CitiusTech

Shaping Healthcare Possibilities

Mitigating GenAI bias in healthcare solutions at scale

Hitangi Matalia,
Sr. Solution Architect, Quality & Validation Service, CitiusTech

Rahul Bhadre,
Tech Specialist, Quality & Validation Service, CitiusTech

Abstract

While Generative AI (GenAI) holds immense potential for healthcare, nearly one-third of healthcare professionals see it as a threat to equity and fairness. Biased AI in healthcare can lead to inaccurate diagnoses, inappropriate treatments, and worsen existing health disparities.

This paper explains the importance of a robust framework and covers the key strategies for mitigating bias in GenAI healthcare solutions. Bias testing throughout the development and deployment ensures fairness and equity, protects patient safety, builds trust and acceptance of AI systems. By implementing these strategies, healthcare organizations can harness the power of GenAI to create ethical, equitable, and effective healthcare solutions.

Don't let bias hinder the potential of AI in healthcare. Read this paper to learn about various types of bias and best practices to mitigate them.

PERSPECTIVE PAPER

Introduction

GenAI is believed to have the greatest potential for enhancing clinical efficiency and improving patient engagement and satisfaction.[1] Despite that, nearly one-third of respondents from a recent survey perceive GenAI as a looming threat to equity and fairness.[2] In healthcare, the consequences of such risks are particularly dire without the right strategic approach and tools. For instance, if GenAI models are trained on data that lacks diversity, they may not accurately represent the needs and experiences of all patients. Moreover, concerns about data privacy and security raise ethical dilemmas about the use of patient information in GenAI models. These issues, if left unaddressed, could undermine the very principles of inclusive and accessible healthcare experiences.

Mitigating challenges and concerns around GenAI biases necessitates a rigorous framework for bias testing. It can empower healthcare stakeholders to make more accurate and informed decisions, safeguard patient safety, and foster equity, thereby paving the way for a fairer, digitally enabled continuum of care.

A closer look at the hidden bias in today's healthcare GenAI tools

The impact of GenAI in healthcare extends far beyond individual patient outcomes. It can reshape the entire healthcare value chain, from research and development to patient care and administrative tasks. The benefits include faster eligibility and benefits verification, chatbot support to get tailored guidance for healthcare needs, intelligent content search, the rise of domain-specific models and solutions, and so much more.[3]

While GenAI offers unprecedented innovation opportunities, it's also essential to acknowledge its limitations and pitfalls if organizations aim to maximize its advantages. Nearly 60% of healthcare leaders from a recent survey, regardless of whether they work at a payer, provider, or HST organization, cite accuracy concerns as a key scale-up challenge for GenAI solutions.[1] This challenge is exacerbated by the fact that 94% of healthcare organizations have not yet developed a GenAI strategy - a prerequisite for addressing accuracy concerns, particularly those related to bias.[4]

The impact of this can be profound and far-reaching. Biased algorithms can lead to inaccurate diagnoses, inappropriate treatments, and disparities in care. For example, if a GenAI model is trained on data that disproportionately represents a certain population, it may be more likely to make errors when applied to individuals from other groups.

Such inaccuracies can exacerbate the existing health disparities and undermine trust in AI-powered healthcare tools. Additionally, biased algorithms can perpetuate harmful stereotypes and reinforce discriminatory practices, further marginalizing vulnerable populations.

To navigate these opportunities and challenges, both C-suite executives and GenAI developers must have a strong understanding of the following key types of bias in healthcare AI:

What leads to biases in GenAI healthcare solutions?

Biases in GenAI healthcare solutions are of various types and can arise from several factors. Here is an overview of the same:

Exhibit 1: Beyond innovation - Understanding GenAI bias in healthcare

Unmasking and addressing the bias: Pathways to fairer healthcare solutions

Given that only 18% of organizations have a dedicated council or board to oversee responsible AI governance, and just one-third require technical talent to possess expertise in GenAI, bias testing becomes even more critical for healthcare solutions.[5]

For GenAI healthcare solutions to benefit all patients equitably, it is imperative to prioritize bias testing throughout the development and deployment of these systems.

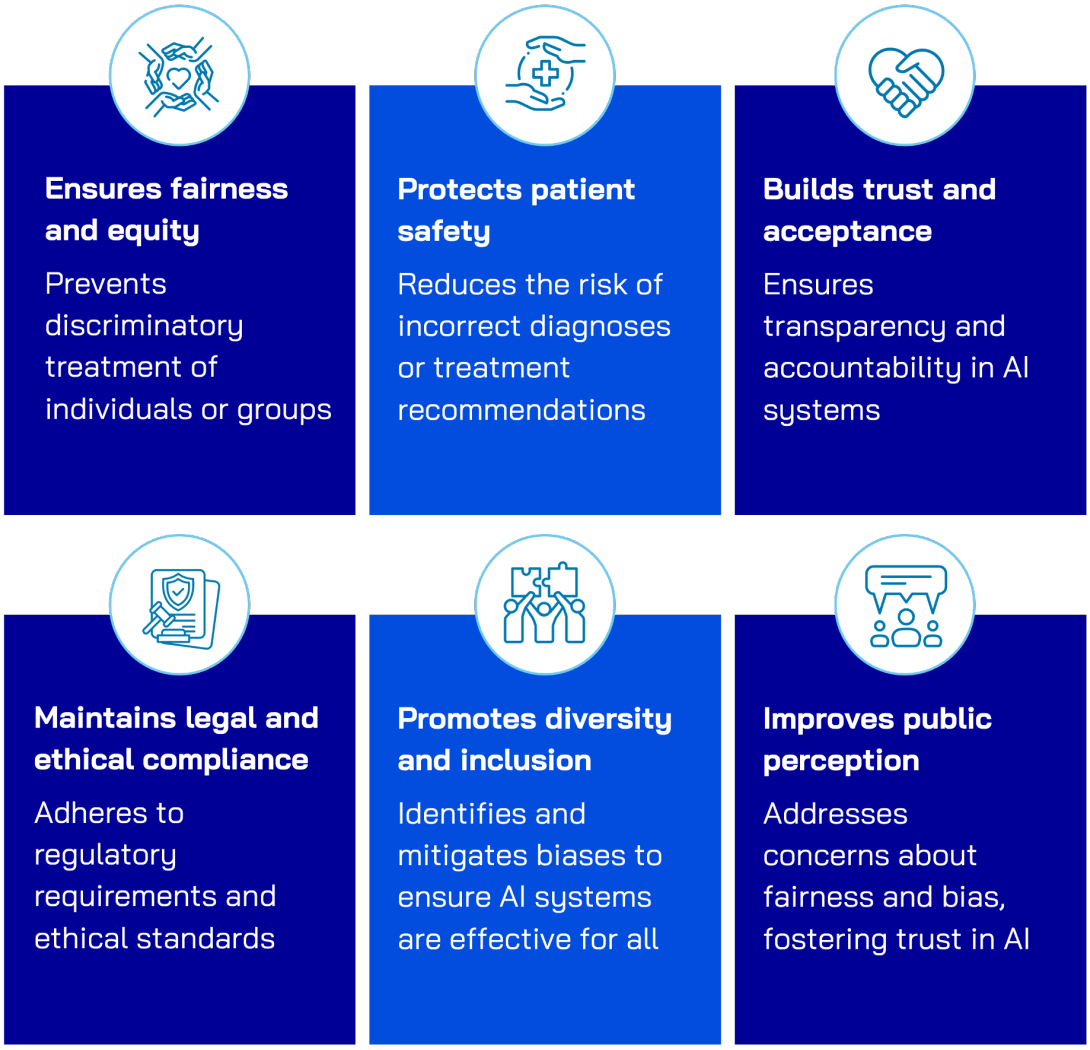

Exhibit 2: Reaching GenAI's blind spot - Why bias testing matters in healthcare innovation

However, optimizing the desired outcomes requires a collaborative approach. Bias mitigation roadmaps should involve software developers, data scientists, domain experts, clinicians, and policymakers working together to ensure ethical and equitable AI implementation. By fostering such a multidisciplinary team, organizations can effectively identify and address biases throughout the AI development lifecycle, from data collection to model deployment.

Key strategies to mitigate bias

The adoption of the following best practices that reinforce quality and trust in AI data models can directly contribute to this strategic approach. Quality and Trust framework solutions are equally essential in this journey.[6] These frameworks provide guidelines and standards for AI development, ensuring that AI systems are transparent, accountable, and trustworthy. Strongly adhering to them can help build AI systems that are not only effective but also ethical and responsible, ultimately promoting healthcare equity and improving patient outcomes for all.

Maintaining data quality and diversity

If a dataset primarily consists of individuals from a single demographic group, the model may exhibit bias towards that group and perform poorly for other populations. Training GenAI models on diverse datasets that accurately reflect the target population can greatly reduce bias that may arise due to underrepresentation or overrepresentation of certain groups.

But before training a model, it is also essential to clean and preprocess the data to remove errors, inconsistencies, and biases. It helps prevent these issues from being propagated into the model's decision-making process. For instance, removing outliers and correcting missing values can improve the accuracy and reliability of the model's predictions. Additionally, detecting and removing biases present in the data can help prevent the model from perpetuating existing inequalities.

Ensuring algorithmic fairness

Identifying and mapping potential sources of bias to specific healthcare workflows, such as Electronic Medical Records (EMRs) or Electronic Health Records (EHRs) is always a best practice. With a thorough understanding of how bias manifests in EMRs and EHRs, we can develop targeted strategies to mitigate its impact. For instance, biased coding practices and algorithms can perpetuate inequalities. Pinpointing critical areas for intervention allows us to address root causes and develop effective mitigation strategies.

In addition to that, it is also important to employ:

- **Fairness metrics:** Employ appropriate fairness metrics to measure bias and evaluate model performance. Common metrics include disparate impact, disparate treatment, and equal opportunity.
- **Bias mitigation techniques:** Implement techniques like adversarial training, fair classification, or reweighting to address bias and ensure equitable outcomes.

Prioritizing transparency and explainability

Model interpretability is essential for understanding how AI models make decisions. The identification of the most important features that contribute to the model's predictions, generating human-readable rules, and using visualizations, can help gain insights into the model's behavior and identify potential biases.

Bias detection tools can also be used to identify and diagnose sources of bias within the model. Statistical analysis and machine learning techniques can be employed to detect patterns of bias in the model's predictions. Additionally, specialized tools and libraries designed for bias detection and mitigation can be used to enhance the process.

Focusing on ethical considerations

Human oversight plays a critical role in preventing unintended consequences and mitigating biases. Ethical review boards can provide oversight and ensure that AI systems are developed and deployed in accordance with ethical principles. Additionally, integrating human experts into the decision-making process can help identify and address potential biases.

Ethical guidelines are also essential for ensuring that AI systems are developed and used responsibly. These guidelines should address key principles such as fairness, accountability, transparency, healthcare compliance regulations, and privacy. Adhering to them ensures that GenAI systems are developed and used in a manner that benefits healthcare stakeholders, both in the short and long term, and avoids harm.

Enabling continuous monitoring and evaluation

Regular assessment is crucial for identifying and addressing emerging bias-related issues. Thus, conducting bias audits, monitoring performance metrics, and collecting feedback from users and stakeholders can provide valuable insights into the model's functioning and identify areas for improvement.

Iterative improvement is another key component of continuous monitoring and evaluation. Regularly updating the training data, refining the model's algorithms and parameters, and incorporating ethical considerations into the development and evaluation process, ensures that the model remains effective, fair, and aligned with ethical principles.

GenAI and healthcare experiences: A balancing act

Bias in healthcare GenAI can exacerbate existing disparities and lead to catastrophic results at scale. However, by implementing effective strategies and fostering a culture of explainable AI, we can ensure that AI-powered healthcare solutions benefit all stakeholders, regardless of where they are in the value chain.

The right GenAI bias testing solutions can identify and address biases early in the development process, preventing them from propagating into the final models. Driven by continuous monitoring and evaluating AI systems for bias, organizations can ensure they are fair, equitable, and reliable.

References

1. Generative AI in healthcare: Adoption trends and what's next | McKinsey
2. What is the future of Generative AI? | McKinsey
3. Five game-changing use cases for Gen AI in healthcare | Forbes
4. Generative AI: the next frontier of healthcare | Forbes
5. Gen AI adoption spikes and starts to generate value | McKinsey
6. Scaling GenAI projects with confidence | CitiusTech

CitiusTech

Shaping Healthcare Possibilities

CitiusTech is a global IT services, consulting, and business solutions enterprise 100% focused on the healthcare and life sciences industry. We enable 140+ enterprises to build a human-first ecosystem that is efficient, effective, and equitable with deep domain expertise and next-gen technology.

With over 8,500 healthcare technology professionals worldwide, CitiusTech powers healthcare digital innovation, business transformation and industry-wide convergence through next-generation technologies, solutions, and products.

www.citiustech.com

Shaping Healthcare Possibilities