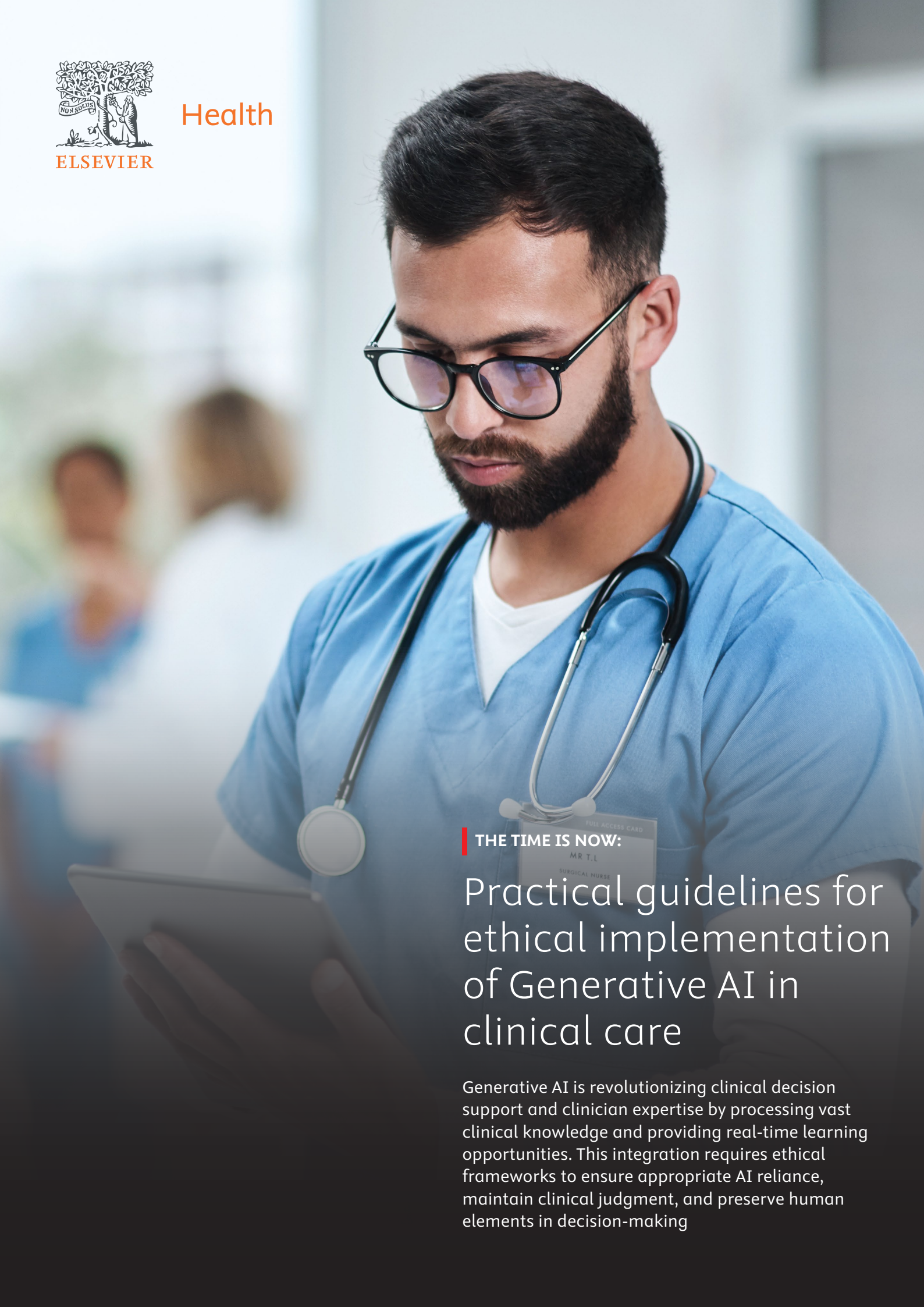


ELSEVIER Health

THE TIME IS NOW:

Practical guidelines for ethical implementation of Generative AI in clinical care

Generative AI is revolutionizing clinical decision support and clinician expertise by processing vast clinical knowledge and providing real-time learning opportunities. This integration requires ethical frameworks to ensure appropriate AI reliance, maintain clinical judgment, and preserve human elements in decision-making

Contributors:

Dr. Arun Khemariya, MBBS, MBA, Senior Clinical Specialist, Elsevier India
Dr. Puja Sakhuja, MD, Chairperson, Director, Head, and Professor, Department of Pathology, Govind Ballabh Pant Hospital, New Delhi
Dr. Edna D'Souza, PhD, Cluster Manager (Healthcare, Life Sciences & Chemicals),British Deputy High Commission, Mumbai
Prof. Sanjay Dhir, Professor, Department of Management Studies, IIT Delhi
H. Sidhnath, Nehal Jain and Navya, Department of Management Studies, IIT Delhi

Editor:

Arun Khemariya

Notes on this Paper:

This paper is reflective of documented and anticipated ethical aspects of Generative AI as of May 20, 2025. Due to the rapid pace of development, use, and harms of Generative AI, we want to acknowledge that this is an inherently dynamic paper, subject to changes in the future.

Executive summary

This white paper presents practical guidelines for the ethical implementation of Generative AI (GAI) in clinical care in a way that's ethical and effective. It's all about tapping into the incredible potential of AI to transform healthcare while keeping patient safety, fairness, and high ethical standards front and centre. As healthcare increasingly integrates GAI, it is crucial to establish a framework that balances innovation with responsibility.

The paper underscores the importance of a balanced approach to integrating Generative AI in clinical settings, highlighting that while AI has the potential enhance patient out-comes and operational efficiency, it must be implemented ethically to mitigate risks and uphold the integrity of healthcare.

Key actions for ethical implementation of generative AI include transparently communicating GAI operations and decision-making processes to build trust, establishing clear accountability for GAI outcomes with human oversight, testing to mitigate biases and ensure equitable treatment of diverse populations. Implementing continuous monitoring and regular evaluation policies for GAI systems, along with robust governance frameworks to protect patient rights and personal information, is essential. Additionally, inter-disciplinary collaboration among clinicians, data scientists, ethicists, and policymakers is crucial for navigating the ethical complexities of healthcare delivery.

This paper is a call for balance. GAI can do amazing things, but only if we roll it out thoughtfully. Stick to these guidelines, and we can embrace the good while keeping risks in check and preserving what makes healthcare human.

Table of contents

Introduction

"Can machines think?", the quintessential question asked by Dr Alan Turing in the marquee paper Computing machinery and intelligence (Turing 1950). In November 2022, a chatbot developed using the Large Language Model GPT-3 brought GAI tools into the spotlight of public awareness.

Artificial intelligence (AI) is revolutionizing economies and sectors across the globe, offering India a unique opportunity to tackle systemic challenges and foster inclusive growth. The promise of artificial intelligence (AI) is vast and multifaceted, with the potential to revolutionize numerous aspects of human life. It promises to drive a new era of creativity and productivity while also posing critical questions for humanity. The introduction of AI into healthcare holds the promise of addressing major challenges in the field, such as diagnosis and screening, therapeutics, preventive treatments, clinical decision-making, public health surveillance, complex data analysis, and predicting disease outcomes. This range of applications is expected to expand continuously in the future.

Origins of Generative Artificial Intelligence

Generative AI is a ground breaking advancement in machine learning, representing a major step forward in AI's capacity to comprehend and engage with intricate data patterns. Generative AI refers to algorithms that can synthesize new content based on the data they have been trained on (Johnson et al. 2024).

Generative Artificial Intelligence (AI) has a rich history spanning several decades, marked by key advancements. In 1952, Arthur Samuel developed the first machine learning algorithm for a checkers game and coined the term "machine learning" (Samuel 1959).

The concept of generating new data began with Frank Rosenblatt's Perceptron in 1957, an early trainable neural network (Rosenblatt 1958). ELIZA, created in 1961, was one of the first examples of generative AI, interacting with users in natural language and crafting empathetic responses, paving the way for modern chatbots. The 1980s saw further progress with generative models like the Boltzmann Machine, which aimed to understand data distributions and generate samples similar to the original dataset (Burns, Nemelka & Arora 2024).

Technological milestones

The evolution of modern generative AI began in the 1980s-1990s with Hopfield networks and Boltzmann machines, which laid the foundation for more advanced models. In the 2000s, deep belief networks marked a major breakthrough in generative modeling. The field was revolutionized in 2014 when Ian Goodfellow introduced Generative Adversarial Networks (GANs), enabling the creation of highly realistic images and data. In 2017, Google's transformer model further advanced natural language processing, significantly enhancing generative AI capabilities (Burns, Nemelka & Arora, 2024).

The emergence of Neural Networks and Generative

The progress in neural networks coincided with GAI evolution, particularly with the introduction of backpropagation, a method used to train deep neural networks. This era also saw the development of more sophisticated AI models capable of learning and generating data.

A breakthrough in GAI came in 2014 with the introduction of Generative Adversarial Networks (GANs) by Ian Goodfellow and his colleagues. GANs consist of two neural networks, a generator and a discriminator, that work together to produce realistic data. This innovation enabled the creation of high-quality images, videos, and audio that appeared authentic. The origins of modern generative AI can be traced back to the early days of artificial intelligence, focusing on probabilistic models and neural networks. Initial efforts were theoretical due to computational limitations and lack of large-scale data (Diro et al., 2025)

Large language models foundation

Large Language Models (LLMs) are specialized AI systems trained on extensive text datasets to understand and generate human-like content, capable of answering questions, summarizing texts, and engaging in open-domain conversations with emergent proper-ties like logical reasoning. Beyond text, generative models can also create images, videos, audio, and computer code (Johnson et al., 2024).

Foundation Models (FMs), a subset of generative AI, perform multiple tasks, including generating text, images, and audio, and serve as the backbone of generative technologies. They encompass LLMs for text, Vision Language Models (VLMs) for visual and textual data, and multimodal models for various data types, facilitating the development of autonomous agents like AutoGPT, which can independently pursue user-defined goals (Liu et al., 2025).

In the early 2020s, the emergence of transformer-based deep neural networks, particularly large language models (LLMs) like GPT-3 and GPT-4, significantly advanced generative AI capabilities. These models can generate coherent and contextually relevant text, images, and videos from natural language prompts, marking a major shift from earlier deep learning techniques in terms of scale and potential impact (Pack, Barrett & Escalante, 2024).

Trained on extensive datasets with billions of parameters, LLMs require regulatory over-sight to address issues of interpretability, fairness, and unintended consequences. They utilize tokens words, subwords, or charactersas the basic units for processing language, making tokenization a crucial aspect of natural language processing (NLP). However, tokenization currently lacks regulation in the healthcare sector.

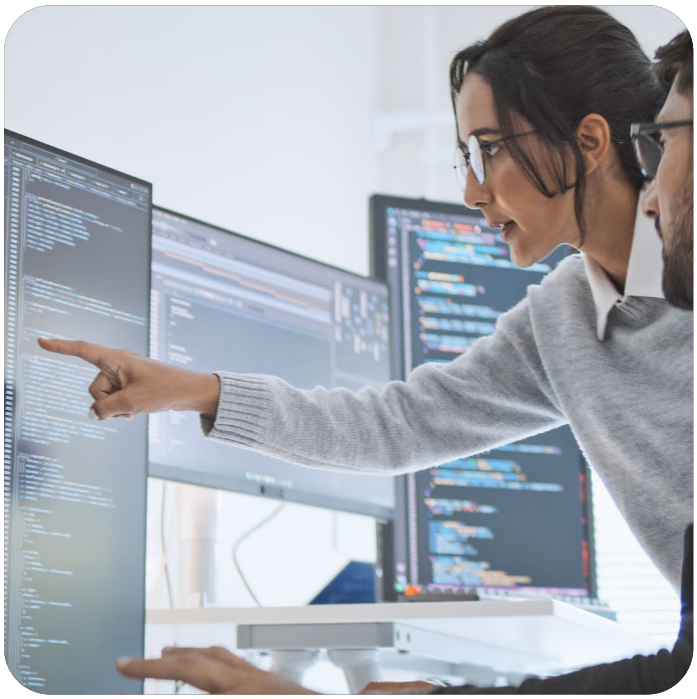
The Indian Outlook on GAI

Potential impact on economy and workforce transformation

The integration of Generative AI (GAI) across various sectors is expected to significantly influence India's economy and workforce. New job roles, such as AI specialists and ethics consultants, will emerge alongside the necessity for widespread upskilling in AI literacy and data analytics. Automation will drive significant productivity gains across sectors like finance and IT, while also augmenting decision-making in fields like healthcare and law. Industry-specific transformations will accelerate, with GAI revolutionizing software development, healthcare diagnostics, and media content creation.

According to McKinsey & Company's report, "The Economic Potential of Generative AI: The Next Productivity Frontier," generative AI could contribute $2.6 trillion to $4.4 trillion annually across various industries, with healthcare as a significant beneficiary. Approximately 75% of this value is focused on four main areas: customer operations, marketing and sales, software engineering, and research and development (McKinsey & Company 2023) report by EY forecasts that generative AI will transform 38 million jobs in India by 2030, increasing economic productivity by over 2.6%, contributing as much as $438 billion to India's GDP by 2030, driven by enhanced productivity, innovation, and the creation of high-value jobs (Ernst & Young 2024).

Similar changes are anticipated globally, particularly in healthcare administration, radiology, and patient engagement. The workforce will need to adapt by acquiring skills in AI governance, data analysis, and human-AI collaboration. Policymakers must address ethical concerns, implement regulations, and revamp education systems to prepare future generations with necessary skills. This technology can enhance diagnostics, automate administrative tasks, and personalize treatments, improving efficiency and reducing costs. While some roles will change due to automation, new opportunities will emerge in AI model training, implementation, and oversight, requiring a workforce with new skills.



Regulatory compliance for Generative Artificial Intelligence (GAI)

AI-driven health care faces significant compliance challenges, primarily concerning data privacy, algorithmic bias, and cross-border regulatory alignment. The Digital Personal Data Protection (DPDP) Act, 2023 mandates lawful data processing, prior authorization, and strict data protection measures to minimize unauthorized access and data breaches in India. Similarly, the Health Insurance Portability and Account-ability Act in the USA, the EU Medical Device Regulation & AI Act, GDPR in the EU, and the World Health Organization AI Ethics Guidelines are critical compliance frameworks that regulate safety, privacy, and ethical AI deployment. These regulations enforce transparency, validation, data protection, and bias mitigation, ensuring alignment with global healthcare governance standards (Ministry of Electronics and Information Technology 2023).

However, cross-border challenges related to mutual recognition, agreements, licensing, and certification create hurdles in integrating AI solutions into existing healthcare frame-works. In addition, algorithmic biases embedded in AI models can raise ethical concerns, necessitating rigorous validation and governance mechanisms to ensure fairness and patient safety (Palaniappan et al. 2024).

Compliance with Indian data protection laws

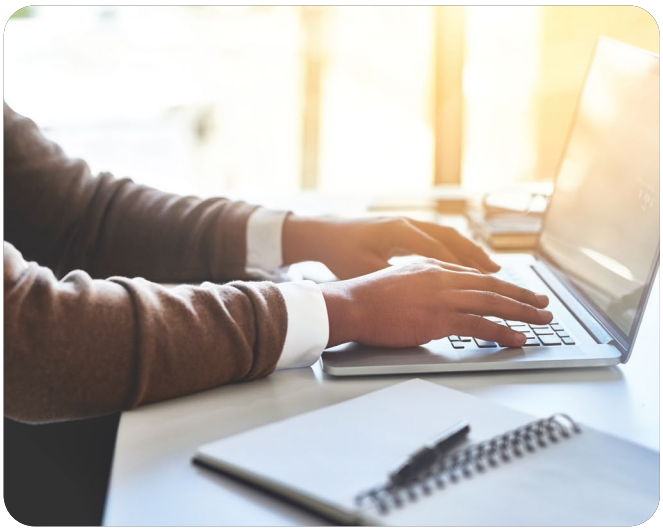
The Digital Personal Data Protection (DPDP) Act, 2023, directly impacts Generative AI (GAI) implementations. GAI systems process personal data, requiring compliance with DPDP regulations, including obtaining explicit user consent before data collection or processing. The Act enforces data minimization, ensuring only essential personal data is used, and purpose limitation, restricting data use to its stated intent. It grants individuals rights to access, rectify, or erase their data, requiring organizations to implement compliance mechanisms. Strong security measures must protect personal data, and impact assessments may be necessary before deploying GAI. Additionally, cross-border data transfers must adhere to DPDP provisions. Additionally, the government is investing in AI research and development through initiatives like the National AI Portal, fostering innovation and collaboration among academia, industry, and government.

As of now, India does not have a specific law directly governing the use of Generative AI (GAI). However, several existing legal frameworks apply to AI technologies, particularly in sectors like healthcare. Apart from the Digital Personal Data Protection Act, the Information Technology Act, 2000 addresses cyber crimes and data protection, while intellectual property laws may impact the ownership of AI-generated content. Additionally, bodies like the Medical Council of India and the Insurance Regulatory and Development Authority of India may influence AI use in their fields. Although ethical guidelines and a national AI strategy are being developed, dedicated AI legislation, including for GAI, is expected as the technology advances.

Generative AI (GAI) in Healthcare

Generative AI (GAI) is shaping up to be a game-changer for healthcare, it holds significant promise for addressing healthcare challenges with a focus on patient-centric solutions. GAI serves as a helping hand to doctors and nurses, giving them sharp insights to make better calls. Its transformative potential has sparked critical discussions about its role, ranging from serving as an invaluable assistant to medical personnel to the possibility of AI systems taking over entire clinical departments. This paradigm shift invites a reimagining of healthcare delivery, where technology and human expertise collaborate to ele-vate patient outcomes and redefine the future of medical practice.

Raising a query like, “What’s the latest on Epilepsy management?”-and the AI doesn’t provide a generic response. Instead, it delves into the most recent studies, guidelines, and data to deliver a clear and concise explanation tailored to your needs. Whether you're curious about the side effects of a new drug or how a rare condition manifests in children compared to adults, feel free to ask!



Generative AI: Clinical use cases

This exploration of clinical use cases highlights the profound impact of Generative AI on modern medicine, showcasing its potential to not only augment healthcare delivery but also redefine the future of patient care.

Clinical decision-making

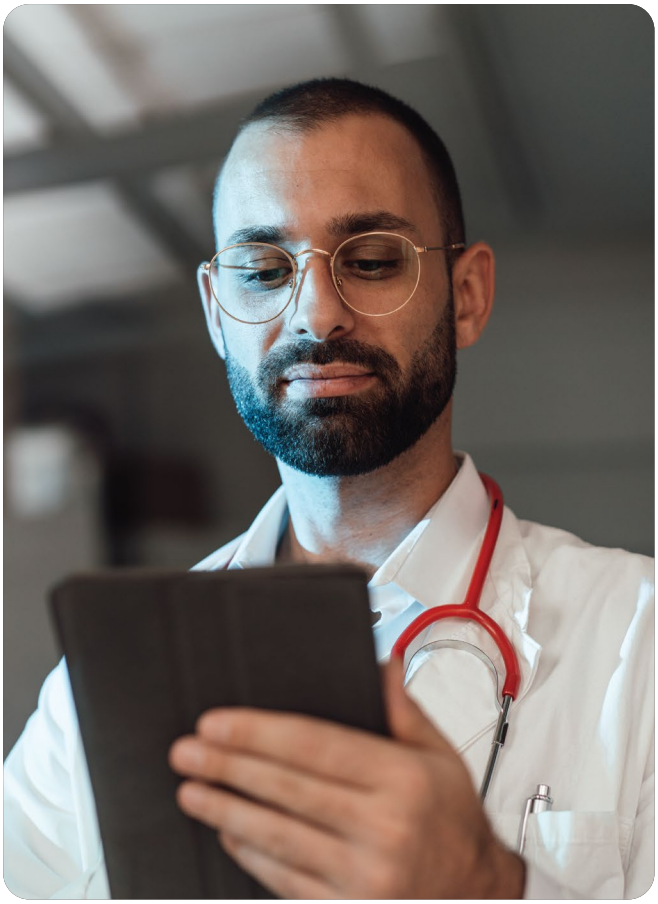
Generative AI (GAI) can significantly enhance the major aspects of clinical decision-making i.e. scientific evidence, clinical judgment and patient involvement.

Evidence-based medicine enhancement with GAI: By rapidly analyzing and synthesizing vast amounts of medical literature to identify the most reliable evidence for clinical decision-making. A significant amount of patient data can be found in published literature from scientific journals. Research articles often include data from electronic health records (EHRS), medical imaging, laboratory results, clinical trials, and genomic studies. . It also aids in developing and regularly updating clinical guidelines to reflect current best practices, ensuring standardized, high-quality care. Additionally, studies on wearable devices, patient surveys, and social determinants of health are frequently published. However, raw patient data can come from these sources directly as well. GAI generated summaries customized to your expertise and interests, help you navigate through information overload, saves time and allows you to focus on patient care.

Patient involvement: GAI personalizes treatment recommendations by considering individual patient profiles, aligning care with the most effective evidence-based options. This personalized approach not only helps patients understand their conditions better but also encourages them to engage in shared decision-making with their healthcare providers.

Point of care use: Clinicians can harness the power of Generative AI (GAI) at the point of care to enhance patient interactions and optimize clinical workflows. During OPD visits, GAI facilitates more meaningful conversations by leveraging patient history and treatment progress, ensuring that all critical health aspects are addressed. This leads to highly personalized and empathetic care.

In the in-patient setting, if a patient's condition is not improving as expected, GAI can identify potential issues and suggest corrective actions. By analyzing factors such as medication adherence, lifestyle choices, and other variables, GAI helps clinicians make necessary adjustments to the treatment plan. This proactive approach guarantees that patients receive the highest standard of care, ensuring the best possible outcomes.



Although clinical decision-making is the most significant application of Generative AI, there are additional potential uses that can enhance various aspects of clinical care.

Multimodal imaging and diagnostics

Generative AI enhances medical imaging by spotting early indicators of disease by analyzing large volumes of medical images to detect subtle patterns and changes that might be missed by human eyes. This also helps GAI to forecast disease progression and suggest potential diagnoses. GAI can create synthetic images for educational purposes and automates the segmentation of abnormalities, thereby improving accuracy and efficiency in both diagnosis and treatment.

Interactive Q&A systems

If you're a clinician with a million things on your mind about patients, lab reports, class schedule, budgets and you've got a quick question about a tricky medical condition, a new treatment or some latest research. GAI is like having a super-smart, always-on-call colleague who never gets tired of your questions

Public health policy

GAI can help make more data driven policy decisions by analyzing patterns and health metrics. Multiple scenarios can be processed and their impact on community health assessed in short periods of time, allowing policymakers to evaluate the effectiveness and impact of different strategies before implementation.

Clinical reasoning

GAI can generate and respond to hypothetical patient cases. Diverse, realistic patient scenarios with varying levels of complexity helps clinicians refine their diagnostic reasoning, clinical decision-making, and treatment planning skills. The scope to introduce progressive symptom development, varying lab results, imaging findings, and patient histories require real-time decision-making, prompting clinicians to rule out possibilities. It functions like a virtual patient that evolves according to the clinician's choices, making the learning experience both immersive and practical.

Clinical research

Generative AI has the capability to simulate clinical trial scenarios, refine trial protocols, and create potential outcomes scenario, helping researchers in making informed, data-driven decisions. It can also generate synthetic patient data based on real data, protecting patient privacy and expanding data sets.

Health economic modelling

With the appropriate human supervision GAI can significantly enhance health economic modeling by automating data processing, improving predictive analysis, and optimizing decision-making. It accelerates literature reviews, extracts relevant cost and clinical data from various sources, and integrates structured and unstructured information to refine economic models Through explainable AI and automated sensitivity analyses, GAI improves transparency, making economic models more accessible to policymakers and stakeholders.

Macroeconomics of GAI Implementation

The macroeconomic implications of integrating General Artificial Intelligence (GAI) into healthcare necessitate a comprehensive cost-benefit analysis within the complex economic framework of a healthcare system. The initial deployment of GAI requires substantial capital investment in advanced computational infrastructure, workforce reskilling, and robust data ecosystems. However, these upfront costs are counterbalanced by significant long-term economic gains, driven by enhanced diagnostic precision, minimized medical errors, and optimized resource allocation across the healthcare sector.

GAI driven innovations such as real-time predictive analytics, hyper-personalized treatment models, and fully automated clinical work flows amplify healthcare productivity and elevate patient outcomes, yielding substantial reductions in aggregate healthcare expenditures. These advancements facilitate a structural shift from a reactive, cost-intensive healthcare paradigm to a proactive, value-based system.

This transition not only strengthens the healthcare sector's economic resilience but also generates positive externalities, spurring growth in allied industries like advanced medical robotics, health data analytics, and next-generation health information systems. (Meslamani 2023)

From a macroeconomic perspective, GAI adoption in healthcare enhances labor productivity by reallocating human capital from routine diagnostic and administrative tasks to higher value roles, while simultaneously reducing systemic inefficiencies. This shift contributes to GDP growth, improves fiscal sustainability by lowering public healthcare spending, and fosters innovation-driven employment in technology and sup-port sectors. However, policymakers must address potential frictional costs, such as short term labor displacement or unequal access to GAI-enabled care, to ensure inclusive economic benefits across the population.

Significance of Ethically Implementing GAI in Clinical Care

The Indian Council of Medical Research (ICMR) released ethical guidelines for the use of AI in biomedical research and healthcare to clarify various issues related to liability, transparency, accountability, and oversight. While highlighting AI's potential to enhance individual and population health, the guidelines also warn about the ethical, legal, and social challenges arising from the complexities of algorithmic learning (Indian Council of Medical Research 2023).

The WHO's guidance on AI regulation discusses the following major factors that influence AI adoption in healthcare: Transparency, Risk management, External validation, Data quality standards, Equity and Human oversight (World Health Organization 2023). These principles also align with the HTI-1 rule's emphasis on "FAVES" (Fair, Appropriate, Valid, Effective, Safe) AI deployment in U.S. healthcare settings.

Here are several key points that highlight the significance of ethically implementing GAI in clinical care, emphasizing its potential to improve patient outcomes while ensuring trust and accountability.

Consistent and reliable knowledge for all stakeholders

GAI can ensure that clinicians, administrators, and patients have access to uniform and reliable information.

Unified Source of Information for Clinicians

GAI can serve as a unified source of information for clinicians by compiling and synthesizing up-to-date clinical knowledge.

Reduction of bias:

Reduction of Bias: Ethically integrating GAI involves actively addressing and mitigating biases in AI algorithms, which can lead to more equitable healthcare delivery across di-verse populations.

Evaluating effectiveness and facilitating improvement

Data collected from GAI systems can be used to assess the effectiveness of clinical interventions and promote process enhancements.

Empowering Patients and Caregivers

Ethical GAI fosters trust and transparency, which can enhance patient engagement. Ethical implementation prioritizes patient safety through rigorous testing, validation, and adherence to clinical standards.

Data Privacy and Security

Ethical GAI ensures compliance with regulations like DPDP or GDPR, safeguarding patient data through encryption, anonymization, and strict access controls.

Challenges in GAI Implementation

Ethical Challenges

Autonomy Gaps: Enhancing patient engagement during GAI-assisted consultations is essential to ensure that individuals are active participants in their healthcare journey. Patients would always prefer a Clinician as their source of information.

Data Privacy Vulnerabilities: Risk of protected health information being used improperly/without consent during AI training and deployment. AI model trained on patient records to enhance predictive diagnostics, leading to legal and ethical issues.

Algorithmic Bias and Health Disparities: Challenges of AI systems amplifying existing healthcare inequities by propagating historical bias. This would lead to misdiagnosis or suboptimal treatment for specific groups.

Unclear Clinical Validation Standards: Healthcare is highly specific to its environment. A GAI model that works effectively in one setting may not perform well in another due to differences in patient demographics, clinical processes, and data collection practices, all of which can greatly influence its accuracy and reliability.

Lack of Explainability: The "black box" problem of ambiguity in sources of GAI responses. Many GAI models provide accurate but opaque predictions, clinicians can't understand the reasoning behind AI recommendation. Lack of explainability undermines trust and can lead to liability issues in case of medical errors.

System Issues: GAI systems work only as accurately as the sources behind the system, lack of proper monitoring and timely updates create a tendency for GAI to perform poorly over time.

Operational Challenges

Resource Burden: Developing and integrating GAI systems into existing healthcare infrastructure can be costly. Expenses for software security, data management, hardware upgrades, staff training and integration with existing systems. Organizations must allocate resources effectively to support the successful integration of AI technologies. Hospital administrators may resist AI adoption due to the high initial investment and uncertainty about the return on investment.

Administrative Pushback & Resistance to Change: AI can seem complex and intimidating, especially to those without a technical background. This can create anxiety and a reluctance to adopt new technologies. Many healthcare professionals value the human connection with their patients. They may fear that GAI will erode the patient-provider relationship. There may also be hesitance stemming from how GAI adoption could impact their current positions and responsibilities.

Transparency & Accountability: Organizations need to establish clear lines of accountability to address potential errors or conflicts resulting from AI recommendations ensuring that AI tools complement and enhance the work of healthcare professionals rather than disrupt established practices.

Regulatory Compliance: Most countries are still developing AI-specific regulations, leading to uncertainty in compliance requirements. Adhering to evolving regulatory frameworks governing AI use in healthcare can be complex and resource intense. Organizations must stay informed about legal requirements and ensure that their GAI implementations comply with relevant laws and guidelines.

Training and Education: Healthcare professionals may require training to effectively use GAI tools and understand their implications. Organizations must invest in education and training programs to equip staff with the necessary skills and knowledge. Without proper training, professionals may struggle to interpret AI-generated data or understand how to use AI tools effectively.

Continuous Monitoring and Evaluation: Ongoing assessment of GAI systems is necessary to ensure their effectiveness and safety. Organizations must establish processes for monitoring AI performance, identifying potential issues, and making necessary adjustments.

Clinical Workflow Disruption: Workflow and team effort forms the foundation of a successful healthcare setup. A potential Prevents the implementation pitfall
where technically sound GAI systems fail because they don't integrate smoothly with clinical workflows.

Technology Burnout: High workloads, complex systems, lack of training, constant alerts, and increased administrative burdens. Poor integration of technologies and frequent technical issues also contribute to stress and fatigue. Addressing these factors through adequate training, streamlined workflows, reduced administrative tasks, and robust technical support can help mitigate these issues.

Pillars of Ethical Implementation

Ethical pillars help to establish a clear set of principles that guide decision-making and behaviour within an organization, particularly in contexts where technology, data, and human interactions intersect. Inclusiveness, fairness, security, and transparency are fundamental principles emphasized in widely recognized responsible AI frameworks (Indian Council of Medical Research 2023).
The following ethical pillars provide a structured framework to guide responsible decision-making and behaviour in the integration of Generative AI (GAI) in healthcare, ensuring transparency, accountability, and patient centered care.

Transparency

Transparency in GAI is vital to build trust, by openly communicating how GAI systems function and the rationale behind their decisions, stakeholders can promote fair and effective use of these technologies, ultimately resulting in improved patient outcomes and increased public confidence. Black boxes remain one of the major hindrances in achieving transparency in GAI recommendations (Box 1)

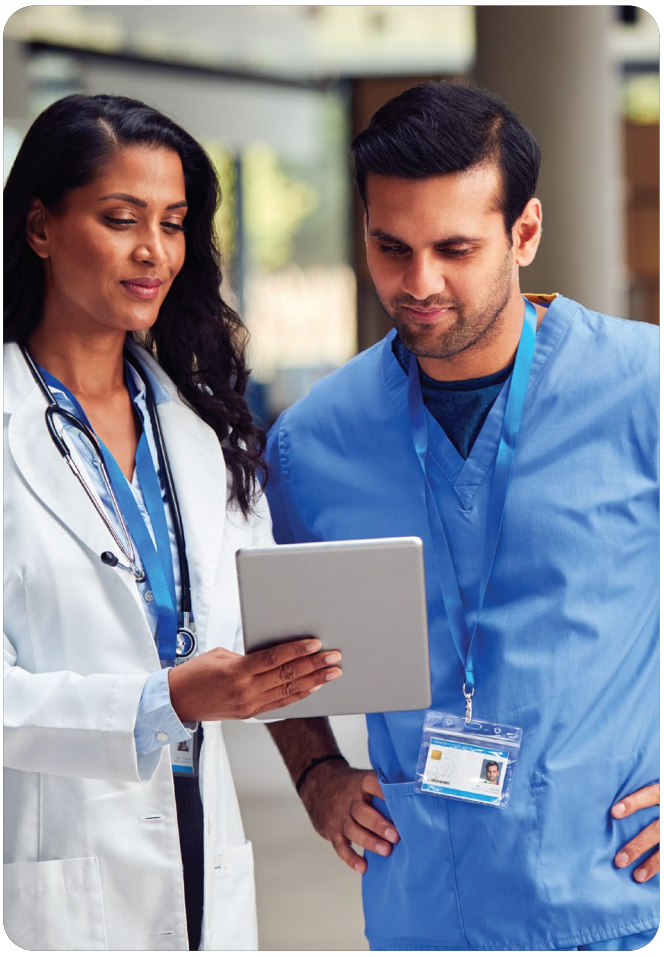
Black box nature of AI models

Generative AI models have a nonlinear structure and a higher accuracy rate, making it unclear what information in the training data allows them to reach different conclusions, referred to as black boxes. This raises concerns about model interpretability, bias, accountability, and patient safety, limiting the adoption of AI in clinical settings (Wani et al. 2024).

Generative AI systems in healthcare are often referred to as "black boxes" due to the difficulty in understanding their decision-making processes. This lack of transparency can hinder acceptance by healthcare practitioners (Gupta et al. 2025).

Adaptive Safeguards

The terms "glass box" or "white box" refer to approaches that emphasize transparency and interpretability in various fields, such as artificial intelligence (AI), software testing, and decision-making models. These approaches allow users to observe the inner workings of a system, rather than treating it as a "black box" where the internal processes are hidden. This transparency helps users trust the system and ensures that the decisions are fair and unbiased (Robson & Baek 2024) Turning the black box into a glass box in healthcare involves enhancing transparency, conducting detailed evaluations of AI components, and implementing systems that provide clear insights into their operations. This approach can build trust and improve the effectiveness of AI in healthcare.



Explainability

A key challenge of generative AI, like that of artificial intelligence in general, is that while it offers significant benefits, there are associated costs. It is inevitable that the system may occasionally deliver inaccurate information. Explainability of AI-generated recommendations is a critical aspect of ensuring trust and transparency. As AI systems increasingly assist in clinical decision-making, it becomes essential that healthcare professionals and patients understand how these recommendations are derived.

Explainable AI (XAI)

Explainable Artificial Intelligence (XAI) refers to a sub-domain of AI techniques that provide a set of approaches or algorithms to produce results that are easy to explain and intuitive for all concerned stakeholders (Keleko et al 2023)

XAI improves explainability and transparency, providing a clear understanding of AIdriven decisions. Techniques such as Shapley Additive Explanations and Local Interpretable Model-Agnostic Explanations offer insights into decision-making by high lighting key contributing factors and validating AI-generated recommendations (Parvathaneni et al., 2022)

Adaptive Safeguards

Ensuring the explainability of AI-generated recommendations requires transparency about the algorithms and data utilized (Box 2), along with clear interpretations of the key factors influencing decisions. By clarifying the connections between inputs and outputs combined with user-friendly explanations, the reliability of AI recommendations improves.

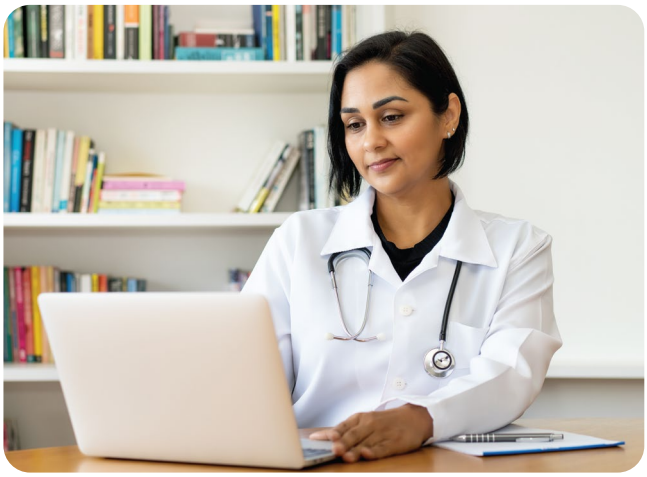
Accountability

The rapid adoption of GAI underscores the urgency of ensuring Accountability and defining Responsibility for AI-Driven Decisions in Healthcare addressing ethical, legal, and operational challenges associated with AI-driven decisions. The closed-source nature of many AI models prevents understanding of AI-generated outputs, complicating accountability, key considerations include algorithmic transparency, liability distribution among stakeholders, and the evolving responsibilities of clinicians and healthcare organizations in supervising AI systems (Hasan et al 2024).

Adaptive Safeguards

Policymakers should define the responsibilities of developers, operators, and users, and establish mechanisms for holding entities accountable for AI errors, misuse, or unintended consequences (Gupta et al 2024)

Clinicians remain accountable for final decisions, necessitating competence in interpreting AI outputs. Healthcare organizations must ensure adequate training and system validation, design interfaces that clearly distinguish AI-generated content and develop clear override protocols when clinicians disagree with AI recommendations. Audits can help demonstrate institutional accountability if an AI system fails to perform as expected or causes harm



Fairness and Equity

Ensuring fairness and equity in AI recommendations is a critical aspect of responsible AI deployment in healthcare. It is crucial to prevent discrimination, build trust and comply with regulations (Menye et al 2024).

Adaptive Safeguards

Avoiding biases, promoting intersectional fairness, and implementing continuous monitoring and human-in-the-loop approaches can help healthcare organizations develop AI systems that deliver equitable and inclusive care. Robust validation of training data to uncover any underrepresentation, creating oversight committees that include diverse community representation is crucial. Such committee can conduct regular bias audits with documented corrective actions and assessing AI performance before implementation.

Collaborative Governance

Ethical AI deployment in health care depends on private-sector innovation and public-sector oversight. While private enterprises provide advanced AI models and scalability, public institutions ensure transparency, accountability, and patient safety through regulatory policies (Organizational Governance of Emerging Technologies: AI Adoption in Healthcare, 2023).

Adaptive Safeguards

Collaboration among technology companies, healthcare providers, advocacy groups, and regulators through interdisciplinary research is crucial for refining AI ethics and establishing transparent, equitable governance frameworks aligned with ethical standards.

Effective Governance Demands Partnerships Between:

- Clinical informaticians converting workflow requirements into technical specifications.
- Data scientists ensuring algorithmic fairness across gender, race, and socioeconomic variables
- Legal teams navigating evolving regulatory landscapes
- Ethicists addressing dilemmas like AI's role in end-of-life decisions
- Regulatory and Policy Considerations
- Global Standards Convergence

Data Privacy

Data privacy is crucial for Generative Artificial Intelligence (GAI) and organizations must protect sensitive information from unauthorized access and breaches. Ethical and security concerns arise from the extensive use of sensitive data, necessitating secure communication and processing. Appropriate balance between data access for development with maintaining confidentiality is essential, as differential privacy techniques can compromise data quality.

Adaptive Safeguards

GAI implementation in healthcare makes it essential to have stringent data governance and privacy measures in place. This encompasses effective data anonymization prior to AI training, well-defined data retention policies with regular audits, clear documentation of patient data usage, protocols for managing incidental findings, and strict adherence to regulations such as DPDP. AI modules can implement various strategies for data anonymization and secure storage, and these practices should be guided by established frameworks and best practices.

Clinical Oversight (Human-in-the-Loop)

While AI excels at data processing, clinician's role is paramount in delivering empathic communication when delivering AI-informed diagnoses. Doctor patient relationship is built on trust, empathy, and personalized care, which are inherently human qualities (Fuehrer et al. 2024).

The Evolving Role of Clinicians in AI Oversight

- Clinical Validation of Algorithmic Outputs
- Verify AI recommendations against patient-specific factors (e.g., comorbidities, social determinants)
- Identify potential data artifacts or biases (e.g., underrepresentation of minority groups in training sets)
- Escalate discrepancies through appropriate channels.
- Radiologists at massachusetts General Hospital, for example, rejected 18% of AI-generated tumor classifications in 2024 due to contextual factors absent from the model's training data.
- Maintaining interdepartmental clinical synergy with open discussions
- Clinical judgement is required for ethical deliberation for edge-case scenarios and integration of GAI insights with historical patient narratives

Adaptive Safeguards

Clinical Validation

it is essential to conduct thorough validation of GAI against clinical standards before deployment. Comprehensive documentation outlining AI limitations and suitable use cases should be created, along with continuous monitoring systems to identify any performance drift. Regular quality audits that compare AI outputs with expert clinical judgment are vital, as are established protocols for reporting and addressing adverse events, ensuring patient safety and maintaining trust. Clinicians have a moral responsibility to educate and train other healthcare professionals on how to effectively use AI tools in their practice.

Patient-centered approach

Patients are less likely to trust and engage with healthcare systems that do not prioritize their needs, which can hinder the adoption of GAI technologies. It may also lead to patients feeling neglected or misunderstood, leading to dissatisfaction with the care they receive.

Adaptive Safeguards

Ensure AI implementation addresses actual clinical needs rather than technology-driven priorities. This includes developing clear protocols for obtaining informed consent when using AI in patient care, establishing mechanisms for patients to discuss GAI based recommendations and create accessible explanations of how AI is used in different clinical contexts.

Continuous monitoring

Generative AI (GAI) systems may become outdated and less effective over time, the quality of care provided by GAI systems may decline without regular updates and refinements, negatively impacting patient outcomes. Organizations that do not embrace continuous improvement may also face resistance to change, making it difficult to implement new technologies and innovations.

Adaptive Safeguards

It's imperative for Clinicians to monitor GAI based recommendations in real time and develop a robust feedback mechanism. This involves comparing outputs with current clinical guidelines and expert opinions. Healthcare organisations have the responsibility of keeping detailed records of AI interactions, decisions, and outcomes. The records can be used for future reference, audits, and continuous improvement efforts.

Towards a "Clinically Explainable, Fair, and Responsible GAI involving Clinicians, Experts, and Patients"

As we harness the transformative potential of generative AI, it is imperative that we take decisive action to embed ethical considerations into every facet of its implementation. By translating our principles into tangible practices, we can fully realize the benefits of generative AI while safeguarding the rights and well-being of all stakeholders. We are committed to establishing a comprehensive Responsible AI Framework to guide the development, deployment, and monitoring of AI systems across our organization. This framework will not only protect our stakeholders but also enhance our innovation capabilities and position us as a leader in ethical AI practices.

An ethically sound policy framework is crucial to guide the development and application of GAI technologies in healthcare. As GAI technologies advance and are increasingly used in clinical decision-making, it is vital to establish processes that address accountability in case of errors to ensure safeguarding and protection. Similar to other diagnostic tools, GAI-based solutions cannot be held accountable for their decisions and judgments. Therefore, it is important to assign accountability and responsibility at all stages of GAI development and deployment in healthcare.

A comprehensive approach that includes technical transparency, shared responsibility models, and proactive institutional governance is essential.

Clinicians should act as skilled interpreters of AI insights, integrating them into holistic patient care rather than merely accepting algorithmic outputs. Healthcare organizations need to become leaders in AI governance, promoting interdisciplinary collaboration and maintaining strict oversight. Globally, regulatory bodies are aligning around principles of fairness, safety, and human oversight, but the fast pace of AI innovation requires adaptable frameworks to tackle unforeseen challenges. Ultimately, the successful integration of AI in healthcare depends on preserving the essential human aspects of medicine while leveraging technology to enhance clinical excellence.

The rapid rise of generative AI necessitates swift, ethical action to align systems with core values like transparency, accountability, fairness, and privacy, moving from theory to practice. Ethical implementation demands clear communication of AI operations and limitations to build trust, human oversight for accountability, and rigorous testing to ensure equitable treatment across diverse populations. Robust governance must safeguard data privacy and regulatory compliance, while continuous monitoring adapts systems to evolving standards. Engaging patients and stakeholders, alongside interdisciplinary collaboration among clinicians, data scientists, ethicists, and policymakers, is essential to responsibly navigate AI's ethical complexities in healthcare.

REFERENCES

Burns, B., Nemelka, B., & Arora, A. (2024). Practical implementation of generative artifi-cial intelligence systems in healthcare: A United States perspective. Future healthcare journal, 11(3), 100166. https://doi.org/10.1016/j.fhj.2024.100166

Diro, A., Kaisar, S., Saini, A., Fatima, S., Hiep, P. C., & Erba, F. (2025). Workplace security and privacy implications in the GenAI age: A survey. Journal of Information Security and Applications, 89, 103960.

Doe, J. (2022, August 15). From black box to glass box: The future of AI ethics. Forbes.

Ernst & Young. (2024). The AIdea of India: Generative AI's potential to accelerate India's digital transformation. EY. https://www.ey.com/en_in/insights/ai/generative-ai-india-2025-report

Fuehrer, S., Weil, A., Osterberg, L. G., Zulman, D. M., Meunier, M. R., & Schwartz, R. (2024). Building Authentic Connection in the Patient-Physician Relationship. Journal of primary care & community health, 15, 21501319231225996. https://doi.org/10.1177/21501319231225996

Gupta et al (2025). AI enhanced healthcare: Opportunities, challenges, ethical consid-erations, and future risk. Responsible and Explainable Artificial Intelligence in Healthcare, 127-153.

Gupta, R., Nair, K., Mishra, M., Ibrahim, B., & Bhardwaj, S. (2024). Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda. In-ternational Journal of Information Management Data Insights, 4(1), 100232.

Hasan, S. S., Fury, M. S., Woo, J. J., Kunze, K. N., & Ramkumar, P. N. (2024). Ethical Appli-cation of Generative Artificial Intelligence in Medicine. Arthroscopy: The Journal of Ar-throscopic & Related Surgery.

Indian Council of Medical Resesarch (ICMR). (2023). Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. ICMR.

Keleko, A. T., Kamsu-Foguem, B., Ngouna, R. H., & Tongne, A. (2023). Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI. Advances in Engineering Software, 175, 103339.

Kim, J. Y., Boag, W., Gulamali, F., Hasan, A., Hogg, H. D. J., Lifson, M., ... & Sendak, M. (2023, June). Organizational governance of emerging technologies: AI adoption in healthcare. In proceedings of the 2023 ACM conference on fairness, accountability, and transparency (pp. 1396-1417).

McKinsey & Company. (2023). The economic potential of generative AI: The next produc-tivity frontier. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-fron-tier

Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applica-tions, and challenges. Informatics in Medicine Unlocked, 101587.

Ministry of Electronics and Information Technology. (2023). Digital Personal Data Pro-tection Act, 2023. Government of India. https://www.meity.gov.in/content/digital-per-sonal-data-protection-act-2023/

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated es-say scoring of English language learner writing: Insights into validity and reliability. Com-puters and Education: Artificial Intelligence, 6, 100234.

Parvathaneni et al. (2022). From Blackbox to Explainable AI in healthcare: Existing tools and case studies. In Mobile Information Systems (Vol. 2022, pp. 1–20) [Journal-article]. https://doi.org/10.1155/2022/8167821

Robson, B., & Baek, O. K. (2024). Glass box machine learning for retrospective cohort studies using many patient records. The complex example of bleeding peptic ulcer. Computers in biology and medicine, 173, 108085. https://doi.org/10.1016/j.compbio-med.2024.108085

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

Turing, A. M. (2021). Computing machinery and intelligence (1950).

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.

World Health Organization. (2023). Regulatory considerations on artificial intelligence for health. https://iris.who.int/handle/10665/373421