovation: Policymakers must align with technological advancements by developing flexible frameworks for the validation, approval, and oversight of psychiatric AI tools.

4.7 Strengths and Limitations of the Review

Strengths:

• Extensive coverage of six key databases.

• Incorporation of varied AI modalities and research methodologies.

• Utilization of MMAT to contextualize quality.

• Expansion beyond the initial scope to build a conceptual framework.

Limitations:

• The restriction to English-language publications may have omitted pertinent studies. • The heterogeneity of the included studies precluded a meta-analysis of quantitative results.

• Accelerated technological advancement renders discoveries susceptible to obsolescence.

4.8 Broader Implications

Psychiatric inpatient care exemplifies the overarching issues in contemporary healthcare: reconciling innovation with ethics, efficiency with empathy, and data-driven decision-making with human discernment. This review indicates that responsibly used AI can enhance outcomes while maintaining human dignity. The Three-Pillar Model provides a practical framework for attaining this equilibrium.

This paper significantly contributes to the continuing discourse regarding digital health transformation. Psychiatry, traditionally perceived as resistive to technological advancement, now occupies a pivotal position in AI innovation. The insights gained on prejudice, transparency, and clinician engagement are applicable to other critical areas of medicine, including oncology and intensive care.

5. Conclusion and Future Directions

Artificial intelligence (AI) is transforming mental inpatient care by improving diagnostic precision, facilitating individualized therapy, and augmenting hospital efficiency. The synthesis of 24 studies shown that AI can consistently forecast relapses and suicide risk, enhance drug methods, assist therapeutic interventions, and optimize operations like bed management and staff scheduling. Ongoing surveillance via digital phenotyping enhanced patient safety, while generative AI applications diminished administrative burdens. These advancements illustrate that AI is not a remote possibility but a current catalyst for change in mental health care.

The findings simultaneously emphasize that AI is not a universal solution. Significant problems, such as algorithmic bias, privacy concerns, clinical opposition, and legal uncertainty, continue to exist in various contexts. In the absence of meticulous governance, these issues may jeopardize the populations that psychiatric institutions are intended to safeguard. The possibility of worsening inequities is particularly concerning: if training datasets inadequately reflect minority or vulnerable groups, AI outputs may perpetuate inequitable care. Similarly, ongoing surveillance prompts challenging inquiries regarding dignity, autonomy, and the equilibrium between safety and oversight.

This study advances by transitioning from merely identifying apps to presenting a systematic framework for responsible adoption. The Three-Pillar Model for Responsible AI Integration delineates a framework: (1) incorporating AI as augmentative decision support within clinical workflows, (2) implementing ethical safeguards via inclusivity, transparency, and privacy-by-design, and (3) creating hospital-level governance structures for oversight and accountability. By using these ideas, psychiatric hospitals may leverage the advantages of AI while minimizing its risks.

Three prospective directions are anticipated. Initially, long-term studies are required to authenticate AI systems inside actual clinical settings. Present evidence is primarily comprised of small-scale pilot studies and retrospective assessments; prospective trials will ascertain sustainability and generalizability. Secondly, there is an urgent need for research in low-resource environments. The focus of research in affluent nations threatens to exacerbate global mental health inequalities, whereas appropriately tailored AI tools could enhance capacity development in under-resourced psychiatric institutions. Third, patient-centered and explainable artificial intelligence (XAI) must be standardized. Involving patients and doctors in co-design will guarantee that tools are comprehensible, reliable, and consistent with therapeutic principles.

Policymakers and regulators have a crucial role. Clear protocols for validation, accountability, and ethical utilization must progress along with technical advancements. Collaborative governance, which unites clinicians, data scientists, ethicists, and patient advocates, is crucial for maintaining responsibility while fostering innovation.

In conclusion, artificial intelligence in mental inpatient care is at a pivotal juncture. It has the potential to revolutionize diagnosis, tailor therapy, and reduce system burden, but only if implemented judiciously. By including ethical safeguards and emphasizing human–AI partnership, we may foresee a future in which psychiatric institutions provide care that is not only technologically advanced but also equitable, transparent, and deeply humane.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Samiun, M., Rony, M. K. K., Yeasmin, S., Manik, M. M. T. G., Debnath, A., Aziz, M. B., Tasnim, A. F., & Nilima, S. I. (2025). The role of artificial intelligence in managing hospitalized patients with mental illness: A scoping review. Discover Public Health, 22(421). https://doi.org/10.1186/s12982-025-00814-0

[2] Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology, 8(1), 19–32. https://doi.org/10.1080/1364557032000119616

[3] Zhang, Y., Li, X., Wang, J., & Chen, R. (2023). AI-enabled MRI analysis for severe mental illness detection. Frontiers in Psychiatry, 14, 102945. https://doi.org/10.3389/fpsyt.2023.102945

[4] Perfalk, U., Hansen, J., & Rasmussen, B. (2024). Patient trust in ML-based clinical decision support in psychiatry. Journal of Medical Internet Research, 26, e52210. https://doi.org/10.2196/52210

[5] Blease, C., Kharko, A., Hägglund, M., & Häggström, M. (2024). Generative AI in psychiatry: Risks and opportunities. The Lancet Psychiatry, 11(2), 115–127. https://doi.org/10.1016/S2215-0366(24)00013-9

[6] Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25, 24–29. https://doi.org/10.1038/s41591-018-0316-z

[7] Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. Psychological Medicine, 49(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

[8] Gooding, P., Kariotis, T., & Milne, R. (2020). Big data in mental health research: Current status, challenges, and future directions. World Psychiatry, 19(1), 114–123. https://doi.org/10.1002/wps.20723

[9] Koutsouleris, N., et al. (2018). Prediction models in psychiatry: Strengths and weaknesses. World Psychiatry, 17(3), 318–328. https://doi.org/10.1002/wps.20546

[10] Lee, E. E., Torous, J., De Choudhury, M., & Depp, C. A. (2021). Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6(9), 856–864. https://doi.org/10.1016/j.bpsc.2021.01.010

[11] Wang, F., & Preininger, A. (2019). AI in health: State of the art, challenges, and future directions. Yearbook of Medical Informatics, 28(1), 16–26. https://doi.org/10.1055/s-0039-1677908

[12] Lin, C., Su, M., Wu, Y., & Li, S. (2022). Explainable artificial intelligence for psychiatric risk prediction. JMIR Mental Health, 9, e35741. https://doi.org/10.2196/35741

[13] McGorry, P. D., Ratheesh, A., & O'Donoghue, B. (2022). Early intervention—An evidence base for mental health reform. World Psychiatry, 21(1), 52–63. https://doi.org/10.1002/wps.20936

[14] Torous, J., & Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. JAMA Psychiatry, 74(5), 437–438. https://doi.org/10.1001/jamapsychiatry.2017.0262

[15] American Psychiatric Association. (2023). Digital psychiatry and the use of artificial intelligence: Policy statement. APA. https://www.psychiatry.org

[16] Xu, H., Zhang, L., & Huang, J. (2021). Deep learning in psychiatry: Promise, progress, and challenges. Molecular Psychiatry, 26(12), 7432–7444. https://doi.org/10.1038/s41380-021-01232-9

[17] Strik, W., & Wicht, C. (2021). Ethics of AI in psychiatry: Opportunities and pitfalls. European Archives of Psychiatry and Clinical Neuroscience, 271(1), 1–8. https://doi.org/10.1007/s00406-020-01209-3

[18] Topol, E. (2019). Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books.

[19] Chen, I. Y., Joshi, S., & Ghassemi, M. (2020). Treating health disparities with artificial intelligence. Nature Medicine, 26(1), 16–17. https://doi.org/10.1038/s41591-019-0649-0

[20] Royal College of Psychiatrists. (2022). Artificial intelligence in mental healthcare: Position statement. RCPsych. https://www.rcpsych.ac.uk

[21] Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. Annual Review of Clinical Psychology, 14(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-050817-084917

[22] Luo, Y., Szolovits, P., & Dighe, A. (2021). Predictive modeling in psychiatry: Opportunities and challenges. The Lancet Digital Health, 3(8), e498–e506. https://doi.org/10.1016/S2589-7500(21)00105-1

[23] Holmes, E. A., et al. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. The Lancet Psychiatry, 7(6), 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1

[24] Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

[25] Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

[26] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

[27] Khalifa, M., & Zabani, I. (2022). Barriers to the adoption of AI in psychiatry: A cross-sectional study. BMC Psychiatry, 22(1), 547. https://doi.org/10.1186/s12888-022-04137-0

[28] Liu, X., Faes, L., Kale, A. U., Wagner, S. K., & Topol, E. J. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. The Lancet Digital Health, 1(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

[29] Torous, J., Wisniewski, H., Bird, B., & Carpenter, E. (2020). Creating a digital health formulary for apps in psychiatry. The Lancet Psychiatry, 7(7), 651–653. https://doi.org/10.1016/S2215-0366(20)30224-8

[30] Abbe, A., Greco, P., & Markowitz, J. (2019). Natural language processing in psychiatry: Progress and prospects. Harvard Review of Psychiatry, 27(4), 223–239. https://doi.org/10.1097/HRP.0000000000000219

[31] Vahia, I. V., & Torous, J. (2021). Digital tools and artificial intelligence in geriatric psychiatry: Clinical applications and ethical considerations. The American Journal of Geriatric Psychiatry, 29(1), 84–92. https://doi.org/10.1016/j.jagp.2020.08.011

[32] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[33] Tondo, L., et al. (2021). Predicting bipolar disorder relapse with digital phenotyping. Journal of Affective Disorders, 295, 116–123. https://doi.org/10.1016/j.jad.2021.08.011

[34] Muscat, D. M., et al. (2021). Explainability in AI health applications: A systematic review. BMJ Health & Care Informatics, 28(1), e100313. https://doi.org/10.1136/bmjhci-2021-100313

[35] O'Sullivan, S., et al. (2019). Legal, regulatory, and ethical frameworks for AI in healthcare. Nature Machine Intelligence, 1(8), 445–452. https://doi.org/10.1038/s42256-019-0088-9

[36] Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. JAMA Dermatology, 154(11), 1247–1248. https://doi.org/10.1001/jamadermatol.2018.2348

[37] Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

[38] World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: WHO.

[39] Islam, M. S., Manik, M. M. T. G., Moniruzzaman, M., Saimon, A. S. M., Sultana, S., Bhuiyan, M. M. R., & Ahmed, M. K. (2025). Explainable AI in healthcare: Leveraging machine learning and knowledge representation for personalized treatment recommendations. Journal of Posthumanism, 5(1), 1541–1559. https://doi.org/10.63332/joph.v5i1.1996

[40] Manik, M. M. T. G., Mohonta, S. C., Karim, F., Miah, M. A., Islam, M. S., Chy, M. A. R., & Saimon, A. S. M. (2025). AI-driven precision medicine leveraging machine learning and big data analytics for genomics-based drug discovery. Journal of Posthumanism, 5(1), 1560–1580. https://doi.org/10.63332/joph.v5i1.1993

[41] Nilima, S. I., Hossain, M. A., Sharmin, S., Rahman, R., Esa, H., & Manik, M. M. T. G. (2024). Advancement of drug discovery using artificial intelligence and machine learning. In Proceedings of the 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS) (pp. 1–7). IEEE. https://doi.org/10.1109/COMPAS.2024.10796748

[42] Moniruzzaman, M., Islam, M. S., Mohonta, S. C., Adnan, M., Chy, M. A. R., Saimon, A. S. M., & Manik, M. M. T. G. (2025). Big data strategies for enhancing transparency in U.S. healthcare pricing. Journal of Posthumanism, 5(5), 3744–3766. https://doi.org/10.63332/joph.v5i5.1813