Hashed & Salted | A Privacy and Data Security Update
October 2023

Navigating Health Data Privacy in AI—Balancing Ethics and Innovation

In the ever-evolving health care landscape, the integration of artificial intelligence (AI) has emerged as a transformative force with the potential to revolutionize patient care, diagnosis, treatment and medical research. AI's ability to swiftly analyze vast amounts of health-related data promises enhanced medical outcomes, cost efficiency and improved patient experiences. Yet, as AI's presence in health care grows, it brings a web of legal considerations. Our Privacy, Security & Data Innovations team has been closely monitoring the privacy challenges companies seeking to harness health data for AI solutions must address and resolve. In this article, we unpack the term "AI" and outline some specific considerations for AI methods, outline the web of privacy laws that will apply when certain forms of health data are used to power AI solutions and close with a checklist lawyers can use to help guide their legal product reviews.

What do we mean when we say 'AI'?

We should start any analysis by level setting on a definition of AI. The IAPP defines artificial intelligence as "an engineered system where machines learn from experience, adjusting to new inputs, and potentially performing tasks previously done by humans. More specifically, it is a field of computer science dedicated to simulating intelligent behavior in computers. It may include automated decision-making."

If you use AI, understanding the methods used will be critical to unpacking the legal implications. The U.S. Department of Health and Human Services (HHS) published a Trustworthy AI Playbook in 2021 that outlines some of the key considerations in the AI methods commonly used in the health care space.

AI Method	Definition	Key Considerations
Machine Learning (ML)	A subfield of AI involving algorithms that enable computer systems to iteratively learn from and then make decisions, inferences or predictions based on data. These algorithms build a model from training data to perform a specific task on new data without being explicitly programmed to do so.	■ Data sets should be reviewed for bias and diversity (to ensure outcomes aren't skewed based on health care demographics). ■ Automated decision-making that will significantly impact individuals (e.g., denial of claims, access to health care and treatment) should have a human review or an opportunity for the individual to opt out. ■ The source of the data. Privacy laws will apply to any identifiable information.
Natural Language Processing (NLP)	A subfield of AI that helps computers understand, interpret and manipulate human language by transforming information into content. It enables machines to read text or spoken language, interpret its meaning, measure sentiment and determine which parts are important for understanding.	■ To the extent that NLP uses personal information for training, consider the principle of "explainability." ■ Consider how incorrect interpretations could negatively impact affected individuals and identify a process to audit outcomes or introduce a human review/check.
Speech Recognition	Voice command technology that allows users to interact with and control technologies by speaking to them.	■ Voice and speech should include a broad range of languages, dialects and accents. ■ To the extent that voiceprints are used, consider how to address applicable biometric laws.
Computer Vision	Intelligent algorithms that perform important visual perception tasks such as object recognition, scene categorization, integrative scene understanding, human motion recognition and material recognition.	Teams should train computer vision models with data representative of the patient population.

Privacy/ethical challenges in AI-driven health care

Whether your AI solution uses patient data or health data more broadly, a variety of state and federal privacy laws will regulate that data if it is identifiable. Even if you use de-identified data, your organization will need strong data governance methods in place to reduce the risk that it will be reidentified. Several privacy laws include restrictions on automated processing (which would include AI) that has "legal or significant effects." "Significant effects" could include bias and discrimination, claims denial, imbalanced access to health care or treatment, inaccurate diagnoses, or even emotional harm in the event a breach reveals health information.

Sourcing your data: Informed consent and data collection

If you are leveraging an AI solution that requires health data, you will need to consider the source of that data and understand the privacy and security implications of its use.

Informed consent. In the United States, the law recognizes that in some cases liability can attach to a health care provider if the patient is not informed of the risks and benefits of proposed treatment and non treatment. This concept is similarly reflected in privacy laws, which require meaningful notice regarding how personal information is used, consent for the collection of certain categories of data and the right to opt out or object to certain processing activities, including automated decision-making. Accordingly, privacy requirements will vary based on the nature of the data and how it is being used.

Protected health information (PHI). If you are using personal information protected by the Health Insurance Portability and Accountability Act (HIPAA), the HIPAA Privacy Rule does not allow for the use of PHI unless there is a specific permission or requirement in the Privacy Rule. Permitted disclosures include:

- Treatment, payment and health care operations
- Public interest and benefit activities (for public health agencies, law enforcement, etc.)

You may be able to use PHI in connection with an AI-based solution designed to deliver treatment, manage payments or support other health care operations. However, when you use PHI to train data sets or otherwise aid in the design and development of those solutions, you may need to obtain patient authorization (or confirm that the data received from a third party was collected pursuant to an authorization).

Health data (non-PHI). Any data collected outside HIPAA (i.e., directly from the individual and/or not from or on behalf of a covered entity) will still be covered by state privacy laws. Since the passing of the California Consumer Privacy Act (CCPA) in 2019, 12 other states—Utah, Connecticut, Virginia, Colorado, Iowa, Indiana, Oregon, Montana, Tennessee, Texas, Florida and Delaware — have passed similar legislation. While many of these states have exceptions, which exclude information and entities regulated by HIPAA, non-PHI health data is typically treated as sensitive personal information that is subject to additional restrictions. Several states require opt-in consent for the use of sensitive personal information, while other states require consent for secondary use cases. To the extent your organization wants to use health data previously collected for one purpose for use in the development of an AI solution, it may need to obtain consent for additional use cases unless any of those use cases fall into a statutory exception. Finally, several states have passed laws specific to health data. Most notably, Washington State's My Health, My Data Act requires HIPAA-like authorization to collect and use consumer health data, which is defined very broadly under the law.

Biometric data. Voiceprints, faceprints, retinal scans and other biometric data will likewise be considered sensitive personal information and subject to the restrictions described above. Beyond the privacy laws regulating sensitive personal information, Illinois, Texas and Washington State have specific laws addressing the collection and use of biometric data. The Biometric Information Protection Act (BIPA) in Illinois is one of the most litigated statutes due to its private right of action. BIPA requires prior written consent to collect and use biometric information. While data collected from patients in a health care setting is carved out of the definition of biometric data under BIPA, all non-HIPAA covered biometric data will be subject to written consent requirements similar to what is required by HIPAA.

Limited data sets or de-identified data. Removing direct identifiers from your data sets may frustrate the purpose of your AI solutions. However, where possible, using limited data sets (defined by HIPAA as PHI from which certain specified direct identifiers of individuals and their relatives, household members and employers have been removed) or de-identifying data per HIPAA or state privacy law standards will enable you to use that data without obtaining consent. Companies may disclose a limited data set for research, health care operations and public health purposes, provided the recipient enters into a data use agreement promising specified safeguards for PHI within the limited data sets. De-identified data generally falls outside the definition of personal information under both HIPAA and state privacy laws. That said, data that is de-identified today could be reidentified tomorrow. Companies will need to implement effective data controls and a governance structure to reduce the risk of reidentification.

Protecting your data

HHS' Office for Civil Rights (OCR) has reported over 239 breaches in 2023, affecting the health care data of more than 30 million individuals within the U.S. Most notably, many of these breaches are a result of third-party hacking. Ransomware companies have focused on hacking and retrieving sensitive patient health data,

including PHI, in an attempt to extort companies for money, as they need to regain access to this information. It is vital to the security of health data that health care providers, research facilities and other companies involved in health care operations sufficiently vet any vendors that may have access to that health data, including providers of AI products and solutions. Analysis of vendor risk may include:

- Deciding whether access to health data is necessary and beneficial to optimize use of the AI product.
- Ensuring that vendors have adequate access controls in place to limit access to health data and any derivative algorithms for individuals who need access to such information for the purposes of providing the AI products or solutions to the organization.
- Ensuring that vendors have implemented policies and procedures that adequately protect against the unauthorized use and disclosure of health data, including PHI, such as implementing the recently developed National Institute of Standards and Technology (NIST) Healthcare Framework. The NIST Healthcare Framework outlines a national set of security standards and guidelines aimed at protecting health care data. Vendors that have adopted and implemented these standards illustrate the establishment of a privacy infrastructure that incorporates data protection, incident response and system security to demonstrate HIPAA compliance.
- Negotiating contractual provisions with vendors that permit the auditing of systems that contain health data and to continue compliance monitoring.
- Requiring vendors to conduct data privacy and security employee awareness trainings.

Beyond vetting your vendors, companies should also examine their internal security posture. HHS has warned HIPAA-regulated entities to "determine the potential risks and vulnerabilities to ePHI before adding any new technology into their organization." HHS recommends health care entities consider the following AI risk mitigation strategies:

- Review NIST's Artificial Intelligence Risk Management Framework.
- Review the MITRE Atlas knowledge base of adversary tactics, techniques and case studies for ML systems.
- Adopt AI-based tools for defense, including penetration testing, threat detection, threat analysis and incident response.
- Provide AI training for cybersecurity personnel.

This guidance is useful even if your organization is not covered by HIPAA or regulated by HHS. Companies should look to NIST, ISO-27001 or CIS Controls to benchmark their security practices. The costs of a breach are escalated when a company is deemed not to have had "reasonable" security controls in place.

Evaluating use cases

Identifying the potential for harms (or significant effects). Companies that are currently performing privacy impact assessments may consider leveraging this process to conduct AI impact assessments, which should be designed to determine the potential harm from the use of AI and to identify strategies for mitigating the risks created by their AI solutions/products. While the U.S. does not currently have an AI-specific regulation or law requiring these assessments, documenting the steps taken to identify and mitigate risks will help provide an internal record of how a company arrived at its position, and it may help explain its position in the event of a regulatory inquiry.

Beware of false advertising. In early 2023, the Federal Trade Commission (FTC) published a blog post, Keep Your AI Claims in Check. In that post, the FTC outlines its concerns about the risks of false or deceptive AI advertising claims. The FTC notes that it is focused on whether:

- Claims about what an AI product can do are being exaggerated. The FTC notes that "we're not yet living in the realm of science fiction, where computers can generally make trustworthy predictions of human behavior. Your performance claims would be deceptive if they lack scientific support or if they apply only to certain types of users or under certain conditions."
- A company is promising that an AI product does something better than a non-AI product without proof.
- A company understands and has addressed the reasonably foreseeable risks. Notably, the FTC warns companies that they can't blame the third-party developer of a technology, even when that technology is a "black box" or otherwise difficult to test.
- A product actually uses AI.

What's next: A quick checklist for lawyers

Data source assessment

- ☐ What are the sources of data?
- ☐ Which laws attach to it?
- ☐ Who owns the data/do we have the full scope of rights required?
- ☐ If data is purchased/licensed from a third party, what audit/assessment ability do we have to evaluate the data quality, integrity and relevance?
- ☐ If data is identifiable, what scope of consent is required and have we obtained it? Can we rely on consent previously provided or provided to a third party?
- ☐ If data is de-identified, what standard of de-identification are we using (HIPAA safe harbor/FTC/expert determination)? What controls are in place to prevent reidentification?

Data security

- ☐ What security controls are applied?
- ☐ Have the teams been trained on the cybersecurity controls?
- ☐ Do we have role-based access controls?
- ☐ Do we have the ability to create an audit trail?
- ☐ What data retention policies are in place?
- ☐ Have we vetted any third-party vendors? Do we have ongoing audit rights?

Identifying risks

- ☐ What is the use case?
- ☐ Are the outcomes explainable?
- ☐ What is the risk that the underlying data set could result in biased outcomes? What demographics are included? Are any demographics over- or under-included?
- ☐ If the outcomes are inaccurate, who would be harmed and how? How can we prevent that harm (e.g., human review)?
- ☐ Can we honor individual privacy rights (opt out, correction, deletion, access)? If not, what alternatives can we offer to protect individual privacy?
- ☐ Can we conduct a bias and fairness assessment?
- ☐ Have we defined data retention policies?
- ☐ Do we have a process for continuous monitoring of outcomes and individual impact?

We will continue to monitor the legal landscape governing the use of AI in health care. Please do not hesitate to reach out to our Privacy, Security & Data Innovations team here at Loeb.

Related Professional

Bianca Lewis . . . . . blewis@loeb.com

7450 REV1 10-24-2023