

<https://doi.org/10.1038/s44401-024-00009-w>

Environment scan of generative AI infrastructure for clinical and translational science

Check for updates

Betina Idnay¹, Zihan Xu², William G. Adams³, Mohammad Adibuzzaman⁴, Nicholas R. Anderson⁵, Neil Bahroos⁶, Douglas S. Bell⁷, Cody Bumgardner⁸, Thomas Campion^{2,9}, Mario Castro¹⁰, James J. Cimino¹¹, I. Glenn Cohen¹², David Dorr⁴, Peter L. Elkin¹³, Jungwei W. Fan¹⁴, Todd Ferris¹⁵, David J. Foran¹⁶, David Hanauer¹⁷, Mike Hogarth¹⁸, Kun Huang¹⁹, Jayashree Kalpathy-Cramer²⁰, Manoj Kandpal²¹, Niranjan S. Karnik²², Avnish Katoch^{23,24}, Albert M. Lai²⁵, Christophe G. Lambert²⁶, Lang Li²⁷, Christopher Lindsell²⁸, Jinze Liu²⁹, Zhiyong Lu³⁰, Yuan Luo³¹, Peter McGarvey³², Eneida A. Mendonca³³, Parsa Mirhaji³⁴, Shawn Murphy³⁵, John D. Osborne³⁶, Ioannis C. Paschalidis³⁷, Paul A. Harris³⁸, Fred Prior³⁹, Nicholas J. Shaheen⁴⁰, Nawar Shara³², Ida Sim⁴¹, Umberto Tachinardi⁴², Lemuel R. Waitman⁴³, Rosalind J. Wright⁴⁴, Adrian H. Zai⁴⁵, Kai Zheng⁴⁶, Sandra Soo-Jin Lee⁴⁷, Bradley A. Malin³⁸, Karthik Natarajan¹, W. Nicholson Price II⁴⁸, Rui Zhang⁴⁹, Yiye Zhang², Hua Xu^{50,55}✉, Jiang Bian^{51,53,54,55}✉, Chunhua Weng^{1,52,55}✉ & Yifan Peng^{2,9,55}✉

This study reports a comprehensive environmental scan of the generative AI (GenAI) infrastructure in the national network for clinical and translational science across 36 institutions supported by the CTSA Program led by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) at the United States. Key findings indicate a diverse range of institutional strategies, with most organizations in the experimental phase of GenAI deployment. The results underscore the need for a more coordinated approach to GenAI governance, emphasizing collaboration among senior leaders, clinicians, information technology staff, and researchers. Our analysis reveals that 53% of institutions identified data security as a primary concern, followed by lack of clinician trust (50%) and AI bias (44%), which must be addressed to ensure the ethical and effective implementation of GenAI technologies.

The burgeoning advancement of generative AI (GenAI) provides transformative potential for healthcare systems globally. GenAI employs computational models to generate new content based on patterns learned from existing data. These models, exemplified by large language models (LLMs), can produce content across various modalities such as text, images, video, and audio^{1–5}. Its ability to generate human-comprehensible text enabled the exploration of diverse applications in healthcare that involve the sharing and dissemination of expert knowledge, ranging from clinical decision support to patient engagement^{6,7}. Integrating GenAI into healthcare can enhance diagnostic accuracy, personalized treatment plans, and operational efficiencies. For instance, GenAI-driven diagnostic tools can analyze medical images and electronic health records (EHRs) to detect diseases, often surpassing the accuracy of human experts^{8–13}. GenAI applications can streamline administrative processes, reduce clinicians' documentation burden, and enable them to spend more time on direct patient care^{14,15}.

However, implementing GenAI technologies in healthcare has several challenges. Issues such as trustworthiness, data privacy, algorithmic bias, and the need for robust regulatory frameworks are critical considerations that must be addressed to ensure the responsible and effective use of GenAI^{16,17}.

Given these promising advancements and associated challenges, understanding the current institutional infrastructure for implementing GenAI in healthcare is crucial. Various stakeholders (e.g., clinicians, patients, researchers, regulators, industry professionals) have different roles and responsibilities in GenAI implementation, ranging from ensuring patient safety and data security to driving innovation and regulatory compliance, and may hold varying attitudes toward GenAI applications that influence their acceptance and utilization of these technologies. Failure to consider these diverse perspectives may hinder the widespread adoption and effectiveness of GenAI technologies.

A full list of affiliations appears at the end of the paper. ✉e-mail: hua.xu@yale.edu; bianji@iu.edu; cw2384@cumc.columbia.edu; yip4002@med.cornell.edu

Previous studies have examined stakeholder perspectives on AI adoption to some extent. For example, Scott et al.¹⁸ found that while various stakeholders generally had positive attitudes towards AI in healthcare, especially those with direct experience, significant concerns persisted regarding privacy breaches, personal liability, clinician oversight, and the trustworthiness of AI-generated advice. These concerns are reflective of AI technologies in general. Specific to GenAI, Spotnitz et al. surveyed healthcare providers and found that while clinicians were generally positive about using LLMs for assistive roles in clinical tasks, they had concerns about generating false information and propagating training data bias¹⁹.

Despite these insights, there remains a gap in understanding the infrastructure required for GenAI integration in healthcare institutions, particularly from the perspective of institutional leadership. The Clinical and Translational Science Awards (CTSA) Program, funded by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) in the United States (US), supports a nationwide consortium of medical research institutions at the forefront of clinical and translational research and practice²⁰. By examining the GenAI infrastructure within CTSA institutions, we can gain valuable insights into how GenAI is being adopted into cutting-edge research environments and help set benchmarks for the broader healthcare community. Furthermore, understanding the challenges faced by CTSA institutions in this context is crucial for developing strategies that promote fair and accessible GenAI implementation^{8,21}.

In this study, we aim to conduct an environmental scan of the infrastructure for GenAI within CTSA institutions by surveying CTSA leaders to comprehensively understand its current integration status. We also highlight opportunities and challenges in achieving equitable GenAI implementation in healthcare by identifying key stakeholders, governance structures, and ethical considerations. We acknowledge the dual roles that respondents may represent, whether in their capacity as leaders within academic institutions (i.e., CTSA), healthcare systems, or both. Hence, we use the term “healthcare institutions” to encompass the broad range of leadership representation and capture a more complete picture of GenAI integration across research-focused and healthcare-delivery institutions. The insights gained from this study can inform the development of national policies and guidelines to ensure the ethical use of GenAI in healthcare; identifying successful GenAI implementation strategies can serve as best practices for other institutions; highlighting gaps in the current GenAI infrastructure can guide future investments and research priorities; and ultimately, a robust GenAI infrastructure can enhance patient care through more accurate diagnoses, personalized treatments, and efficient healthcare delivery.

Results

The US CTSA network contains over 60 hubs. We sent email invitations to 64 CTSA leaders, each responding on behalf of a unique CTSA site, with 42 confirming participation. Ultimately, we received 36 complete responses, yielding an 85.7% completion rate. Only fully completed responses were included in the analysis, as the six unfinished responses had 0–65% progress and were excluded. The survey questions are available in Supplementary Material A. Of the 36 completed responses, 15 (41.7%) represented only a CTSA, and 21 (58.3%) represented a CTSA and its affiliated hospital.

Stakeholder identification and roles

Figure 1a shows that senior leaders were the most involved in GenAI decision-making (94.4%), followed by information technology (IT) staff, researchers, and physicians. Cochran's Q test revealed significant differences in stakeholder involvement ($Q = 165.9, p < 0.0001$). Post-hoc McNemar tests (see Methods) with Bonferroni correction showed senior and departmental leaders were significantly more involved than business unit leaders, nurses, patients, and community representatives (all corrected $p < 0.0001$; Supplementary Table 1). Nurses were also less engaged than researchers and IT staff (corrected $p < 0.0001$).

We further split our analysis based on whether institutions have formal committees or task forces overseeing GenAI governance to provide insights into how governance models may impact GenAI adoption. 77.8% (28/36) respondents reported having formal committees or task forces overseeing GenAI governance, 19.4% (7/36) did not, and 2.8% (1/36) were unsure. We grouped those without formal committees for analysis to simplify the comparison and focus on clear distinctions between institutions with and without established governance structures. Institutions without formal committees did not involve patients and community representatives as stakeholders in the decision-making and implementation of GenAI (Fig. 1a).

Further, the decision-making process for implementing GenAI (Fig. 1b) was primarily led by cross-functional committees (80.6%), with clinical leadership also playing a key role (50.0%). Institutions without formal committees were led more by clinical leadership. Specific mentions include the dean, CTSA and innovation teams, researchers, and health AI governance committees. Cochran's Q test revealed significant differences in leadership involvement ($Q = 46.8, p < 0.0001$), especially between cross-functional committees and both regulatory bodies and other stakeholders (corrected $p < 0.0001$; Supplementary Table 2).

Decision-making and governance structure

The decision-making process for adopting GenAI in healthcare institutions varied (Fig. 1c). A centralized (top-down) approach was used by 61.1%

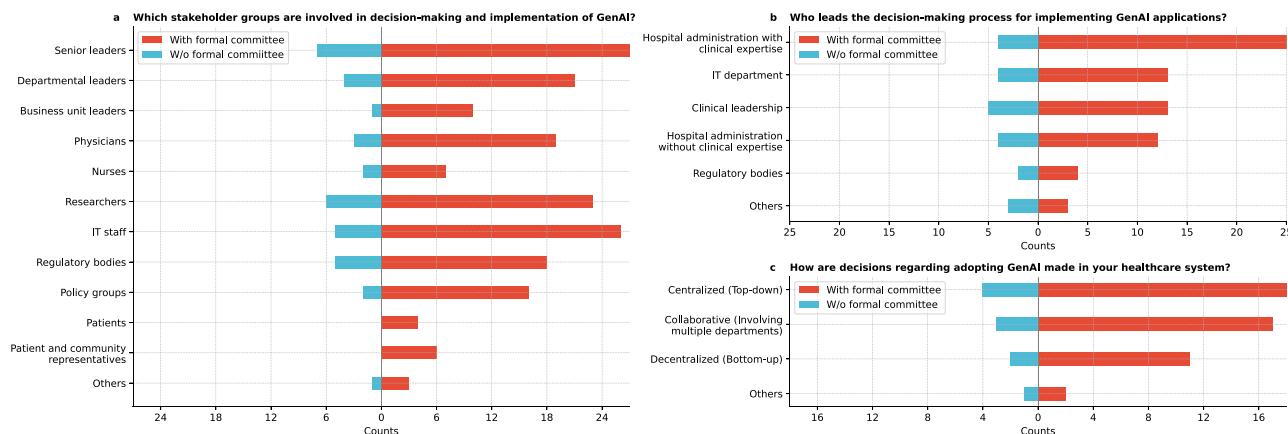


Fig. 1 | Results on stakeholder identification and roles. **a** Which stakeholder groups are involved in your organization's decision-making and implementation of GenAI? **b** Who leads the decision-making process for implementing GenAI

applications in your organization? **c** How are decisions regarding adopting GenAI made in your healthcare institution?

Fig. 2 | Thematic analysis of governance and leadership structures in GenAI deployment across CTSA institutions with featured responses. This figure illustrates two primary domains of governance and leadership structures: AI Governance and Policy (blue) and Strategic Leadership and Decision-Making (orange), divided into seven subcategories. Segment sizes reflect the prevalence of each approach. Annotated quotes provide qualitative insights into governance strategies, showcasing the diversity of institutional practices in GenAI deployment.



(22/36) of respondents, while 8.3% (3/36) mentioned alternative methods, such as decisions based on the tool's nature or a mix of centralized and decentralized approaches.

Thematic analysis of statements about governance structures in organizations with formal committees identified two major themes (Fig. 2). “AI Governance and Policy” reflects institutions’ structured approaches to ensure responsible GenAI implementation. Institutions often establish multidisciplinary committees to integrate GenAI policies with existing frameworks, aligning AI deployment with organizational goals and regulatory requirements and focusing on legal and ethical compliance. “Strategic Leadership and Decision Making” highlights the crucial role of leadership in GenAI initiatives. High-level leaders drive GenAI integration through strategic planning and resource allocation, with integrated teams from IT, research, and clinical care fostering a culture of innovation and collaboration. Excerpts on these governance practices are detailed in Supplementary Table 3.

Regulatory and ethical considerations

Regulatory body involvement in GenAI deployment varied widely across institutions (Fig. 3a). Federal agencies were engaged in 33.3% (12/36) of organizations. A significant portion (55.6%) identified other bodies, including institutional review boards (IRBs), ethics committees, community advocates, and state agencies. Internal governance committees and university task forces were also explicitly mentioned.

Regarding ethical oversight (Fig. 3b), 36.1% (13/36) of respondents reported an ethicist's involvement in GenAI decision-making; 27.8% (10/36) mentioned an ethics committee, while 19.4% (7/36) reported neither, and 16.7% (6/36) were unsure. Ethical considerations were ranked based on importance (Fig. 3c), with “Bias and fairness” (mean rank 2.31) and “Patient Privacy” (mean rank 2.36) being the top priorities.

Stage of adoption

Institutions were at varying stages of GenAI adoption (Fig. 3d), with 75.0% (27/36) in the experimentation phase, focusing on exploring AI's potential, building skills, and identifying areas for value addition. Integrating existing systems and workflows was met with mixed responses (Fig. 3e), with 50.0% (18/36) rating it as neutral.

Workforce familiarity with LLMs also varied (Fig. 3f), with 36.1% (13/36) of respondents reporting slight familiarity and 25.0% (9/36)

reporting moderate familiarity. Workforce training on LLMs was uneven, with only 36.1% (13/36) having received training, while 44.4% (16/36) considered but did not receive training, and 19.4% (7/36) neither received nor considered training. The demand for further training was evident, with 83.3% (30/36) finding it desirable or even more (Fig. 3g). The respondents who indicated receiving further LLM training for their workforce was undesirable were from institutions without a formal committee.

Vendor collaboration was crucial, with 69.4% (25/36) of institutions partnering with multiple vendors, ranging from one to twelve, to implement GenAI solutions. Notable vendors included major service providers, established EHR vendors, and various startups. Some respondents noted that discussions are often confidential or lack comprehensive information on enterprise-wide vendor engagements. Additionally, 25.0% (9/36) have considered vendor collaboration but have not engaged, while only 5.6% (2/36) have neither considered nor pursued such partnerships.

Budget trends

Regarding funds allocation for GenAI projects, 50.0% (18/36) of respondents reported that ad-hoc funding was allocated mostly from institutions with formal committees (Fig. 3h). Most institutions without formal committees reported that no funds had been allocated for GenAI projects (62.5%; 5/8). Since 2021, 36.1% (13/36) of respondents were unsure about budget changes, 19.4% (7/36) noted the budget remained roughly the same, and 44.5% reported budget increases ranging from 10% to over 300% (Fig. 3i).

Current LLM usage

Institutions were adopting LLMs with varied strategies (Fig. 4a), with 61.1% (22/36) using a combination of both open and proprietary LLMs, 11.1% (4/36) using open LLMs only, and 25.0% (9/36) using proprietary LLMs only. Only 2.8% (1/36) reported not using any LLMs. Significant differences exist ($Q = 28.7, p < 0.0001$) between the types of LLMs used. Post-hoc tests revealed significant differences (Supplementary Table 4) between using open and proprietary LLMs versus open LLMs only (corrected $p = 0.0032$), indicating a notable preference for combining different LLM types in some institutions. No significant differences were found among specific open or proprietary LLM types ($Q = 2.4, p = 0.4936$), suggesting that institutions did not exhibit strong preferences between particular open or proprietary LLM models. Institutions developing open LLMs prioritized technical

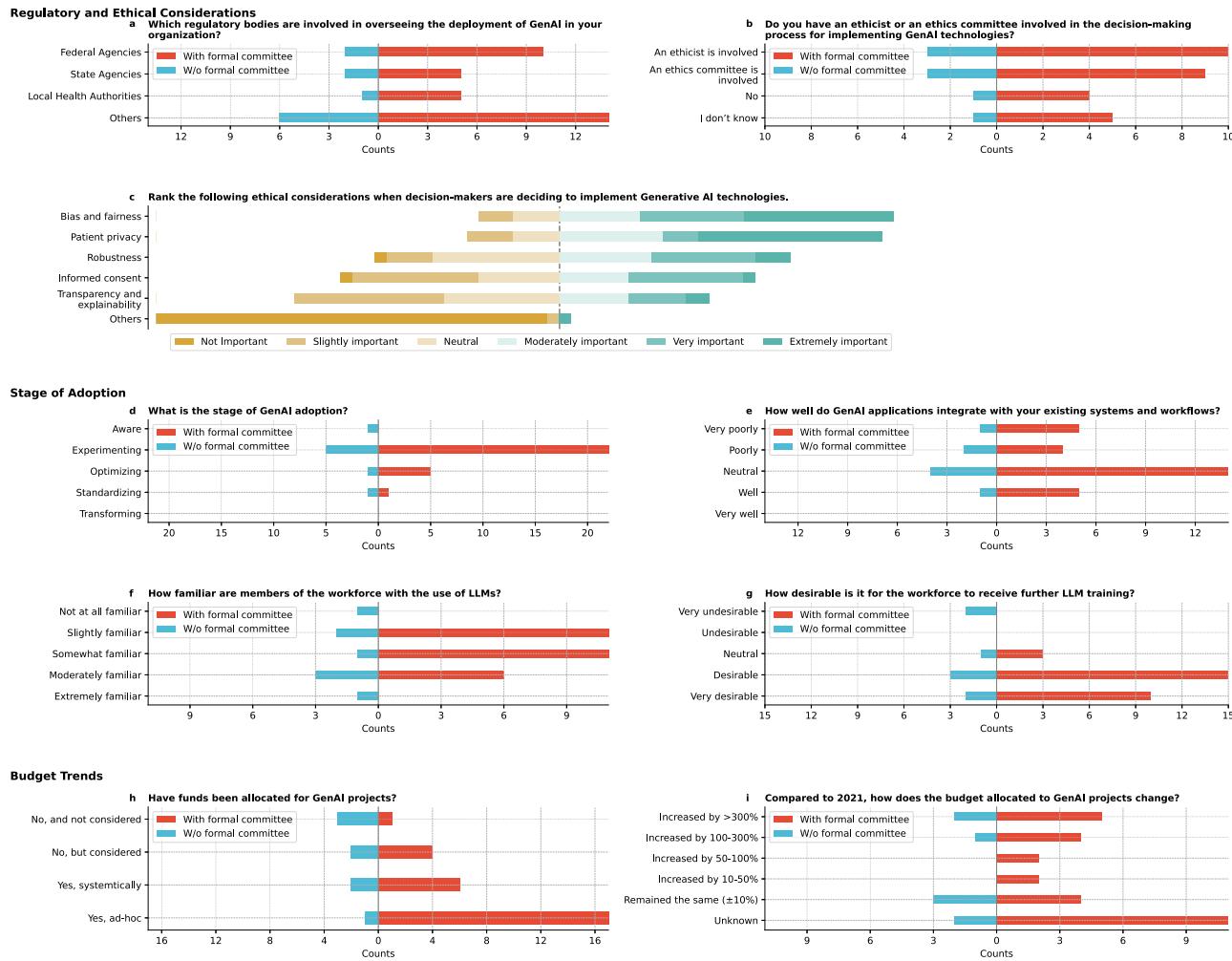


Fig. 3 | Results on regulatory, ethical, and budget considerations. **a** Which regulatory bodies are involved in overseeing the deployment of GenAI in your organization? **b** Do you have an ethicist or an ethics committee involved in the decision-making process for implementing GenAI technologies? **c** Please rank the following ethical considerations from most important (1) to least important (6) when decision-makers are deciding to implement GenAI technologies. **d** What is the stage of GenAI adoption?

the stage of GenAI adoption in your organization? **e** How well do GenAI applications integrate with your existing systems and workflows? **f** How familiar are members of the workforce with the use of LLMs in your organization? **g** How desirable is it for the workforce to receive further LLM training? **h** Have funds been allocated for GenAI projects? **i** Compared to 2021, how does the budget allocated to GenAI projects in your organization change?

architecture and deployment (61.1%), followed by customization and integration features (50.0%, Fig. 4c). Some institutions focused on research and experimentation, comparing open to proprietary LLMs, with interests in medical education and cost-effectiveness. Technical architecture and deployment are prioritized over clinician or patient buy-in (corrected $p = 0.0024$; Supplementary Table 5).

Regarding GenAI deployment (Fig. 4b), private cloud and on-premises self-hosting were the most common approaches (both 63.9%), suggesting that most institutions have both approaches but do not take a hybrid approach. Some institutions specified using local supercomputing resources or statewide high-performance computing infrastructure. Statistical analysis ($Q = 42.6$, $p < 0.0001$) indicated a preference for more controlled environments, with private cloud and on-premises self-hosting significantly more favored than public cloud (corrected $p = 0.0022$ and $p = 0.0060$, respectively; Supplementary Table 6).

For institutions adopting proprietary LLMs, the critical factors for decision-making include technical architecture and deployment (61.1%), and scalability and performance (Fig. 4c). Respondents noted the importance of ease of deployment, especially in partnerships with established EHR vendors, and the advantage of existing Health Insurance Portability and Accountability Act (HIPAA) Business Associate Agreements with major cloud service providers. Statistical analysis ($Q = 57.4$, $p < 0.0001$) revealed

significant differences, particularly between technical architecture and deployment and monitoring and reporting and AI workforce development (both corrected $p = 0.0113$; Supplementary Table 7). Scalability and performance were significantly more prioritized than LLM output compliance and AI monitoring and reporting (corrected p values = 0.0405).

Finally, LLMs were applied across diverse domains, with common uses in biomedical research (66.7%), medical text summarization (66.67%), and data abstraction (63.9%, Fig. 4d). Co-occurrence analysis showed frequent overlaps in these areas (Supplementary Table 8). Medical imaging analysis was the most common use case for institutions without formal committees overseeing GenAI governance. Significant differences (Supplementary Table 9) were observed in using LLMs for data abstraction compared to drug development, machine translation, and scheduling and between biomedical research and drug development, machine translation, and scheduling (corrected p values < 0.05).

LLM evaluation

Respondents prioritized accuracy and reproducible and consistent answers when evaluating LLMs for healthcare (Fig. 4e; Supplementary Table 10), each receiving the highest mean rating of 4.5. Healthcare-specific models and security and privacy risks were also deemed important, though responses varied. An analysis of variance (ANOVA) test revealed significant

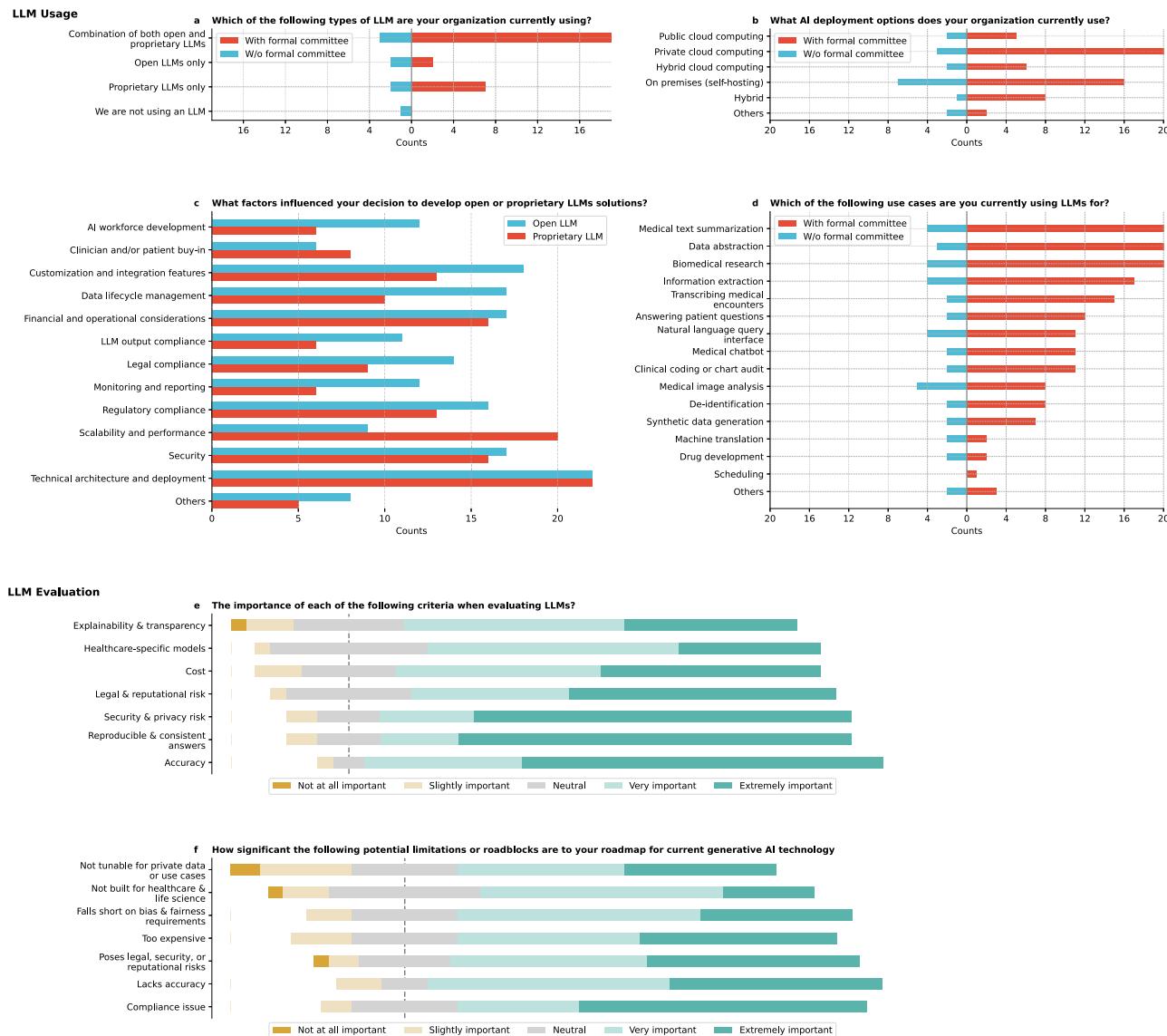


Fig. 4 | Results on LLMs usage. **a** Which of the LLMs are you currently using? **b** What AI deployment options does your organization currently use? **c** You indicated that your organization is using open LLMs (blue) or proprietary LLMs (red). What factors influenced your decision to develop internally/to go with commercial solutions? **d** Which of the following use cases are you currently using LLMs for? **e** On

a scale from 1 to 5, please rate the importance of each of the following criteria when evaluating LLMs. 1 means “Not at all Important,” and 5 means “Extremely Important”. **f** On a scale of 1 to 5, please rate how significant the following potential limitations or roadblocks are to your roadmap for current generative AI technology, with 1 being not important and 5 being very important.

differences among the importance ratings ($F = 3.4, p = 0.0031$). Post-hoc Tukey's honestly significant difference (HSD) tests showed a significant difference between accuracy, explainability, and transparency ($p = 0.0299$).

Regarding potential roadblocks to adopting GenAI in healthcare, regulatory compliance issues were rated as the most significant concern, with a mean rating of 4.2 (Fig. 4f; Supplementary Table 11). While ‘Too expensive’ and ‘Not built for healthcare and life science’ were less of a concern, they still posed challenges for some respondents, though there are no significant differences among these ratings ($F = 2.0, p = 0.0606$).

Projected impact

Participants rated the anticipated impact of LLMs on various use cases over the next 2–3 years (Fig. 5a; Supplementary Table 12), with the highest mean ratings for natural language query interface, information extraction, and medical text summarization (4.5 each), followed by transcribing medical encounters (4.3). Data abstraction (4.3) and medical image analysis (4.2) were also highly rated, while synthetic data generation, scheduling (3.5 each), and drug development (3.4) received

lower ratings. Additional use cases, such as medical education and decentralized clinical trials, suggest an expanding scope for LLM applications.

Further, respondents reported increased operational efficiency (44.4%) as the most commonly observed improvement, with faster decision-making processes noted by 13.9% (Fig. 5b). However, none reported improved patient outcomes. Other reported improvements included increased patient satisfaction and enhanced research capacity, although some noted it was too early to prove such benefits. Significant differences among these improvements were observed ($Q = 38.9, p < 0.0001$; Supplementary Table 13), particularly between better patient engagement and improved patient outcomes (corrected $p = 0.0026$).

Regarding GenAI implementation concerns (Fig. 5c), data security was identified as a major issue by 52.78% of respondents, followed by a lack of clinician trust (50.0%) and AI bias (44.44%). Cochran's Q Test confirmed variability in these concerns ($Q = 33.3, p < 0.001$). Other challenges included the time required to train models, lack of validation tools, inadequate provider training, and concerns about organizational trust. Some respondents

Projected Impact

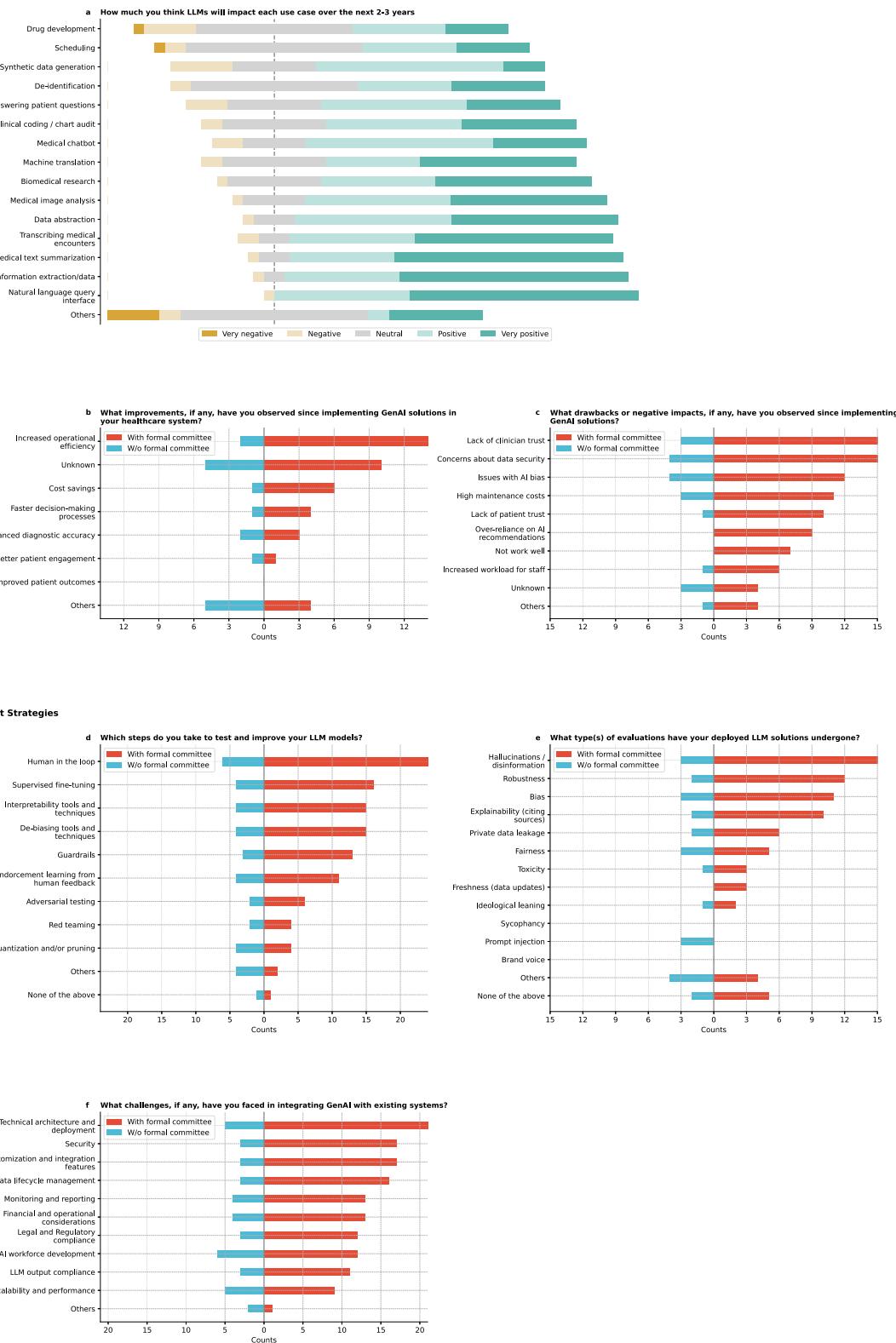


Fig. 5 | Results on projected impact and enhancement strategies. **a** On a scale of 1 to 5, please rate how much you think LLMs will impact each use case over the next 2–3 years. 1 means very negative, and 5 means very positive. **b** What improvements, if any, have you observed since implementing Generative AI (GenAI) solutions in your healthcare institution? **(c)** What drawbacks or negative impacts, if any, have

you observed since implementing GenAI solutions? **d** Which steps do you take to test and improve your LLM models? **e** What type(s) of evaluations have your deployed LLM solutions undergone? **f** What challenges, if any, have you faced in integrating GenAI with existing systems?

also noted that their observations were based on internal experiences, with no implementations yet in production.

Enhancement strategies

Respondents identified several strategies for testing and improving LLMs in healthcare, with human-in-the-loop being the most common (83.3%, Fig. 5d). Significant differences (Supplementary Table 14) were noted between human-in-the-loop and methods like quantization and pruning and Reinforcement Learning with human feedback²² (corrected $p < 0.0001$). Significant differences were found between adversarial testing²³ and human-in-the-loop and guardrails and human-in-the-loop (corrected $p = 0.0067$).

In evaluating deployed LLMs (Fig. 5e), the most common assessments focused on hallucinations or disinformation (50.0%) and robustness (38.9%). However, 19.4% (7/36) of respondents indicated no evaluations had been conducted. Cochran's Q Test revealed significant variation in the importance of these evaluations ($Q = 77.1, p < 0.0001$), with post-hoc analysis (Supplementary Table 15) showing significant differences between explainability and prompt injection (i.e., a technique where specific prompts or questions are used to trick the GenAI into bypassing its specified restrictions, revealing weaknesses in how it understands and responds to information), and between fairness versus ideological leaning and prompt injection (corrected $p = 0.0040$).

Integrating GenAI into healthcare presents several challenges (Fig. 5f), with technical architecture and deployment cited most frequently (72.2%). Interestingly, AI workforce development is the most common challenge for institutions without a formal committee. Data lifecycle management was noted as a critical limitation by 52.8% (19/36) of respondents. Challenges often overlap, with technical architecture and deployment closely linked to security, scalability, and regulatory compliance issues. Additional gaps were also highlighted, such as the absence of a training plan and a limited workforce. Significant variability was observed ($Q = 45.4, p < 0.0001$), with post-hoc analysis indicating that technical architecture and deployment were more prevalent than LLM output compliance (i.e., the trustworthiness of the LLM output) and scalability and performance (corrected $p = 0.0269$; Supplementary Table 16).

Additional insights into GenAI integration

Nine respondents provided additional insights into the complexities of integrating GenAI into healthcare. They emphasized the challenges posed by the rapid pace of technological change, which complicates long-term investment and integration decisions. Organizational approaches to GenAI vary; some institutions aggressively pursue it, while others have yet to implement it on a broader scale despite individual use. The integration of GenAI has improved collaboration between researchers, physicians, and

administrators, but slow decision-making and a significant gap in AI workforce skills remain critical issues. The evolving nature of AI initiatives makes it difficult to fully capture current practices, highlighting the need for a comprehensive approach that addresses technological, organizational, and workforce challenges.

Discussion

This study provides a snapshot of GenAI integration within CTSA institutions, focusing on key stakeholders, governance structures, ethical considerations, and associated challenges and opportunities. Table 1 summarizes the key recommendations from the findings. Senior leaders, IT staff, and researchers are central to GenAI integration, with significant involvement from cross-functional committees highlighting the multi-disciplinary collaboration required for effective implementation. However, findings suggest minimal involvement of nurses, patients, and community representatives in the current GenAI implementation decision-making process, which raises concerns about inclusiveness, which is essential to aligning technologies with the needs of all stakeholders^{18,24}. Most institutions adopt a centralized, top-down governance structure, streamlining decision-making but potentially limiting flexibility for departmental needs²⁵. While formal committees or task forces suggest emerging governance frameworks, the variability across institutions indicates that best practices are still evolving.

According to the respondents, ethical and regulatory oversight of GenAI implementation varies across institutions, with some involvement from federal agencies, IRBs, and ethics committees. Prioritization of ethical considerations such as patient privacy, data security, and fairness in AI algorithms reflects the awareness of the significant challenges in deploying GenAI in healthcare. Our findings also reveal variability in the reported involvement of regulatory bodies, with less frequent mentions of engagement from local health authorities. However, we did not collect detailed information on the specific roles of these agencies or distinguish between different types of regulatory engagement. This limitation suggests a need for more explicit and consistent oversight frameworks to address the unique risks associated with GenAI. Despite these gaps, this study emphasizes the importance of developing comprehensive policies and guidelines to navigate the ethical landscape of GenAI technologies in healthcare.

Collaboration with vendors is common among CTSA institutions, with partnerships reported with major cloud service providers and established EHR vendors. However, the variability in the extent of these collaborations and the need for comprehensive information on enterprise-wide vendor engagements suggest challenges in coordinating AI implementation efforts across institutions. Further, the ad-hoc funding allocation for GenAI projects indicates that AI integration is still in its infancy, with institutions likely testing the waters before

Table 1 | Summary of key findings and recommendations for GenAI implementation in healthcare

Key finding	Recommendation
Stakeholder involvement	Involve senior leaders, IT staff, researchers, clinicians, and patients to ensure a representative and effective decision-making process.
Governance structure	Establish formal GenAI governance committees to ensure structured oversight.
Decision-making	Cross-functional committees should lead decision-making for GenAI adoption, balancing stakeholder involvement.
Popular enhancement strategies	Use human-in-the-loop and supervised fine-tuning as primary enhancement strategies for LLM models.
Cloud architecture preferences	Prefer private cloud or on-premises hosting to maintain control over security, scalability, and regulatory compliance in GenAI deployment.
Ethical considerations	Prioritize bias and fairness, patient privacy, and data security when integrating GenAI into healthcare institutions.
Budget allocation	Encourage institutions to establish systematic funding mechanisms for GenAI projects to support long-term investments.
LLM usage	Adopt a combination of open and proprietary LLMs, depending on the technical and scalability requirements of the institution.
Workforce training	Implement comprehensive training programs to enhance GenAI literacy and bridge skill gaps within the healthcare workforce.
Projected impact and improvements	Focus on operational efficiency and decision-making speed while addressing the gap in direct improvements to patient outcomes.

Table 2 | Summary of key challenges in GenAI implementation across CTSA institutions

Challenge	Description
Stakeholder inclusion	Nurses, patients, and community representatives have limited involvement in the decision-making processes, particularly in institutions without formal committees.
Governance structure	Variability in governance models, with some institutions lacking formal GenAI oversight committees, may impact structured decision-making.
Leadership in decision-making	Institutions without formal committees rely more on clinical leadership rather than cross-functional committees, potentially affecting the balance of stakeholder input.
Ethical oversight	Varying degree of involvement of ethicists and ethics committees can create gaps and disparity in fairness, privacy, and data security in the broad scientific community for clinical and translational science.
Workforce readiness	Variability in workforce familiarity with LLMs, with some institutions having insufficient training and preparedness for GenAI integration.
Training and skill gaps	Significant gaps in formal GenAI training plans, with many institutions struggling to build internal capabilities to manage GenAI tools effectively.
Technical integration	Difficulties in integrating GenAI into existing systems, with mixed responses about how well these technologies integrate into current workflows.
Funding and resources	Many institutions rely on ad-hoc funding mechanisms for GenAI projects, creating uncertainty in long-term resource allocation and support for AI initiatives.
Vendor collaboration	Limited transparency and variability in vendor collaborations, with some institutions facing challenges coordinating enterprise-wide AI implementation.
Data security and trust	Major concerns regarding the security of GenAI systems and lack of clinician trust, particularly in institutions without formal governance structures.
AI bias and mistrust	Concerns about bias in GenAI outputs and mistrust from clinicians and patients could affect the adoption and effective use of GenAI technologies.
Compliance and legal risks	Regulatory compliance and accuracy are major concerns, with institutions needing to navigate legal and reputational risks associated with GenAI deployment.

committing to substantial investments. Implementing LLMs in healthcare settings presents significant challenges, particularly in technical architecture, deployment, customization, and security, requiring a comprehensive and coordinated approach across departments for successful integration²⁶. Additionally, data interoperability challenges, especially for multi-state or multi-jurisdictional institutions, further complicate these efforts, emphasizing the need for standardized frameworks to facilitate seamless integration across diverse technical and regulatory environments.

To evaluate their GenAI technologies, some institutions are using strategies like human-in-the-loop oversight, supervised fine-tuning, and interpretability tools to enhance GenAI transparency and reliability while also employing de-biasing techniques to mitigate biases, ensuring that GenAI outputs are continuously monitored and refined by human experts^{27,28}. Evaluation practices emphasize robustness and accuracy, with assessments for hallucinations, disinformation, and bias crucial to ascertaining GenAI systems function effectively in real-world healthcare settings^{29,30}. However, some institutions' lack of comprehensive evaluations suggests the early stages of LLM adoption and potential shortcomings in initial adoption, highlighting the need to improve their resources or expertise before widespread adoption.

The respondents are optimistic about the projected impact of LLMs on healthcare, particularly in areas like medical text summarization, query interfaces, and information extraction, which are expected to streamline workflows, enhance information access, and improve documentation efficiency^{31,32}. However, the gap between anticipated benefits and actual outcomes, such as the limited direct improvements in patient outcomes, highlights ongoing challenges. This discrepancy emphasizes the need for a focused evaluation of how GenAI tools can directly impact patient health and care quality. Emerging LLM applications in medical education, decentralized trials, and digital twin technologies (i.e., virtual replicas of physical systems used for real-time simulation and analysis in healthcare) suggest an expanding scope for these tools. While their impact in specialized domains like drug development remains uncertain, recent evidence points to promising advancements that could enhance the utility of LLMs in this area³³. Despite the enthusiasm, significant concerns about data security,

clinician trust, high maintenance costs, AI bias, and lack of patient trust complicate LLM integration into healthcare institutions.

Integrating LLMs into healthcare institutions is further complicated by high maintenance costs, AI bias, and lack of patient trust. Evaluations within institutions prioritize accuracy, reliability, and security, with respondents emphasizing the critical need for dependable and secure AI outputs to maintain trust and patient safety³⁴. Legal and reputational risks, along with the need for explainability and transparency, are also highly rated, indicating a significant focus on the ethical and legal implications of AI deployment. However, the importance of these criteria varies, reflecting diverse contexts and priorities across institutions. Despite high expectations for LLMs, the study identified significant roadblocks and considerations for widespread adoption (Table 2). These challenges underscore the complex landscape where multiple factors must be managed simultaneously.

Further, the study reveals that most institutions are still in the experimentation phase of GenAI adoption, exploring the technology's potential and building the necessary skills for its practical adoption. Mixed levels of familiarity with LLMs among the workforce and stakeholders indicate a significant need for further AI workforce training and clinician engagement to enhance GenAI literacy, ensuring that key stakeholders can manage GenAI effectively. Without proper training, healthcare professionals may struggle to fully leverage these tools, potentially leading to inefficiencies, errors, or privacy or security violations (e.g., inappropriately uploading data)^{35,36}. Previous work suggests a multifaceted and multi-sectorial approach to address these gaps and facilitate knowledge sharing, including implementing structured training programs, offering hands-on workshops, developing mentorship opportunities, and partnering with vendors to provide tailored training specific to the healthcare setting³⁷. This opens the possibility that NCATS and other NIH institutes may want to consider collaborative initiatives to address the questions raised in this research. Additionally, the CTSA network's emphasis on knowledge sharing could facilitate smoother GenAI adoption across institutions³⁸, particularly for late adopters. By encouraging the dissemination of best practices and lessons learned from early adopters³⁹, the CTSA network can help institutions with fewer resources or those facing governance challenges navigate the complexities of GenAI implementation more

efficiently. Furthermore, the insights from this study could inform strategies for GenAI adoption in non-CTSA institutes and contribute to shaping the global GenAI landscape, where diverse institutional structures and resource availability demand adaptable and scalable approaches.

The study has limitations, including variability in respondents' knowledge and the evolving nature of GenAI practices, which may not capture ongoing progress or changes beyond the survey period. Additionally, the reliance on responses from senior leaders, who may not have full visibility into all aspects of GenAI integration within their institutions, introduces the risk of misreporting or incomplete information. The focus on CTSA institutions may limit the generalizability of the findings to other healthcare organizations, particularly for institutions with fewer resources where these implementation and governance challenges may be especially difficult to address. The survey also did not distinguish between live GenAI systems and those still in development, which limits our ability to assess the operational readiness and deployment status of these tools fully across institutions. Further, we acknowledge that this study did not address energy costs and sustainability concerns, which are important considerations for GenAI technologies and should be explored in future work. Additionally, reliance on self-reported data introduces possible biases.

In conclusion, the study highlights the complex and evolving landscape of GenAI integration in CTSA institutions. By identifying successful strategies and highlighting areas for improvement, this research provides an actionable roadmap for institutions seeking to navigate the complexities of AI integration in healthcare to ensure ethical, equitable, and effective implementation, ultimately contributing to advancing patient care and the broader goals of precision medicine.

Methods

Study design

This study uses an online survey to conduct an environmental scan of GenAI infrastructure within CTSA institutions through multiple choice, ranking, rating, and open-ended questions to understand GenAI integration, including stakeholder roles, governance structures, and ethical considerations.

Survey instrument development

The survey, administered through the Qualtrics platform (Qualtrics, Provo, UT), was intended to take ~15 minutes to complete. Initially developed through a comprehensive review of current literature on AI in healthcare, the survey covered topics such as stakeholder roles, governance structures, ethical considerations, AI adoption stages, budget trends, and LLM usage. The survey was reviewed by experts (SL, BM, KN, WP, RZ, YZ) in health informatics, clinical practice, ethics, and law, who provided feedback that informed revisions to improve clarity and comprehensiveness. A small group piloted the final version to identify any remaining issues. The survey questions are available in the Supplementary File.

Participant recruitment

Participants were recruited in July 2024 through targeted outreach to key stakeholders at CTSA sites using purposive and snowball sampling⁴⁰. Email invitations were sent to senior leaders involved in GenAI implementation and decision-making within the CTSA network (<https://ccos-cc.ctsa.io/resources/hub-directory>), with follow-up reminders to maximize response rates.

Data collection

Data were collected from July to August 2024. CTSA leaders who responded to the initial invitation received a follow-up email with the survey link. A PDF version of the survey was provided to help participants prepare by reviewing questions offline before completing the survey online. Participants could return to the survey if necessary.

Data analysis

Quantitative data from the survey were analyzed using various methods. Multiple-choice and multiple-answer questions were summarized with frequency distributions and percentages. In addition, multiple-answer questions were also analyzed using co-occurrence and pattern analysis to identify common selections and combinations among stakeholder groups. Cochran's Q test identified overall differences among response proportions, with post-hoc analysis using pairwise McNemar tests with Bonferroni corrections⁴¹. Ranking questions were analyzed by calculating mean ranks, with lower mean ranks indicating higher importance. Likert-scale items were summarized using measures of central tendency and dispersion, with an ANOVA test to check for significant differences in ratings across different use cases, followed by Tukey's HSD test for post-hoc pairwise comparisons while controlling for the family-wise error rate⁴².

Qualitative data from open-ended survey questions was analyzed using thematic analysis⁴³. This process involved coding the data to identify common themes and patterns. Two researchers (BI, ZX) independently coded the data, and a third researcher (YP) resolved disagreements through consensus.

Data availability

No datasets were generated or analyzed during the current study.

Code availability

All data have been included in the manuscript and supplementary materials.

Received: 7 December 2024; Accepted: 22 December 2024;

Published online: 25 January 2025

References

1. Huang, J. et al. Generative artificial intelligence for chest radiograph interpretation in the emergency department. *JAMA Netw. Open* **6**, e2336100 (2023).
2. Matsubayashi, C. O. et al. Artificial intelligence for gastric cancer in endoscopy: from diagnostic reasoning to market. *Dig. Liver Dis.* **56**, 1156–1163 (2024).
3. Saaran, V., Kushwaha, V., Gupta, S. & Agarwal, G. A Literature review on generative adversarial networks with its applications in Healthcare: Proceedings of CIS 2020, Volume 1. In: *Congress on Intelligent Systems* (eds. Sharma, H., Saraswat, M., Yadav, A., Kim, J. H. & Bansal, J. C.) vol. 1334, 215–225 (Springer, Singapore, 2021).
4. Kazerouni, A. et al. Diffusion models in medical imaging: a comprehensive survey. *Med. Image Anal.* **88**, 102846 (2023).
5. Wei, R. & Mahmood, A. Recent advances in variational autoencoders with representation learning for biomedical informatics: a survey. *IEEE Access* **9**, 4939–4956 (2021).
6. Brown, T. et al. Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877–901 (2020).
7. Touvron, H. et al. LLaMA: open and efficient foundation language models. *arXiv* <https://arxiv.org/abs/2302.13971> (2023).
8. Alowais, S. A. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med. Educ.* **23**, 689 (2023).
9. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
10. Kim, J. et al. Large language models outperform mental and medical health care professionals in identifying obsessive-compulsive disorder. *NPJ Digit. Med.* **7**, 193 (2024).
11. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
12. Hao, B. et al. A GPT-based EHR modeling system for unsupervised novel disease detection. *J. Biomed. Inform.* **157**, 104706 (2024).
13. Amini, S. et al. Prediction of Alzheimer's disease progression within 6 years using speech: a novel approach leveraging language models. *Alzheimers Dement.* **20**, 5262–5270 (2024).

14. Jiang, F. et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* **2**, 230–243 (2017).
15. Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V. & Biancone, P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med. Inform. Decis. Mak.* **21**, 125 (2021).
16. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
17. Polevikov, S. Advancing AI in healthcare: a comprehensive review of best practices. *Clin. Chim. Acta* **548**, 117519 (2023).
18. Scott, I. A., Carter, S. M. & Coiera, E. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health Care Inform.* **28**, 102555 (2021).
19. Spotnitz, M. et al. A survey of clinicians' views of the utility of large language models. *Appl. Clin. Inform.* **15**, 306–312 (2024).
20. Liverman, C. T., Schultz, A. M., Terry, S. F. & Leshner, A. I. The CTSA program at NIH: opportunities for advancing clinical and translational research. National Academies Press (US); 2013. PMID: 24199260 (2013).
21. Yin, J., Ngiam, K. Y. & Teo, H. H. Role of artificial intelligence applications in real-life clinical practice: systematic review. *J. Med. Internet Res.* **23**, e25759 (2021).
22. Stiennon, N. et al. Learning to summarize from human feedback. *Adv. Neural Inf. Process. Syst.* **33**, 3008–3021 (2020).
23. Hao, B. et al. Distributionally robust image classifiers for stroke diagnosis in accelerated MRI. In: *Lecture Notes in Computer Science*, 768–777 (Springer Nature Switzerland, Cham, 2023).
24. Sujan, M. A., White, S., Habli, I. & Reynolds, N. Stakeholder perceptions of the safety and assurance of artificial intelligence in healthcare. *Saf. Sci.* **155**, 105870 (2022).
25. Argyres, N. S. Technology strategy, governance structure and interdivisional coordination. *J. Econ. Behav. Organ.* **28**, 337–358 (1995).
26. Denecke, K., May, R., LLMHealthGroup, Rivera & Romero, O. Potential of large language models in health care: Delphi study. *J. Med. Internet Res.* **26**, e52399 (2024).
27. Bakken, S. AI in health: keeping the human in the loop. *J. Am. Med. Inform. Assoc.* **30**, 1225–1226 (2023).
28. Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascárán, J. & Fernández-Leal, Á. Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.* **56**, 3005–3054 (2023).
29. Williamson, S. M. & Prybutok, V. The era of artificial intelligence deception: unraveling the complexities of false realities and emerging threats of misinformation. *Information* **15**, 299 (2024).
30. Menz, B. D., Modi, N. D., Sorich, M. J. & Hopkins, A. M. Health disinformation use case highlighting the urgent need for artificial intelligence vigilance: weapons of mass disinformation. *JAMA Intern. Med.* **184**, 92–96 (2024).
31. Tripathi, S., Sukumaran, R. & Cook, T. S. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. *J. Am. Med. Inform. Assoc.* **31**, 1436–1440 (2024).
32. Gebreab, S. A., Salah, K., Jayaraman, R., Habib ur Rehman, M. & Ellaham, S. LLM-based framework for administrative task automation in healthcare. In: 2024 12th International Symposium on Digital Forensics and Security (ISDFS) 1–7 (IEEE, 2024).
33. Yang, J. et al. Automating biomedical literature review for rapid drug discovery: Leveraging GPT-4 to expedite pandemic response. *Int. J. Med. Inform.* **189**, 105500 (2024).
34. Choudhury, A. & Chaudhry, Z. Large language models and user trust: focus on healthcare. *arXiv* <https://arxiv.org/abs/2403.14691> (2024).
35. Hazarika, I. Artificial intelligence: opportunities and implications for the health workforce. *Int. Health* **12**, 241–245 (2020).
36. Charow, R. et al. Artificial intelligence education programs for health care professionals: scoping review. *JMIR Med. Educ.* **7**, e31043 (2021).
37. Frehywot, S. & Vovides, Y. An equitable and sustainable community of practice framework to address the use of artificial intelligence for global health workforce training. *Hum. Resour. Health* **21**, 45 (2023).
38. Fishman, A. Y. et al. Moving from prove to improve: a collaborative continuous quality improvement process for advancing clinical and translational science. *J. Clin. Transl. Sci.* **8**, 1–21 (2024).
39. Escobar-Rodríguez, T. & Romero-Alonso, M. The acceptance of information technology innovations in hospitals: differences between early and late adopters. *Behav. Inf. Technol.* **33**, 1231–1243 (2014).
40. Biernacki, P. & Waldorf, D. Snowball sampling: problems and techniques of chain referral sampling. *Sociol. Methods Res.* **10**, 141–163 (1981).
41. Stephen, D. & Adruse, S. A. Z. Cochran's Q with pairwise McNemar for dichotomous multiple responses data: a practical approach. *Int. J. Eng. Technol.* **7**, 4 (2018).
42. Mircioiu, C. & Atkinson, J. A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy* **5**, 26 (2017).
43. Nowell, L. S., Norris, J. M., White, D. E. & Moules, N. J. Thematic analysis: striving to meet the trustworthiness criteria. *Int. J. Qual. Methods* **16**, 1609406917733847 (2017).

Acknowledgements

This work was supported by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) under grant numbers UL1TR002384, UM1TR004789, UM1TR004399, UL1TR001412, UL1TR001449, ULTR002345, UM1TR004404, UL1TR001866, UM1TR004909, UL1TR003017, UL1TR002377, UL1TR001430, UL1TR001425, and UL1TR001873; National Library of Medicine (NLM) of NIH under grant numbers T15LM007079, T15LM012495, and R25LM014213; and the National Institute on Alcohol Abuse and Alcoholism of NIH grant numbers R21AA026954 and R33AA0226954. This study was funded by the Novo Nordisk Foundation under grant number NNF23SA0087056. This study was also funded in part by the Department of Veterans Affairs, and the NIH Intramural Research, NLM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

B.I.: methodology, software, validation, data curation, formal analysis, writing, Z.X.: data curation, formal analysis, writing, W.G.A., M.A., N.R.A., N.B., D.S.B., C.B., T.C., M.C., P.L.E., J.W.F., D.J.F., D.H., M.H., M.K., A.K., A.L., C.G.L., L.L., C.L., J.L., Y.L., P. McGarvey, P. Mirhaji, J.D.O., P.A.H., F.P., N.J.S., I.S., L.R.W., R.J.W., A.H.Z., K.Z.: survey participants, manuscript revision, J.J.C., I.G.C., D.D., T.F., K.H., J.K.-C., N.S.K., Z.L., E.A.M., S.M., I.C.P., N.S., U.T.: manuscript revision, S.S.J.L., B.M., K.N., W.N.P. II, R.Z., Y.Z.: design of questions, manuscript revision, H.X., J.B., C.W., Y.P.: supervision, conceptualization, writing, funding acquisition.

Competing interests

I.G.C. is a member of the Bayer Bioethics Council, the Chair of the ethics advisory board for Illumina, and an advisor for World Class Health. He was also compensated for speaking at events organized by Philips with the Washington Post, by the Doctors Company and Danaher, attending the Transformational Therapeutics Leadership Forum organized by Galen Atlantica. He has been retained as an expert in health privacy, gender-affirming care, and reproductive technology lawsuits. Y.Z. serves as the editor but is not involved in the decision-making process for this paper.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44401-024-00009-w>.

Correspondence and requests for materials should be addressed to Hua Xu, Jiang Bian, Chunhua Weng or Yifan Peng.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA. ²Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA. ³Department of Pediatrics, Boston Medical Center, Boston, MA, USA; Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, USA. ⁴Oregon Clinical and Translational Research Institute, Oregon Health and Science University, Portland, OR, USA. ⁵Department of Public Health Sciences, University of California, Davis, Davis, CA, USA. ⁶Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ⁷Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. ⁸Department of Pathology and Laboratory Medicine, University of Kentucky College of Medicine, Lexington, KY, USA. ⁹Clinical and Translational Science Center, Weill Cornell Medicine, New York, NY, USA. ¹⁰Division of Pulmonary, Critical Care and Sleep Medicine, University of Kansas School of Medicine, Kansas City, KS, USA. ¹¹Department of Biomedical Informatics and Data Science, Heersink School of Medicine, University of Alabama, Birmingham, AL, USA. ¹²Harvard Law School, Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, Harvard University, Cambridge, MA, USA. ¹³Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA. ¹⁴Center for Clinical and Translational Science, Mayo Clinic, Rochester, MN, USA. ¹⁵Technology and Digital Solutions, Stanford Medicine, Stanford University, Stanford, CA, USA. ¹⁶Center for Biomedical Informatics, Rutgers Cancer Institute, New Brunswick, NJ, USA. ¹⁷Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA. ¹⁸Altman Clinical and Translational Research Institute (ACTRI), University of California San Diego, La Jolla, CA, USA. ¹⁹Department of Biostatistics and Health Data Science, School of Medicine, Indiana University, Indianapolis, IN, USA. ²⁰Department of Ophthalmology, CCTSI, University of Colorado, Aurora, CO, USA. ²¹Center for Clinical and Translational Science, Rockefeller University Hospital, Rockefeller University, New York, NY, USA. ²²AI4Health4All Center, Center for Clinical & Translational Science, and Department of Psychiatry, University of Illinois Chicago, Chicago, IL, USA. ²³Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA. ²⁴Penn State Clinical and Translational Science Institute, Hershey, USA. ²⁵Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. ²⁶Division of Translational Informatics, Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, USA. ²⁷Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA. ²⁸Duke Clinical Research Institute, Duke University, Durham, NC, USA. ²⁹Department of Population Health, Virginia Commonwealth University, Richmond, VA, USA. ³⁰Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ³¹Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. ³²Georgetown-Howard Universities Center for Clinical and Translational Science, Washington, DC, USA. ³³Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ³⁴Institute for Clinical Translational Research, Albert Einstein College of Medicine, New York, NY, USA. ³⁵Department of Neurology, Mass General Brigham, Somerville, MA, USA. ³⁶Department of Medicine, University of Alabama, Birmingham, AL, USA. ³⁷College of Engineering and Faculty of Computing & Data Sciences, Boston University, Boston, MA, USA. ³⁸Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. ³⁹Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA. ⁴⁰Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA. ⁴¹Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. ⁴²Department of Biostatistics, Health Informatics and Data Sciences, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁴³Department of Biomedical Informatics, Biostatistics, and Medical Epidemiology, School of Medicine, University of Missouri, Columbia, MO, USA. ⁴⁴Department of Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁴⁵Division of Health Informatics and Implementation Science, Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, MA, USA. ⁴⁶Department of Informatics, University of California, Irvine, Irvine, CA, USA. ⁴⁷Department of Medical Humanities and Ethics, Columbia University, New York, NY, USA. ⁴⁸Michigan Law School, University of Michigan, Ann Arbor, MI, USA. ⁴⁹Division of Computational Health Sciences, Medical School, University of Minnesota, Minneapolis, MN, USA. ⁵⁰Department of Biomedical Informatics and Data Science, Yale School of Medicine, Yale University, New Haven, CT, USA. ⁵¹Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA. ⁵²The Irving Institute for Clinical and Translational Research, Columbia University, New York, NY, USA. ⁵³Present address: Biostatistics and Health Data Science, School of Medicine, Indiana University, IN, USA. ⁵⁴Present address: Regenstrief Institute, Indianapolis, IN, USA. ⁵⁵These authors contributed equally: Hua Xu, Jiang Bian, Chunhua Weng, Yifan Peng.

✉ e-mail: hua.xu@yale.edu; bianji@iu.edu; cw2384@cumc.columbia.edu; yip4002@med.cornell.edu