At the nexus of health care and Generative AI

Accounting for trust in development, deployments, and monitoring

Deloitte AI Institute™
Deloitte Center for Health Solutions

About the Deloitte Center for Health Solutions

The source for health care insights

The Deloitte Center for Health Solutions (DCHS) is the research division of Deloitte LLP's life sciences and health care practice. The goal of DCHS is to inform stakeholders across the health care system about emerging trends, challenges, and opportunities. Using primary research and rigorous analysis, and providing unique perspectives, DCHS seeks to be a trusted source for relevant, timely, and reliable insights.

About the Deloitte AI Institute™

The Deloitte AI Institute helps organizations connect the different dimensions of a robust, highly dynamic and rapidly evolving AI ecosystem. The AI Institute leads conversations on applied AI innovation across industries, with cutting-edge insights, to promote human-machine collaboration in the "Age of With".

The Deloitte AI Institute aims to promote a dialogue and development of artificial intelligence, stimulate innovation, and examine challenges to AI implementation and ways to address them. The AI Institute collaborates with an ecosystem composed of academic research groups, start-ups, entrepreneurs, innovators, mature AI product leaders, and AI visionaries, to explore key areas of artificial intelligence including risks, policies, ethics, future of work and talent, and applied AI use cases. Combined with Deloitte's deep knowledge and experience in artificial intelligence applications, the Institute helps make sense of this complex ecosystem, and as a result, delivers impactful perspectives to help organizations succeed by making informed AI decisions.

No matter what stage of the AI journey you're in; whether you're a board member or a C-Suite leader driving strategy for your organization, or a hands on data scientist, bringing an AI strategy to life, the Deloitte AI institute can help you learn more about how enterprises across the world are leveraging AI for a competitive advantage. Visit us at the Deloitte AI Institute for a full body of our work, subscribe to our podcasts and newsletter, and join us at our meet ups and live events. Let's explore the future of AI together.

www.deloitte.com/us/AIInstitute

The health care industry thrives on trust—trust in providers and payers, medicines and therapies, supply chains, and all the technologies that enable enterprise operations and delivery of care.

Many hospital and health systems are in the process of digital transformation, adopting modern technologies to address challenges in the workforce, operational efficiency, personalized care, and more. Amid this, there appears to be a general wariness around the risks associated with artificial intelligence (AI). Many business leaders likely know well the sobering examples of health care AI deployments that introduced risks effecting enterprise trustworthiness and even health equity. In some cases, the adoption of a technology meant to improve the organization may have served to diminish the all-important quality of trust.

With the arrival of a new kind of AI—Generative AI—the stakes are even higher, and health care leaders are appropriately concerned about that risks this type of AI could create. Generative AI is different from other types of AI in that it can intake a natural language prompt, consult a variety of data, and create a novel output that appears to be human generated. Models can output data in six modalities: text, image, video, audio, computer code, and 3D rendering. The large language models (LLMs) that have captured public attention are a type of Generative AI.

Already, 75% of leading health care companies are experimenting with Generative AI or attempting to scale use cases, according to Deloitte's 2024 Life Sciences and Health Care Generative AI Outlook Survey.[1] This activity comes in tandem with the adoption of complementary technologies, including cloud computing, data modernization and analytics, and the Internet of Things (IoT). This is to the good, as the greatest potential value in Generative AI is not found in a vacuum or standalone deployment. Instead, a digital transformation as a whole is what can permit hospital and health care systems to create truly differentiating Generative AI use cases.

The challenge, as with all AI, is to identify and mitigate the risks, preserve trustworthiness, and promote the ethical use of technology. Generative AI use cases can inject substantial, sometimes unobvious risks. To help elucidate them and inform Generative AI governance and oversight, consider some of the value-driving use cases and their implications for trust.

Generative AI across the health care value chain

While Generative AI use cases are proliferating, some of the most powerful are likely yet to be imagined. The following examples serve as a window into the risks health care organizations may encounter when building their Generative AI programs.

Drug discovery

The process for developing new treatments can be both time-consuming and costly. Part of the challenge is analyzing vast biomedical datasets and teasing out the insights and correlations that inform clinical development. Bringing Generative AI to bear, medical data can be queried for patterns or anomalies that point toward a discovery, and drug compounds can be simulated in a virtual environment to rapidly validate drug candidates without needing to manufacture them in the real world. The potential benefit is not only new insights but the capacity to transform drug discovery, permitting scale and speed that eclipses human capabilities. What is more, in a time of inflation, shrinking margins, and rising costs for supplies and labor,[2] **technologies like Generative AI can fuel innovation while limiting the costs associated with more traditional drug discovery approaches.**

The value is evident, but what of the risks? Medical data is subject to regulations, with expectations for patient privacy and data security. When a model accesses proprietary, sensitive, or regulated data, the organization needs guardrails and processes in place to help ensure the model only consumes data that can be legally and compliantly analyzed, as well as preventing model outputs from inadvertently leaking protected information to unauthorized users. Data of this kind is also a target for nefarious actors, requiring robust cybersecurity measures to help guard the information in line with health data privacy rules and laws.

But there may be a more fundamental challenge: the quality of the data the model consumes during training and after deployment. Poor data quality translates to poor model outputs. Biased or incomplete datasets may lead a model to output (with total confidence) a novel treatment design that shows clinical promise. Relying on that output and moving toward pre-clinical testing, the enterprise may discover that the design is flawed and unworkable due to poor data, meaning the time investment and R&D costs to that point were wasted. This shows the importance of data modernization for Generative AI. It is not just the model that can create risk. It is the net result of multiple technologies working together, revealing that risk and trust should be assessed and addressed holistically and with a mind toward technology governance.

Diagnostics

One of the most powerful and perhaps surprising capabilities of Generative AI models is the capacity to look across large (often siloed) datasets and extract patterns or correlations that might otherwise go unseen. For example, current research shows that a person's pitch and strength in speaking can be analyzed to diagnose type 2 diabetes.[3] Generative AI appears well-suited to this task, given that it is multimodal (i.e., can consume and output multiple data types).[4] Other Generative AI diagnostic applications may include retinal imaging for disease detection or even a generalized AI agent that can review an individual's data and biometrics to identify a range of diseases or conditions.

Generative AI-enabled tools and products point toward a future where medical professionals could leverage machine capabilities to achieve something greater than what either could do independently. But in this age of human-machine collaboration, Generative AI outputs require human validation, particularly for something as impactful as diagnosis and treatment recommendations. The reality is that all Generative AI models are subject to errors or inaccuracies (by virtue of how Generative AI works), and trust in the viability and value of these models hinges in part on how the human stakeholder assesses and validates the outputs. There are also important questions around the laws and rules governing medical services, liability, and all the factors related to diagnosing and treating maladies.

Payer and provider capacity

For health care payers and providers, the labor-intensive and time-consume activities around prior authorizations and customer claims are ripe for automation, and Generative AI's capabilities can unlock greater efficiencies not possible with other types of AI. For example, providers can use Generative AI to prepare a prior authorization submission, with the model autonomously consulting data on requirements, guidelines, and a patient's medical records to help ensure requirements are met. Payers for their part can use Generative AI to accelerate prior authorization decisions, as well as monitor for fraud by flagging anomalies in a provider's coding practices.

Greater time- and resource-efficiencies support operations and also move toward improved customer service, which can enhance the patient experience and promote trust in the organization. Similar value-driving outcomes may be captured with a Generative AI-fueled claims assistant for patients in payer call centers. Automated, hyper-personalized, empathetic customer service can support the patient experience while expanding capacity for live human agents to focus on the most complex cases.

Across these use cases, however, are common risks around output reliability, impartiality, and security. Output validation should be built into the processes surrounding these deployments, and the underlying data should be interrogated for bias, completeness, and accuracy.[5] What is more, security and governance grow out of the connected infrastructure that permits Generative AI deployments—from the IoT phone systems in call centers to the cloud applications that permit data access and storage to the robustness of the model itself.

This view of Generative AI as a component of a technology ecosystem that collectively supports trustworthy, ethical deployments **can orient decision making and investments as health care systems pursue digital transformations.**

Supply chain management

The COVID-19 pandemic revealed the complexity and fragility of the global supply chain. Constraints around medical device availability (e.g., ventilators), pharmaceutical production, and tier N supplier reliability impacted the capacity to deliver care and underscored the importance of supply chain visibility and forecasting. Generative AI can help. By consuming data around geographic characteristics, disease prevalence, socio-economic factors, and logistical realities, **Generative AI can be used to create micromarket demand forecasts and optimized supply chain strategies.** This could have grand implications not just for the delivery of care but also cost efficiency for patients, health care organizations, insurers, and governments.

There are associated risks. Supply chain managers should understand how the Generative AI model derived its outputs and recommendations, including the data it consulted and how it arrived at its conclusions. Put another way, the model should be explainable to stakeholders in a meaningful way. Supply chain managers are unlikely to have much technical understanding of how a Generative AI model works, but to rely on and validate its outputs, they should have a level of Generative AI fluency that allows them to collaborate effectively with the machine. More broadly, AI literacy is important for every stakeholder who touches the AI lifecycle.

Shared technology value

In some ways, Generative AI can be seen as an interface between the user and the data they seek to query. For health care patients, this can open the door to a range of use cases that can improve the patient experience and even patient outcomes. A Generative AI-enabled personalized agent could be used for preliminary biomedical reporting (e.g., disease symptoms) or for inquiring about a treatment plan. This could take the shape of a generalized AI agent trained to perform in multiple contexts, or agents could be fine-tuned for use in a discrete part of the patient's health care experience. The possible outcome includes shared technology value, where patients receive greater personalized service, workforce capacity is expanded, enterprise efficiency and profitability are enhanced, and at the broadest level, public health is promoted while economic activity is invigorated.

A similar shared technology value can be found within the health care organization. The suite of tools and infrastructure that constitute a Generative AI-ready, modernized tech stack can offer new opportunities to capture data about employee performance and behavior. By collecting and analyzing granular details about the workforce, the organization can create the plans and incentives to improve performance, promote workplace safety and wellbeing, personalize continuous learning, and identify and support future leaders. As well as improving organizational operations, data-driven insights support enhanced productivity and worker happiness.[6]

True shared technology value hinges on consent to collect and use personal data. Patients may require a level of technical transparency, such as notice of how their information is stored and used. Customers (as well as regulators and lawmakers) generally want confidence that a Generative AI model and enabling infrastructure meet data security and privacy rules and laws. For the workforce, there may be a growing sensitivity to data privacy, use, and security as the organization collects greater, more granular troves of data about their employees, and data privacy concerns may be more acute depending on gender, according to Deloitte's 2023 Connected Consumer Survey.[7]

The insight is that in the course of developing and using Generative AI, the capabilities it affords can create user uncertainty or concern, and driving toward shared technology value means establishing the governance mechanisms that can account for data privacy, security, and indeed, all of the domains of trustworthy technology.

Toward a framework for trustworthy Generative AI in health care

Identifying and understanding Generative AI risks can help organizations make incisive decisions around how to promote trust in technologies and build a strong foundation for governance and compliance.

A framework can be helpful in this regard. Deloitte leverages the Trustworthy Generative AI framework[8] to evaluate technology use cases, both internally and with our clients. One of the reasons a framework is important is that many Generative AI deployments are inherently new and unique to the organization deploying them. Even open-source models with well-understood limitations (e.g., model "hallucination") tend to present new risks and trust implications when trained on enterprise data in a secure environment. In short, **no two organizations are the same, nor are their Generative AI deployments.** It is important to assess each use case in light of the organization's ethical priorities and culture, the potential risks across the AI lifecycle, and the applicable rules and laws (which can differ between geographies).

As health care organizations move forward with Generative AI and construct a framework for trust, ethics, governance, and compliance, **some of the important questions include:**

Which sensitive data sets are necessary for model training, and what laws or regulations apply to that data? Can all stakeholders access all information, and if not, what are the mechanisms needed to restrict data access? Is there the potential for the model to leak sensitive data in its outputs, and if so, what tactics can be used to mitigate that risk?

To what degree do stakeholders need to understand how a Generative AI deployment works and the data it consumes? To what extent do users need to be aware of and consent to their data being consumed by the model? Is there a tradeoff between transparency, accuracy, and user value? How is model function explained in a meaningful way to stakeholders with varying degrees of technical literacy?

Does the underlying data contain latent bias? How could that impact the quality or fairness of model outputs? What are the legal, regulatory, or brand implications of biased Generative AI outputs, and conversely, what are the advantages of impartiality?

Which stakeholders are responsible for considering and managing risk and ethics across the AI lifecycle? Are these stakeholders aware of their responsibility? Which roles or groups could be created internally to support trustworthy technology, such as a Chief Trust Officer or an AI governance board composed of internal stakeholders?

How provably reliable are model outputs? What are the documented processes for output validation, and where are the waypoints for evaluating model function in the face of new, rare, or unfamiliar data?

What security infrastructure surrounds sensitive or propriety data and Generative AI deployments? Is data storage, access, retrieval, and synthesis aligned with rules and laws for data security across all geographies? Are there cybersecurity risks, whether due to poor network security or a targeted attack (e.g., prompt spoofing) from a bad actor? How can the enterprise evidence model security and safety to legal and regulatory authorities?

How is post-deployment monitoring performed, and which stakeholders are accountable? Is monitoring a documented, repeatable process or is it nascent and ad hoc? What are the reporting channels for communicating concerns or issues, and how is that feedback integrated into the AI lifecycle?

Where are the opportunities to co-create or co-develop Generative AI use cases with the users who will be impacted? What are the varying perspectives and concerns voiced by patients, customers, vendors, and the workforce, and how can their input be fed into Generative AI development?

What changes or enhancements are necessary to ensure robust data governance? Does the organization's technology modernization plan include changes to roles, responsibilities, processes, and complementary technologies in a way that supports data governance? What are the implications of weak data governance for trustworthy Generative AI?

Generative AI represents a transformative step forward in human-machine collaboration. **This is a moment for innovation and creativity, where health care organizations can look beyond the status quo and imagine how market needs and technology maturity can open the door to entirely new services, products, business models, and approaches to patient care.** By accounting for risk, promoting trust, and ensuring compliance, health care enterprises can extract the greatest potential from Generative AI, for their organization and the patients they serve.

Reach out for a conversation.

Beena Ammanath
Global Deloitte AI Institute Leader
Deloitte LLP
bammanath@deloitte.com

Dr. Jay Bhatt
Deloitte Center for Health Solutions
and Deloitte Health Equity Institute
Executive Director
Deloitte Services LP
jaybhatt@deloitte.com

Endnotes

[1] Asif Dhar et al, "From Code to Cure, How Generative AI Can Reshape the Health Frontier," Deloitte, 2023.

[2] Tina Wheeler, "2023 Outlook for Health Care," Deloitte Health Forward Blog. December 13, 2022.

[3] Jaycee M. Kaufman et al. "Acoustic Analysis and Prediction of Type 2 Diabetes Mellitus Using Smartphone-recorded Voice Segments," Mayo Clinic Proceedings: Digital Health, Original Article, Volume 1, Issue 4, P534-544, December 2023, published online October 17, 2023.

[4] Ibid., Asif Dhar et al.

[5] Beena Ammanath, "Technology Trust Ethics," Deloitte.

[6] Art Mazor et al, "Beyond Productivity, The Journey to the Quantified Organization," Chapter 2, Deloitte, May 2023.

[7] Susanne Hupfer, "Tech Companies Have a Trust Gap to Overcome—Especially with Women," Deloitte Center for Technology, Media & Telecommunications, 2023.

[8] Beena Ammanath, "Building Trustworthy Generative AI," Deloitte, 2023.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

About Deloitte
Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.