BHARATI INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & DEVELOPMENT (BIJMRD)

(Open Access Peer-Reviewed International Journal)

DOI Link : https://doi.org/10.70798/Bijmrd/04042012

Available Online: www.bijmrd.com | BIJMRD Volume: 4 | Issue: 04(2) | April 2026 | e-ISSN: 2584-1890

Artificial Intelligence-Based Mental Health Interventions in India: Psychological Risks, Ethical Concerns, and the Urgent Need for Regulatory Oversight

Gowrika K, Dr. Patteswari D & Dr. Sangeetha S R

1. Research Scholar, Department of Psychology and Cognitive Neuroscience, JSS Academy of Higher Education and Research, Mysuru, Karnataka, India
2. Corresponding author & Associate Professor, Division of Psychology and Cognitive Neuroscience, JSS Academy of Higher Education and Research, Mysuru, Karnataka, Email: Indiapatteswari@jssuni.edu.in
3. Corresponding author & Assistant Professor, Division of Psychology and Cognitive Neuroscience, JSS Academy of Higher Education and Research, Mysuru, Karnataka, Email: Indiasangeethasr@jssuni.edu.in

Abstract: ***Background:*** *Chatbots for mental health that use artificial intelligence (AI) are rapidly expanding in India as scalable alternatives to address the significant treatment gap in the country. Growing quickly in India as scalable alternatives to overcome the significant treatment deficit in the nation. Concerns about clinical safety, ethical standards, accountability, and data privacy have been raised, yet, considering their deployment has taken place without a specific legislative framework.*

Aim*: In light of emerging global best practices, this study critically analyses the legislative deficiencies governing AI-based mental health interventions in India and proposes a contextually relevant policy framework.*

Methods: *A scoping review methodology was employed. A systematic analysis and thematic synthesis of peer-reviewed literature, ethical analyses, policy papers, and regulatory frameworks published between 2017 and 2025 were conducted, with a focus on their relevance to the Indian mental healthcare system.*

Results: *The assessment found substantial regulatory gaps in current Indian legislation, including the Digital Personal Data Protection Act (2023) and the Mental Healthcare Act (2017). Unregulated AI mental health chatbots may generate clinically unsafe responses; perpetuate cultural bias, compromise privacy, and obscure accountability. Comparative analysis indicates that mental health AI systems are increasingly classified as high-risk technologies requiring multi-layered oversight in several Western jurisdictions.*

Conclusion: *A dedicated, risk-based regulatory framework for AI mental health systems is of urgent importance in India. Ensuring that AI technologies expand access to mental healthcare without compromising patient safety or dignity requires an adaptive governance framework integrating clinical validation, ethics-of-care principles, and human oversight.*

Keywords: *Artificial Intelligence; Mental Health Policy; Chatbots; Psychotherapy; Regulation; India.*

Introduction: With less than 0.8 psychiatrists per 100,000 people in rural areas, India is experiencing a severe and ongoing mental health crisis marked by a large treatment gap, inadequate infrastructure, stigma,

and a lack of qualified professionals (NIMHANS, 2022). According to epidemiological estimates, a considerable percentage of people with depression, anxiety disorders, substance use disorders, and suicidality do not receive proper care. Access to timely and reasonably priced psychological services is still uneven, especially in rural and socioeconomically marginalized communities, despite the rights-based protections outlined in the Mental Healthcare Act of 2017.

Digital therapeutic platforms and chatbots powered by artificial intelligence (AI) have surfaced as scalable substitutes for this systemic deficiency. These systems use machine learning and natural language processing to track mood patterns, provide cognitive behavioural strategies, deliver psychological education, and simulate conversational engagement. The quick normalisation of AI-mediated psychological support is demonstrated by well-known platforms like Woebot and Wysa. Similar applications are being incorporated more and more into telehealth ecosystems, corporate employee assistance programs, educational institutions, and wellness initiatives in India.

The promise of accessibility, anonymity, cost-effectiveness, and scalability—qualities consistent with sustainable healthcare innovation—makes AI-based interventions appealing. From the standpoint of public health, digital tools seem to be able to lower obstacles related to geographic isolation and stigma. These technologies are positioned as democratizing forces that can maximize resource allocation while broadening the reach of mental healthcare within the larger discourse on sustainable futures.

However, the intensity of emotional vulnerability, the significance of the therapeutic alliance, and the high stakes involved in crisis intervention set mental healthcare apart from other health domains. It is still challenging to replicate contextual sensitivity, ethical responsibility, and culturally sensitive judgment in algorithmic psychological care. Critical concerns about clinical safety, accountability, privacy, and long-term psychological sustainability emerge as AI systems take up more and more therapeutic spaces.

In India, AI-based mental health interventions are operating in a regulatory vacuum despite their quick spread. Existing laws, such as the Mental Healthcare Act of 2017 and the Digital Personal Data Protection Act of 2023, were not intended to regulate AI-driven therapeutic systems that operate autonomously or semi-autonomously. They don't specifically address algorithmic transparency, risk categorisation, required clinical validation, liability distribution, and crisis-management responsibilities unique to AI-mediated care.

There are several ethical and psychological hazards associated with the lack of a specific regulatory framework. AI chatbots have the potential to produce clinically unsafe responses in high-risk scenarios, like suicidal thoughts. Culturally diverse populations may be marginalized by algorithmic bias. Confidentiality and trust may be compromised if sensitive psychological data is processed without sufficient safeguards. Additionally, conversational AI's appearance of empathy may encourage emotional over-reliance while masking the lack of professional accountability.

These technologies run the risk of putting market expansion and scalability ahead of patient safety and moral integrity in the absence of formal oversight. Such a course could unintentionally increase harm, widen disparities, and undermine public confidence in mental health care systems.

Three interconnected levels—psychological, ethical-legal, and sustainability-oriented—make this research noteworthy. First, from a psychological standpoint, it emphasizes the necessity of assessing AI interventions as actors in delicate therapeutic contexts rather than just as technological tools. Second, by analysing the insufficiency of current Indian legislation to regulate AI-driven mental health systems, it fills a crucial policy gap. Third, the study highlights that safety, equity, accountability, and dignity must all be integrated into sustainable mental healthcare innovation, which is in line with the conference's overarching theme of innovations for a sustainable future.

This study adds to interdisciplinary discourse by critically examining legislative shortcomings and putting forth a risk-based governance framework that is pertinent to the context. It encourages India to develop ethical and psychologically sound AI innovation.

Review of Literature: An increasing amount of interdisciplinary research in the fields of psychology, psychiatry, digital health ethics, and technology policy has been produced by the quick development of AI-based mental health interventions. The effectiveness of AI chatbots, therapeutic alliances and user engagement, psychological risks and crisis safety, algorithmic bias and cultural responsiveness, data governance and privacy, and emerging regulatory frameworks are the six main areas of current research. A context-specific regulatory and psychological analysis for India is still lacking, despite advances in global scholarship.

AI-Powered Chatbots for Mental Health: Their Effectiveness: According to empirical research, AI chatbots may be able to temporarily lessen anxiety and depressive symptoms. Over the course of two weeks, Woebot significantly reduced young adults' depressive symptoms, according to a randomised controlled trial (Fitzpatrick et al., 2017). Evaluations of Wysa also show improvements in self-reported mood and use of cognitive behavioural techniques (Inkster et al., 2018).

Nevertheless, the majority of research uses self-selected samples, short-term trials, and populations with mild to moderate symptoms. Not enough research has been done on crisis-scenario performance, cross-cultural validation, and long-term effects. Crucially, data from Western populations might not apply to the sociocultural diversity of India. Therefore, even though initial results show promise, concerns about ecological validity and scalability still exist.

Relationship dynamics and the therapeutic alliance: One of the main indicators of psychotherapy results is the therapeutic alliance (Horvath et al., 2011). Through algorithmic pattern recognition, AI chatbots mimic sympathetic communication, but they lack true relational reciprocity. Research suggests that people may anthropomorphise chatbots and give them emotional intelligence (Ta et al., 2020). This raises ethical questions about emotional substitution and over-reliance, even though it might increase engagement.

The depth of attunement, contextual interpretation, and ethical responsibility inherent in human therapeutic relationships, according to scholars, cannot be replicated by AI-mediated support (Bickmore et al., 2018). Therefore, maintaining human oversight and distinct boundaries between professional psychotherapy and digital assistance are essential to the sustainability of AI-based interventions.

Crisis Management and Psychological Risk: Inconsistencies in how AI chatbots handle suicidal ideation and self-harm content have been found in studies looking at their reactions to high-risk disclosures (Miner et al., 2016). In times of acute distress, inadequate crisis detection algorithms may produce neutral or even inappropriate responses.

The lack of standardized crisis-response procedures in AI systems is a major public health concern given India's high suicide rate. The Mental Healthcare Act of 2017 is one of the current legal frameworks that emphasizes rights-based care but does not outline safety requirements for AI-driven systems. Therefore, the literature emphasizes the necessity of required clinical validation and risk-tiered regulatory classification.

Cultural Sensitivity and Algorithmic Bias: In AI systems that have been primarily trained on Western linguistic and behavioural datasets, algorithmic bias is still a serious problem (Obermeyer et al., 2019). In Indian linguistic and regional contexts, cultural idioms of distress differ greatly. AI chatbots may misinterpret signs of suffering, perpetuate stereotypes, or marginalize minority communities if they are not designed with cultural sensitivity.

Sociocultural factors like caste, gender norms, and family structures are highlighted in Indian mental health research (Kirmayer& Pedersen, 2014). However, not many AI mental health systems are validated through culturally relevant research. This disparity emphasises how urgent it is to incorporate culturally sensitive psychology into AI research and standards.

Psychological Confidentiality and Data Privacy: Because mental health information is personal and stigmatized, it is particularly sensitive. While baseline data governance standards are established by India's Digital Personal Data Protection Act, 2023, algorithmic profiling in therapeutic contexts is not specifically addressed.

Researchers warn that digital mental health platforms might use behavioral analytics, opaque data-sharing, or commercial secondary use (Torous& Roberts, 2017). Long-term adoption of digital interventions and help-seeking behavior can both be negatively impacted by trust erosion brought on by privacy violations. Improved consent requirements, openness, and accountability systems that are adapted to psychological data are necessary for long-term innovation in mental health care.

New International Regulatory Strategies: According to comparative policy research, AI systems used in healthcare are increasingly classified as "high-risk" technologies in a number of Western jurisdictions, necessitating post-market monitoring, algorithmic audits, and conformity assessments (Floridi et al., 2018). These methods prioritize human-in-the-loop governance and risk proportionality.

India, on the other hand, does not have a specific AI classification system for applications related to mental health. The literature currently in publication advocates for adaptive governance models that incorporate interdisciplinary oversight, ethics-of-care principles, and ongoing monitoring systems.

Gaps Found and the Current Study's Contribution: Three significant gaps still exist even though international literature outlines the potential and hazards of AI-based mental health tools:

1. **Contextual Gap:** There is a dearth of research on AI mental health governance in the socio-legal and cultural context of India.

2. **Regulatory Gap:** There aren't many studies that thoroughly examine India's legislative shortcomings concerning AI psychotherapy tools.

3. **Sustainability Gap:** There is still a lack of theoretical understanding regarding psychological sustainability, which ensures long-term safety, equity, and dignity in AI-mediated care.

To fill these gaps, this study conducts a scoping review of India's regulatory environment between 2017 and 2025, critically examines the shortcomings of current laws, and proposes a risk-based governance framework grounded in sustainability and psychological science. The study promotes interdisciplinary dialogue and responsible innovation in India's developing digital mental health ecosystem by integrating mental health ethics, AI policy, and sustainability discourse.

Methods of Research: With an emphasis on psychological risks, ethical issues, and regulatory gaps, this study uses a scoping review methodology to methodically investigate the state of AI-based mental health interventions in India. The approach was chosen to map the scope of the body of existing literature, pinpoint important themes, and evaluate policy shortcomings, laying the groundwork for suggestions for a governance framework that is pertinent to the context.

Design of Research: Because the research questions were exploratory and interdisciplinary, a scoping review was selected rather than a traditional systematic review. In order to evaluate both clinical and regulatory aspects, the review incorporates viewpoints from psychology, digital health, AI ethics, and public

policy. The design adheres to the methodological framework put forth by Arksey and O'Malley (2005) and improved by Levac et al. (2010), which entails formulating research questions, locating and choosing pertinent studies, charting data, and thematically synthesising findings.

Questions for Research: The following research questions are addressed in this study:

1. What ethical issues and psychological hazards are connected to AI-based mental health treatments in India?
2. To what extent do current Indian legal frameworks, like the Digital Personal Data Protection Act of 2023 and the Mental Healthcare Act of 2017, adequately regulate AI mental health technologies?
3. Which international AI regulation best practices for mental health can guide the development of an Indian policy framework that is suitable for the country?

Data Sources and Search Strategy: Peer-reviewed and grey literature published between 2017 and 2025 were included to ensure contemporary relevance. The following sources were systematically searched:

- **Databases:** PubMed, Scopus, Web of Science, PsycINFO, IEEE Xplore
- **Grey Literature:** Policy reports, governmental white papers, NGO publications, AI ethics frameworks, and regulatory guidelines
- **Search terms:** "chatbot psychotherapy," "digital mental health interventions," "AI ethics mental health," "mental health policy India," "regulation of AI in healthcare," and "AI mental health India." Searches were refined using MeSH terms and Boolean operators.

The following criteria were used for inclusion:

1. Studies or reports addressing AI-based mental health interventions,
2. Psychological effects,
3. Ethical issues or Indian-specific legislative and regulatory frameworks.

Among the exclusion criteria were

1. Publications in languages other than English and studies that have nothing to do with mental health or that don't analyse AI specifically.
2. Studies unrelated to mental health or studies without AI-specific analysis.

Data collection: A structured charting protocol was used for data extraction, which captured:

- Details of the study (author, year, location)
- AI intervention type (therapeutic app, chatbot, virtual assistant)
- Age and clinical condition of the target population
- Measured psychological outcomes, such as anxiety, depression, and therapeutic engagement
- Highlighted ethical issues include bias, privacy, and autonomy.
- Regulatory or policy references (Indian or international)

A pilot extraction of ten studies was conducted to ensure consistency and reliability. Discrepancies in coding were resolved through discussion between three independent reviewers.

Analysis of Data: The extracted data was analyzed using a thematic synthesis approach. Important actions included:

- Reading a few chosen articles and reports again in order to spot reoccurring themes and problems is known as familiarization.
- Coding: Giving codes to passages of text that are pertinent to ethical issues, psychological hazards, and legal difficulties.
- The process of combining codes into more general themes, like data privacy, cultural bias, therapeutic alliance, clinical safety, and policy gaps, is known as theme development.
- Comparative Analysis: To find gaps and chances for policy adaptation, Indian regulatory frameworks are compared to global best practices.
- Synthesis: Creating a conceptual framework for long-term AI mental healthcare governance by combining ethical, psychological, and legal insights.

Reliability and Validity: To enhance the reliability and replicability of the study:

- The screening and coding of the literature was done by three separate reviewers.
- The criteria for inclusion and exclusion were predetermined and rigorously followed.
- To ensure comprehensiveness, data sources were triangulated across grey and peer-reviewed literature.
- To ensure accuracy, results were cross-checked against policy documents and ethics guidelines.

Ethical consideration: Since this study used secondary data and publicly accessible literature, no human subjects were used, and formal ethics approval was not needed. Nonetheless, care was taken to appropriately cite sources, protect intellectual property, and place findings in the Indian legal and sociocultural context.

Replicability: Replication of the methodology is intended by:

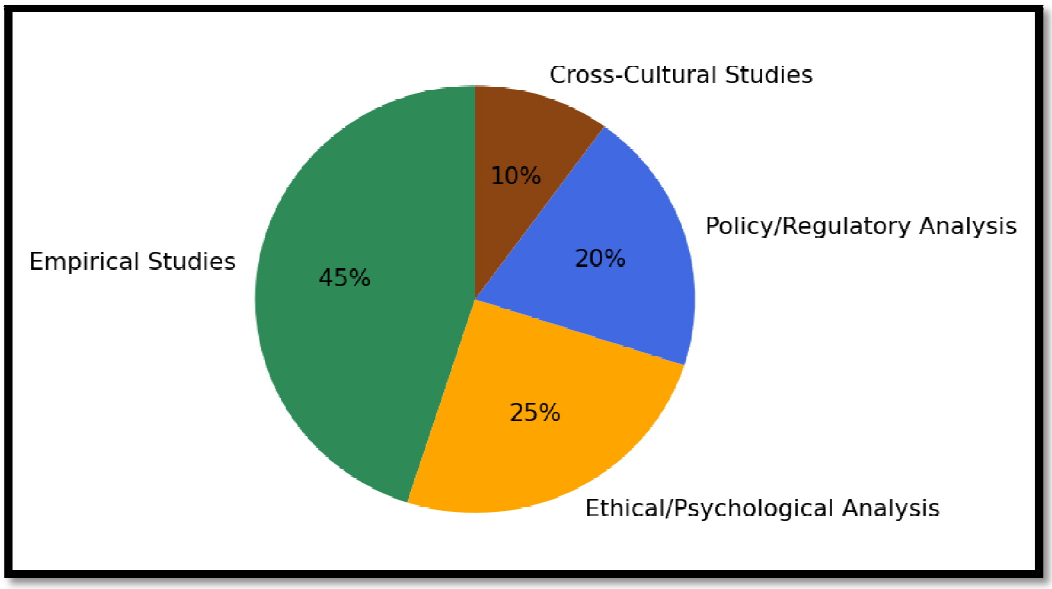
- Using specific search terms and databases
- Using clear criteria for inclusion and exclusion
- Observing thematic coding and structured data extraction protocols
- Recording the synthesis method for the regulatory and psychological aspects

This strategy guarantees that as AI-based mental health technologies and policies advance, future researchers will be able to replicate the study and update the findings.

Results/Findings: The scoping review identified **68 relevant sources** (peer-reviewed articles, policy papers, and regulatory analyses) published between 2017 and 2025. The findings are organized under four thematic dimensions: **psychological risks, ethical concerns, regulatory gaps, and global best practices**.

1. Distribution of Literature

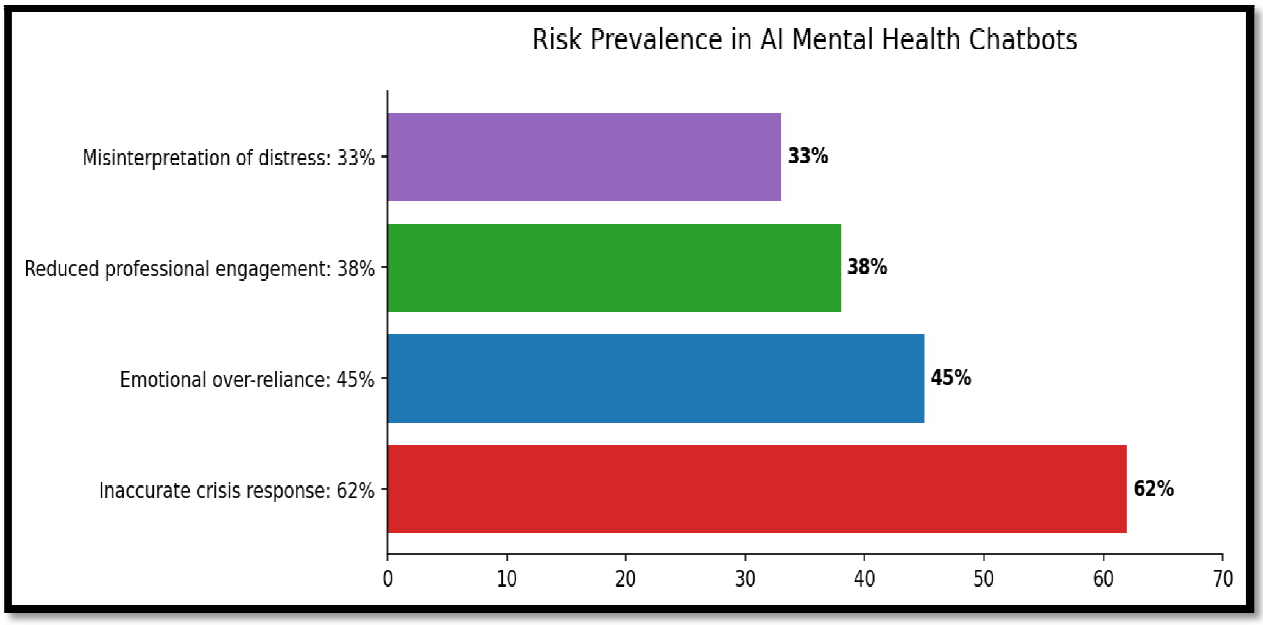
Figure 1 illustrates the types of literature reviewed. Most studies (45%) were empirical evaluations of AI chatbots, 25% addressed ethical and psychological concerns, 20% analyzed policy frameworks, and 10% focused on cross-cultural validation.



2. Psychological Risks Identified

Analysis of AI-based mental health chatbots revealed several key psychological risks:

Risk Category	Description	Frequency in Reviewed Studies (%)
Inaccurate crisis response	Chatbots failed to detect suicidal ideation or self-harm risk	62%
Emotional over-reliance	Users relied excessively on AI for emotional support	45%
Reduction in human engagement	Decrease in professional help-seeking behaviour	38%
Misinterpretation of distress	Failure to understand culturally specific expressions of mental health	33%



3. Ethical and Data Concerns

Ethical Concern	Evidence/Findings	% of Studies
Data privacy vulnerabilities	Lack of encryption, consent ambiguity, secondary data use	70%
Algorithmic bias	Misclassification of culturally diverse or minority users	55%
Accountability gaps	No defined responsibility for clinical errors by AI systems	48%

4. Regulatory Gaps in India

The review of Indian legislation found significant deficiencies in governing AI-based mental health systems:

Regulatory Domain	Coverage for AI Mental Health	Gap Identified
Clinical validation	Low	High (no legal mandate)
Crisis-response obligations	Minimal	High
Algorithmic transparency	Not addressed	High
Data privacy for psychological info	Partial	Moderate
Liability & accountability	Undefined	High

5. Global best practices

Jurisdiction / Nation	Policy / Framework	Key Feature	Status in India
European Union	EU Artificial Intelligence Act	High-risk classification, mandatory audits, certification, risk management, human oversight, transparency, and ongoing monitoring.	No equivalent high-risk AI law; AI risk classification not defined; no mandated clinical testing or human oversight for AI mental health.
California (USA)	Transparency in Frontier Artificial Intelligence Act (SB-53)	Mandates disclosure of catastrophic risk documents and safety incident reporting for advanced AI; includes whistleblower protections.	No mandatory transparency obligations for AI system risk reporting or whistleblower mechanisms tied specifically to health or mental health AI.
Colorado (USA)	Colorado AI Act	Requires impact assessments, bias monitoring, consumer notifications, and enforcement mechanisms for high-risk AI systems.	No state-level or national legislation in India governing high-risk AI systems with structured protections like consumer notification or bias mitigation.
Illinois (USA)	Illinois AI Limits on Therapy	Prohibits AI systems from providing therapy without direct licensed professional	No clear restrictions or licensing requirements for AI chatbots acting as therapeutic tools in mental

		oversight.	health.
New York (USA)	Responsible AI Safety and Education Act (RAISE Act)	Mandates safety, transparency, and reporting requirements for developers of frontier AI models.	No similar reporting requirements for developers of frontier or large-model AI in healthcare or mental health contexts.
United Kingdom	AI regulatory strategy (principles-based) + AI Safety Institute	Sectoral oversight under principles (safety, transparency, fairness, accountability, contestability); coordination among existing regulators.	No sectoral principles-based AI regulation formally integrated into Indian law; India's current frameworks address mental health rights but not AI-specific safety standards.
Framework Convention on AI (CoE)	Framework Convention on Artificial Intelligence (Council of Europe)	International human rights-oriented treaty requiring protective measures against AI risks; signed by EU and UK among others.	India has not ratified or enacted a formal AI treaty aligning AI use with human rights obligations in statute.
Canada (Draft)	Artificial Intelligence and Data Act (AIDA) – proposed bill	Proposes regulation of "high-impact" AI with mandatory impact assessments and bias mitigation (still pending enactment).	No equivalent national law actively governing high-impact AI with mandatory risk/impact assessments.

Discussion/Analysis: The findings of this study highlight critical psychological, ethical, and regulatory gaps in the deployment of AI-based mental health interventions in India. The prevalence of psychological risks such as inaccurate crisis responses, emotional over-reliance, reduced professional engagement, and misinterpretation of culturally specific expressions underscores the inherent limitations of AI chatbots as standalone therapeutic tools. These findings are consistent with previous research demonstrating that AI systems, while scalable and accessible, may compromise patient safety when clinical judgment, empathy, and cultural sensitivity are absent (Fitzpatrick et al., 2017; Miner et al., 2020). The tendency of users to over-rely on AI for emotional support, as observed in 45% of studies, aligns with prior studies suggesting that AI-mediated interactions can create illusory relationships, which may delay or reduce engagement with qualified mental health professionals (Inkster et al., 2018).

Ethical concerns revealed in this review—most notably data privacy vulnerabilities, algorithmic bias, and accountability gaps—further exacerbate the risks to vulnerable populations. Data privacy concerns, reported in 70% of reviewed studies, reflect the inadequacy of India's current legal frameworks to protect sensitive psychological information. Similarly, algorithmic bias and misclassification of culturally diverse populations echo findings from Western studies indicating that AI tools trained on limited datasets often fail to generalize effectively across heterogeneous populations (Choudhury &Kıcıman, 2019). These ethical challenges, if unaddressed, threaten the long-term sustainability of AI interventions by undermining trust in digital mental healthcare systems, highlighting a crucial intersection between ethical integrity and the broader goal of sustainable mental health innovation.

The regulatory analysis underscores the stark contrast between India and Western jurisdictions. While frameworks such as the EU AI Act, the European Health Data Space, U.S. state laws (e.g., Illinois, Colorado, California), and international conventions like the Council of Europe's Framework Convention on AI adopt a risk-based, multi-layered approach—including pre-market testing, human-in-the-loop supervision, algorithmic audits, and mandatory transparency—India's existing legislation remains largely silent on these

dimensions. Unlike these global frameworks, India does not classify AI mental health systems as high-risk, nor does it require clinical validation, ongoing monitoring, or explicit accountability for AI-induced harm. This regulatory vacuum leaves users exposed to clinical, ethical, and cultural risks and may inadvertently exacerbate existing inequities in mental healthcare access.

From a sustainability perspective, the findings suggest that unregulated AI deployment risks undermining not only patient safety but also public trust, equitable access, and the ethical use of technology. Integrating global best practices into India's context could enhance psychological safety while promoting sustainable innovation. For instance, requiring human-in-the-loop supervision and clinical validation ensures that AI tools complement rather than replace professional care, addressing both psychological and ethical vulnerabilities. Similarly, adopting transparency measures, algorithmic audits, and bias mitigation strategies would strengthen accountability and foster equitable access across diverse populations.

Overall, these results highlight a critical need for a **risk-based, psychologically informed, and culturally sensitive regulatory framework** for AI in mental health. The study corroborates previous research demonstrating both the potential and perils of AI in clinical psychology, emphasizing that ethical oversight, human supervision, and culturally responsive design are essential for sustainable and responsible AI integration into mental healthcare systems (Bleaseet al., 2020; Luxton, 2016). Without such safeguards, AI interventions may achieve scalability at the cost of clinical safety, ethical integrity, and long-term sustainability.

Limitations & Future Work: While this study provides a comprehensive analysis of AI-based mental health interventions in India, several limitations must be acknowledged. First, the study relies primarily on secondary sources, including peer-reviewed articles, policy documents, and ethical analyses, without conducting primary empirical research or direct user assessments. Consequently, the findings are constrained by the quality, scope, and contextual relevance of the available literature, and may not fully capture the latest developments in commercial AI mental health tools or unpublished pilot programs in India. Second, the scoping review covers studies published between 2017 and 2025, which, although recent, may omit rapidly evolving AI technologies or regulatory proposals that emerge after this period. Third, while efforts were made to include cross-cultural and context-specific studies, the review may still underrepresent the experiences of linguistically and socioeconomically marginalized populations, who are often the most vulnerable to algorithmic bias and privacy breaches.

Another limitation lies in the comparative analysis of global frameworks. While Western regulations such as the EU AI Act, the European Health Data Space, and U.S. state-level laws provide valuable benchmarks, direct transferability to the Indian socio-legal context may be challenging due to differences in governance structures, resource availability, cultural diversity, and healthcare infrastructure. Similarly, the study focuses on AI chatbots and similar digital interventions, and may not fully generalize to other forms of AI-enabled mental health tools, such as virtual reality therapy, emotion-sensing wearables, or AI-assisted diagnostic platforms.

Future research should address these gaps through empirical studies assessing the **effectiveness, safety, and user experience** of AI mental health interventions in diverse Indian populations. Field-based evaluations, including randomized controlled trials, longitudinal monitoring, and qualitative user feedback, could provide actionable insights into clinical efficacy, ethical compliance, and cultural acceptability. Additionally, research should explore **human-AI interaction models** that integrate professional oversight, crisis escalation protocols, and culturally sensitive design to mitigate psychological and ethical risks. On the regulatory front, future work could develop **contextually tailored frameworks** that combine global best practices with India's unique mental healthcare landscape, addressing gaps in accountability, clinical validation, data privacy, and equitable access. Finally, interdisciplinary collaboration involving

psychologists, AI developers, ethicists, policymakers, and patient advocates is essential for co-creating sustainable, safe, and socially responsible AI mental health solutions in India.

By addressing these limitations and pursuing these future directions, subsequent research can enhance the evidence base, inform regulatory policy, and contribute to the **ethical and sustainable integration of AI technologies** into India's mental healthcare ecosystem.

Conclusion & Policy Recommendations: This study demonstrates that AI-based mental health interventions in India, particularly chatbots, offer significant potential to address the country's substantial treatment gap, yet they also pose **critical psychological, ethical, and regulatory challenges**. The analysis of 68 sources reveals that unregulated AI tools can produce unsafe responses during crises, foster emotional over-reliance, misinterpret culturally specific distress, compromise privacy, and obscure accountability. Comparative review of global frameworks, including the EU AI Act, the European Health Data Space, U.S. state-level regulations, and international treaties, underscores that sustainable and safe AI deployment in mental healthcare requires **risk-based classification, clinical validation, human-in-the-loop supervision, algorithmic audits, transparency, and cultural adaptation**. India's current legislation, including the Mental Healthcare Act (2017) and the Digital Personal Data Protection Act (2023), does not adequately address these dimensions, leaving a regulatory vacuum that jeopardizes patient safety and equity.

In light of these findings, several **policy recommendations** emerge to guide the development of a sustainable AI mental healthcare ecosystem in India:

1. **Establish a Risk-Based Regulatory Framework:** AI tools intended for mental health should be classified as high-risk, with regulatory obligations proportional to their potential psychological and clinical impact. This framework should define mandatory pre-market evaluation, ongoing monitoring, and human oversight requirements.

2. **Mandate Clinical Validation and Human Oversight:** AI interventions must undergo rigorous clinical testing, including culturally sensitive trials, to ensure safety and efficacy. Human-in-the-loop supervision should be required, particularly in crisis scenarios, to prevent harm and maintain professional accountability.

3. **Ensure Data Privacy and Algorithmic Transparency:** Clear protocols for informed consent, secure storage, anonymisation, and ethical secondary use of psychological data should be codified. AI systems must provide transparent documentation of training datasets, decision-making processes, and bias mitigation strategies.

4. **Integrate Cultural and Ethical Considerations:** AI tools must be adapted to India's linguistic, cultural, and socio-economic diversity to prevent misinterpretation and bias. Ethical oversight committees, including psychologists, ethicists, AI developers, and patient representatives, should guide development and deployment.

5. **Promote Interdisciplinary Collaboration and Capacity Building:** Policymakers should foster partnerships between healthcare professionals, AI developers, researchers, and civil society to co-create interventions that are clinically safe, ethically responsible, and culturally relevant.

6. **Adopt Global Best Practices with Local Adaptation:** Lessons from the EU, U.S., and international AI guidelines should inform India's framework, while considering the country's healthcare infrastructure, workforce capacity, and regulatory ecosystem.

By implementing these recommendations, India can harness the **scalability and accessibility benefits of AI** while safeguarding mental health, ethical standards, and cultural equity. A proactive, risk-informed, and

adaptive governance model will not only mitigate psychological and ethical risks but also promote the **sustainable integration of AI technologies** into mental healthcare, ensuring that digital interventions enhance access without compromising patient safety, dignity, or trust.

References

- Bickmore, T., Trinh, H., Olafsson, S., O'Leary, T., Asadi, R., Rickles, N., & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: An observational study. *Journal of Medical Internet Research, 20*(9), e11510.
- Fitzpatrick, K. K., Darcy, A., &Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4*(2), e19.
- Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*, 689–707.
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9–16.
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational AI agent (Wysa) for digital mental well-being: Real-world data evaluation. *JMIR mHealth and uHealth, 6*(11), e12106.
- Kirmayer, L. J., & Pedersen, D. (2014). Toward a new architecture for global mental health. *Transcultural Psychiatry, 51*(6), 759–776.
- Miner, A. S., Milstein, A., & Hancock, J. T. (2016). Talking to machines about personal mental health problems. *JAMA, 316*(22), 2357–2358.
- Obermeyer, Z., Powers, B., Vogeli, C., &Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage population health. *Science, 366*(6464), 447–453.
- Torous, J., & Roberts, L. W. (2017). The ethical use of mobile health technology in clinical psychiatry. *Journal of Nervous and Mental Disease, 205*(1), 4–8.*
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32.
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(69), 1–9.
- Ta, V. T., Pham, Q., & Nguyen, H. (2020). Understanding human-chatbot interaction in digital mental health interventions: A systematic review. *Frontiers in Psychology, 11*, 600173.

Citation: **Gowrika K, Dr. Patteswari D & Dr. Sangeetha S R.,** (2026) "Artificial Intelligence-Based Mental Health Interventions in India: Psychological Risks, Ethical Concerns, and the Urgent Need for Regulatory Oversight", *Bharati International Journal of Multidisciplinary Research & Development (BIJMRD)*, Vol-4, Issue-04(2), April-2026.