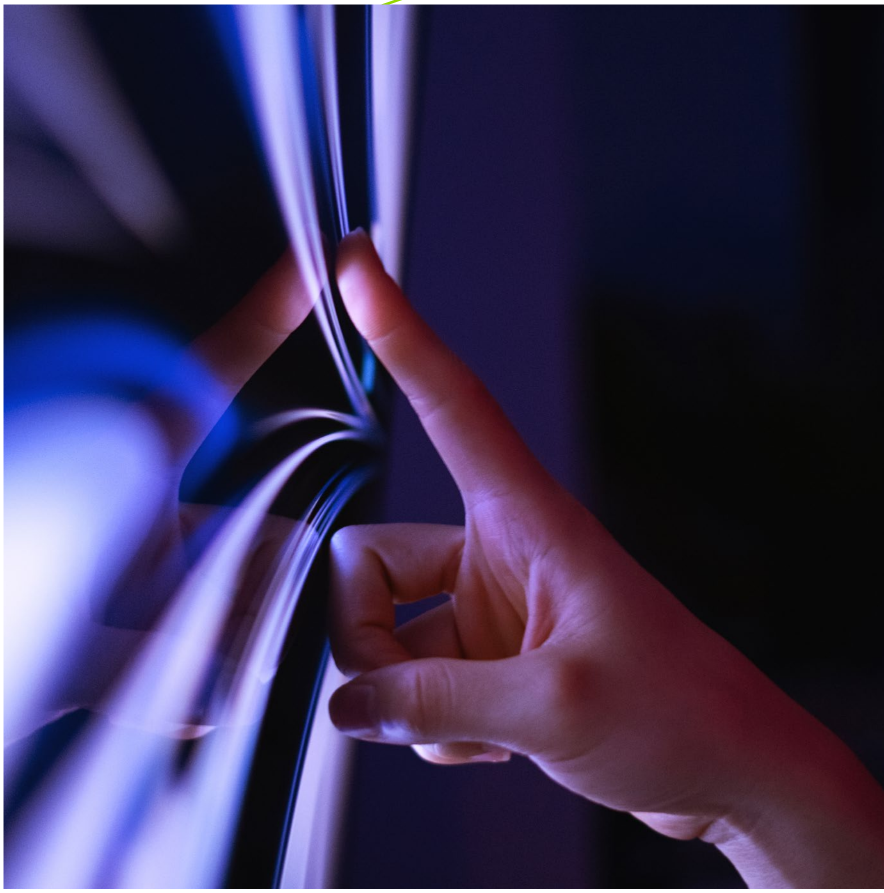
# Generative AI in Healthcare: Data Privacy and Accuracy Challenges

## Generative AI can revolutionize processes. But healthcare organizations must tackle its data privacy and data accuracy issues.



Although not a new concept, generative AI is a popular buzzword. With new capabilities, this technology has the potential to radically transform healthcare. From its ability to create first drafts for email responses to its promise of automating rudimentary coding, there are countless ways it can be deployed to boost productivity, improve accuracy, and support better decision-making.

Despite its potential, generative AI technologies have well-known risks around data privacy and data accuracy. These risks make adoption in highly regulated industries like financial services, insurance, pharmaceuticals, and healthcare particularly complicated.

Several concerns have recently made news. Italy's data protection watchdog banned OpenAI in March 2023 due to suspected European Union General Data Protection Regulations (GDPR) violations and concerns over its lack of age-gating. The ban was reversed in late April. But other countries' privacy and data watchdogs have also raised concerns—including the Office of the Privacy Commissioner of Canada and the United States Senate Judiciary Subcommittee on Privacy, Technology, and the Law, which held hearings on generative AI in May.

To address these concerns, a number of regulations are emerging to govern the use of generative AI, including the EU's AI Act[1], which regulates products that are high risk, and President Biden's work towards developing an AI Bill of Rights.[2] Many industries will also need to look to their sector's existing data privacy regulations to govern how they adopt the technology.

In this paper, we'll examine generative AI's data privacy and data accuracy risks by looking at the challenges the highly regulated healthcare industry might face in adoption. We'll examine the risks with generative AI, look at potential use cases, and discuss ways organizations can mitigate and tackle roadblocks to generative AI adoption in any sector.

## Data Accuracy Risks

The emerging AI technology that many industries are looking to adopt involves large language models (LLMs), the type of generative AI model that powers ChatGPT. These models train on large unlabeled datasets. The goal of their training is to help the models understand the connections between words—which the systems represent as an algorithmic equation. After this unsupervised training is complete, these models are often extensively fine-tuned via supervised learning or techniques like reinforcement learning with human feedback (RLHF), to help the model better understand misinformation in its dataset and what kinds of responses are preferred. To create ChatGPT, for example, OpenAI hired 6,000 annotators to label an appropriate subset of data. The company's machine learning engineers then used that data to fine-tune the model to teach it to generate specific kinds of replies.

Despite these efforts toward data accuracy, the way that LLMs work—essentially by predicting the next likely word in a sentence—means that the models are prone to what's often called "hallucinations." AI hallucinations are the false information that LLMs often generate, either in response to a straightforward question or when given disinformation in a prompt. For example, an LLM might say that a famous actor was born in the wrong year, starred in movies that don't exist, or is married when the actor is actually single. It might also claim that the Golden Gate Bridge is in Italy if asked where in Europe it's located.

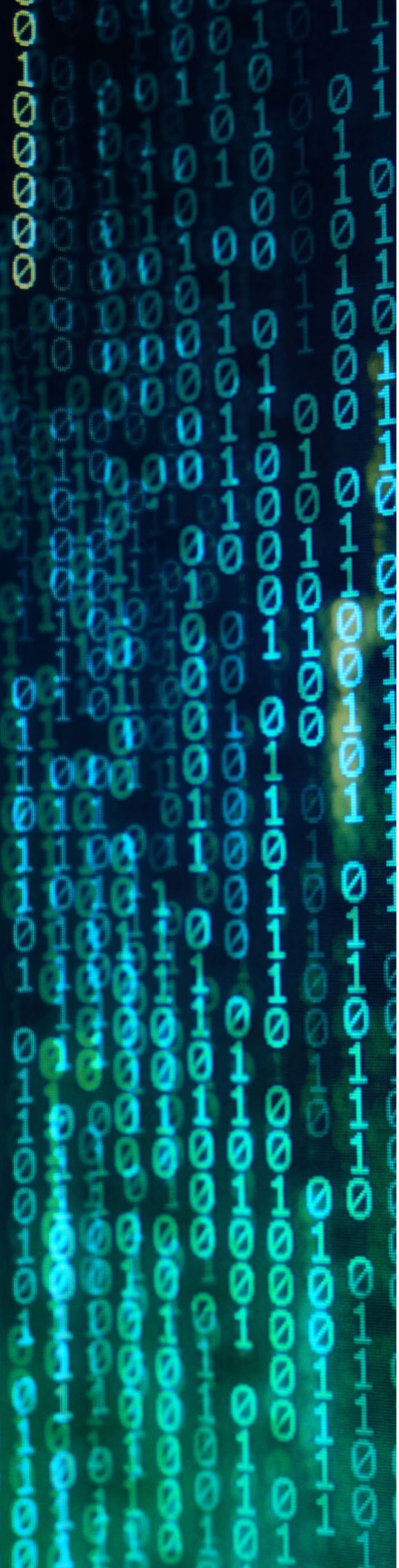
While that sort of misinformation doesn't have life-or-death consequences, misinformation in certain use cases for AI in the healthcare sector could. For example, using generative AI to help with diagnostics or treatment recommendations could open up liability issues if a patient was misdiagnosed or a model exhibited bias, another common problem. Other issues could also emerge in instances where ineffective prompts might lead to inaccurate responses. This could be a problem in cases where health systems rely on chatbots to help patients decide what kind of medical professional to contact for a particular complaint. Also, general purpose LLMs like ChatGPT and GPT-4 that are trained on a broad dataset don't have a significant amount of medical literature in their training set. They are not designed for clinical care delivery and are a bad fit for many types of use cases in healthcare compared to more niche models trained on a higher percentage of medical literature.

Still, LLMs show exciting potential for certain uses in healthcare. Tasks like answering patient queries, helping manage healthcare workflows, assisting in scheduling appointments, sending personalized reminders, providing aftercare instructions, writing doctors' notes, and supporting healthcare professionals in decision-making could work well so long as the data accuracy issues can be appropriately monitored.

Generative AI, with its ability to create synthetic data and content, presents challenges of misrepresentation or misuse of information. However, its potential for diagnosing diseases, predicting patient outcomes, or accelerating drug discovery is noteworthy. Adopting robust data governance, thorough vetting processes, and maintaining transparency can help mitigate these risks, enabling broader and safer adoption of generative AI across sectors.

## Data Privacy Risks

Data accuracy issues in LLMs might create liability in healthcare if the wrong information is given out. Still, the more salient concerns for the sector are expected to be the data privacy risks associated with LLMs. Currently, many AI models employ user-input data as part of their training data, analyzing that data without human input and drawing their own conclusions. Even if AI companies try to anonymize that user data, it's not always effective. Unredacted and unencrypted data could be seen by the AI company's employees, could surface in prompts, or could even be part



of a data breach, as happened in March[3] when some OpenAI users could see titles from other users' chat history.

Although OpenAI's application programming interfaces (APIs) demonstrate a commitment to security standards, with Systems and Organization Controls (SOC) 2 Type 2 compliance, and facilitate clients' HIPAA compliance via business associate agreements, numerous other LLMs still need to achieve this regulatory standard in healthcare. Some prominent LLMs include GPT-3 and Transformer models, Microsoft's Turing-NLG, Google's Meena, and T5. These LLMs offer advanced natural language processing capabilities beneficial to healthcare, such as interpreting medical documents or patient interactions. However, despite these advantages, until these models achieve necessary compliance, their utility in healthcare remains limited, underscoring the need for LLMs to align with regulatory requirements for broader healthcare applications.

Healthcare providers must avoid AI tools that haven't been approved by their health system as meeting requirements for storing protected health information (PHI) and compliance with HIPAA, along with any applicable state privacy laws. Healthcare entities should also update their policies and procedures when it comes to privacy and security to ensure that they're accessing PHI in compliant ways when leveraging generative AI.

## Tackling Generative AI's Privacy and Accuracy Risks Today

Despite these risks, there are ways healthcare organizations interested in implementing generative AI today can remain compliant and optimize the advantages offered by these promising technologies. The first thing to do is to ensure you fully understand the benefits and drawbacks of the LLMs being used. While similar, each model performs better on specific tasks and topics depending on its training and dataset.

### Ways to use LLMs more safely and effectively in the healthcare sector today:

- Fine-tune a general model on additional data.
- Seek out a specialty LLM model focused on healthcare use cases.
- Refine a specialty model on data specific to your use case.
- Limit use of LLMs to use cases where accuracy isn't critical.
- Combine LLMs with other AI models to enhance capabilities, for example, integrating with image recognition AI for improved diagnostics.
- Get explicit patient consent for care or communication that involves generative AI.
- Continually monitor the performance of the LLMs to ensure they are working as intended and making accurate predictions or interpretations.

## Tackling Generative AI's Privacy and Accuracy Risks in the Future

What does the future hold for risk management in generative AI? As the field matures, expect many of the regulatory issues that are currently stumbling blocks to adoption in certain sectors to be addressed, opening the field to broader industrial use.

Healthcare organizations looking to adopt generative AI technologies to transform their workflows or services must navigate privacy concerns, data accuracy risks, and regulatory requirements, while updating processes and policies to accommodate this new technology. Guidehouse is well-positioned to help organizations create a generative AI strategy specific to industry needs, use cases, and compliance requirements. With our cross-functional expertise in key sectors, proficiency in risk management, and generative AI experience and knowledge, we're uniquely equipped to help organizations navigate risks and develop a strategy to take advantage of generative AI's potential while maintaining compliance.

## How LLMs will be safer and more effective for the healthcare sector in the future:

1. **Advanced Regulation Compliance:** Future LLMs will be designed with built-in mechanisms for better adherence to regulatory standards like HIPAA, GDPR, and others, enhancing data security and patient privacy.
2. **Improved Interpretability:** Development in AI interpretability will allow for better understanding of LLMs' decisions, leading to enhanced trust and safer usage.
3. **Precision Medicine:** With advances in AI, LLMs will enable more accurate personalized care. They'll be capable of interpreting complex medical data to create individualized treatment plans, improving patient outcomes.
4. **Enhanced Training Algorithms:** Future LLMs will use advanced training methodologies that help in fine-tuning for specific healthcare tasks, ensuring more accurate and reliable performance.
5. **Integrative AI Solutions:** Combining LLMs with other AI technologies like image recognition or predictive analytics will create comprehensive solutions that improve diagnostics and treatment efficacy.
6. **Ethical AI Frameworks:** As AI ethics evolve, future LLMs will be built with ethical considerations in mind, minimizing biases and ensuring fair and equitable use.
7. **User Feedback Loop:** LLMs will employ feedback mechanisms for continuous learning and improvement, enhancing their effectiveness over time.

[1] "The Artificial Intelligence Act", whitehouse.gov, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[2] "Blueprint for an AI Bill of Rights", whitehouse.gov, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[3] "March 20 ChatGPT outage: Here's what happened", openai.com, May 24, 2023, https://openai.com/blog/march-20-chatgpt-outage.

### Contacts

**Bob Dunmyer, Partner**
Data, Analytics & Intelligence
bdunmyer@guidehouse.com

**Bassel Haidar, Director**
Data, Analytics & Intelligence
bhaidar@guidehouse.com

### About Guidehouse

Guidehouse is a leading global provider of consulting services to the public sector and commercial markets, with broad capabilities in management, technology, and risk consulting. By combining our public and private sector expertise, we help clients address their most complex challenges and navigate significant regulatory pressures, focusing on transformational change, business resiliency, and technology-driven innovation. Across a range of advisory, consulting, outsourcing, and digital services, we create scalable, innovative solutions that help our clients outwit complexity and position them for future growth and success. The company has more than 16,000 professionals in over 55 locations globally. Guidehouse is a Veritas Capital portfolio company, led by seasoned professionals with proven and diverse expertise in traditional and emerging technologies, markets, and agenda-setting issues driving national and global economies. For more information, please visit guidehouse.com.

guidehouse.com/services/data-analytics-intelligence

@ghtechsolutions

linkedin.com/showcase/guidehouse-technology-solutions/

outwit complexity™

This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors. GH-241 WP Generative AI in Healthcare