More Data Please!

The Challenges of Applying Health Information Privacy Laws to the Development of Artificial Intelligence

By Adam H. Greene, Davis Wright Tremaine LLP

Artificial intelligence (AI)[1] has become part of our daily lives, from greeting us in the morning through smart home devices, creating shopping lists, playing music, setting timers, and alerting us of a traffic jam on our expected route home. AI offers substantial potential benefits for health care. AI can assist individuals with identifying whether a problem merits a trip to the doctor. It can watch an ill newborn and alert parents and doctors of signs of distress. AI can analyze huge datasets and identify patterns that otherwise may go unnoticed, such as unexpected side effects of drugs or contributing factors to improved outcomes.

The challenge is that AI will not improve health care in a vacuum. AI developers need tremendous amounts of health information to teach AI the vocabulary and grammar of medicine and the structure and meaning of electronic health record (EHR) and claims data. Generally speaking, the more data, the better the results.

AI's hunger for data, however, can create challenges under existing privacy laws. Federal and state legislatures and agencies did not draft current health information privacy laws, such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,[2] with AI in mind. As a result, we are left applying old law to new technology—not a new or unique problem, particularly in health care, but always a challenge.

Attorneys counseling clients with AI-related health care activities may wish to consider:

- Is de-identified information feasible;
- Does AI "use" health information;
- Is there a permissible purpose for the AI activity;
- How does the Privacy Rule's "minimum necessary" standard apply;

- What is the impact of laws placing greater restrictions on information about certain sensitive conditions; and
- Is notice, consent, and/or opt-out legally required or advisable?

This article focuses on privacy issues concerning protected health information (PHI). Attorneys also should consider various other legal issues related to AI, such as addressing potential inherent biases,[3] and challenges with validating and explaining AI.[4]

Can the AI Developer Use De-Identified Data?

A threshold question is whether parties can perform an AI activity using de-identified data. HIPAA provides two means of de-identification:

(1) The "Safe Harbor" method, in which the covered entity or business associate removes 18 categories of identifiers and does not have actual knowledge that the remaining information can identify an individual;[5] or

(2) The "Expert Determination" method, in which a covered entity or business associate obtains a written determination from an appropriate statistical expert that the risk of identification of an individual is very small.[6]

Many reasons exist to use de-identified data, where feasible. First, under the Privacy Rule's "minimum necessary" standard, a covered entity or business associate must make reasonable efforts to use or disclose only the minimum necessary amount of PHI for an intended purpose, such as the AI activity.[7] Accordingly, the AI activity should use de-identified information when it is feasible to do so. Second, once the information is de-identified, HIPAA will not apply. This would give the parties involved in the AI activity the most flexibility. For example, the Privacy Rule's prohibition on the sale of PHI is not applicable to de-identified information, allowing a covered entity or business associate to receive remuneration in exchange for disclosing the data to an AI developer. Third, de-identifying the data best protects the privacy interests of the individuals, reducing privacy concerns and risk should a breach of the data occur.

Numerous challenges with de-identification, however, often make it infeasible. For unstructured data, a covered entity or business associate may find it difficult to ensure that they remove all of the 18 categories of identifiers. Attempting to de-identify this type of information through automation may lead to imperfect results and hiring persons to do so is time consuming and costly (and also likely to lead to imperfect results). Alternatively, hiring a de-identification expert involves the cost of the engagement, delay in engaging the expert (as they are in short supply), and delay in the expert reviewing the data and rendering a decision.

Additionally, de-identification under HIPAA does not necessarily constitute de-identification under other laws, such as the California Consumer Privacy Act or the European Union's General Data Protection Regulation.[8]

> AI developers need tremendous amounts of health information to teach AI the vocabulary and grammar of medicine and the structure and meaning of electronic health record (EHR) and claims data.

Is AI Processing a "Use" of Protected Health Information?

Another threshold question is whether AI's processing of PHI even qualifies as a "use" or "disclosure" of PHI. In the Privacy Rule's commentary, the Department of Health and Human Services (HHS) advised that "computer processing" of data does not constitute a "use" that is subject to the Privacy Rule:

> *Comment:* One commenter observed that the definition [of "use"] could encompass the processing of data by computers to execute queries. It was argued that this would be highly problematic because computers are routinely used to identify subsets of data sets. It was explained that in performing this function, computers examine each record in the data set and return only those records in the data set that meet specific criteria. Consequently, a human being will see only the subset of data that the computer returns. Thus, the commenter stated that it is only this subset that could be used or disclosed.
>
> *Response:* We interpret "use" to mean only the uses of the product of the computer processing, not the internal computer processing that generates the product.[9]

Based on this commentary, one reasonably can argue that an algorithm combing through terabytes of EHR data to "learn" does not constitute a "use" of the data for purposes of the Privacy Rule if no human ever lays eyes on the PHI.

There are reasons to be cautious with this position, however. The commentary is guidance, which does not have the force of law, and HHS may have changed its interpretation since December 2000 in light of significant technological advances. Additionally, HHS guidance regarding ransomware provides that computer processing (the action of an outside party's malware accessing and encrypting data) constitutes a "disclosure" of PHI.[10] If malware encrypting PHI without a human seeing the results is a "disclosure," then HHS could interpret that AI processing the PHI without a human viewing the information is a "use." This may indicate that HHS departed from its prior position and likewise considers internal computer processing of

A threshold question is whether parties can perform an AI activity using de-identified data.

PHI to be a "use" of the information. Furthermore, a regulator or court could distinguish between search queries—which do not identify data that do not meet search parameters—and AI, which arguably is using all of the data to "learn."

Although the argument that AI's analysis of PHI is not a "use" may prove helpful, the remainder of this article will take the conservative approach and treat AI's application of algorithms to PHI as a "use" of PHI that requires a permission under HIPAA.

What Is the AI Activity's Purpose?

When parties must use or disclose PHI for the AI activity, they should identify the purpose of the activity. This will dictate the application of HIPAA. An entity subject to HIPAA may not use or disclose PHI unless specifically permitted or required by HIPAA.[11]

Treatment

One potential purpose is treatment. An example would be a HIPAA-covered health care provider using AI to assist with determining the best course of treatment for a patient. HIPAA would permit this activity as a use of PHI for the covered entity's treatment purposes without an individual authorization.[12]

To qualify as "treatment," the activity must involve a health care provider[13] and be on behalf of a single individual, rather than a population.[14] For example, the use of AI to review a population and identify patients who would benefit from an alternative treatment is a population-level activity and, therefore, qualifies as "health care operations" (discussed below) rather than "treatment."[15]

Although HIPAA provides great latitude for uses and disclosures for treatment, a covered entity likely would need a HIPAA-compliant business associate agreement (BAA) with the AI vendor.[16] Although HIPAA carves out treatment disclosures in the definition of "business associate," this exception is limited to disclosures to health care providers.[17] Unless the AI vendor qualifies as a "health care provider," then the exception would not apply and HIPAA would require a BAA.

Payment

Another potential purpose of AI is payment. Examples would include a health care provider or its billing company using AI to identify an appropriate code for a health care service, or a health plan using AI to identify billed services that may be medically unnecessary. HIPAA would permit this activity as a use of PHI for the covered entity's payment purposes.[18]Again, HIPAA likely would require a compliant BAA with an AI vendor or an authorization

Health Care Operations

HIPAA also permits a covered entity or business associate to apply AI to PHI for a covered entity's health care operations.[19] HIPAA broadly defines "health care operations" to encompass a range of activities.[20] A covered entity or business associate could use AI technology to conduct almost any of these activities, from quality assessment and improvement activities, to planning where a health care provider should expand its footprint.

"Health care operations" are limited to "the activities of the covered entity to the extent that the activities are related to covered functions."[21] Although a business associate may use or disclose PHI to support a covered entity's health care operations, HIPAA does not treat activities for the benefit of the business associate itself as "health care operations." Additionally, the use of AI for purposes unrelated to "covered functions"—the functions that make an entity a health care provider, health plan, or health care clearinghouse—also would not be "health care operations."[22] For example, a health care provider using AI to improve sales at its gift shop seemingly does not relate to its covered functions and, therefore, arguably may not be "health care operations."

"Health care operations" include the use of PHI to create de-identified data.[23] It does not matter how an entity then will use the de-identified information. For example, a covered entity's use of PHI to create de-identified information that the covered entity will sell to an AI developer would constitute a health care operation. A business associate may use PHI to create de-identified information if permitted by the applicable BAA.[24] HHS has issued guidance permitting a business associate to de-identify information for its own benefit, rather than the benefit of the covered entity, if permitted by the BAA to do so.[25]

Development of AI: Health Care Operations, Research, or None of the Above?

The million-dollar question—or maybe billion-dollar question—is whether development of AI could qualify as "health care operations" or "research." The definition of "health care operations" includes "[c]onducting quality assessment and improvement activities, including outcomes evaluation and development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities."[26] It also includes "population-based activities relating to improving health or reducing health care costs" and "protocol development."[27]

In contrast, HIPAA defines "research" as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[28]

If the purpose of developing AI is to use the AI to improve the quality of health care, evaluate outcomes, or develop clinical guidelines, then the activity may qualify as either health care operations or research. HHS has articulated that the line between the two is whether the "primary purpose" of the activity is to contribute to generalizable knowledge.[29]

For example, if a health care provider engages with a company to develop AI primarily to improve the quality of health care for the health care provider's patients, then the use and disclosure of PHI to develop the AI arguably is a health

care operation. The health care provider would need to enter into a BAA with the AI developer.

In contrast, if a health care provider engages with a company to develop AI primarily for purposes of contributing to generalizable knowledge, then the use and disclosure of PHI would be for research purposes. The activity generally would require: (1) the HIPAA-compliant authorizations of the patients; (2) an institutional review board or privacy board waiving the authorization requirement; or (3) the health care provider limiting the PHI to a limited data set and entering into a data use agreement with the AI developer.[30]

HHS acknowledges that there may be circumstances where an activity begins as health care operations and then transitions to research.[31] HHS directs that, in these cases, "the covered entity should document the change in status of the activity to establish that they did not violate the requirements of [the Privacy Rule]."[32] Accordingly, if a partnership between a covered entity and AI developer begins primarily to improve outcomes for the covered entity's population, but later transitions to primarily seeking to contribute to generalizable knowledge, then the parties should document the transition and begin to comply with the requirements for research.

There is little question that, if the parties' primary purpose is to publish an academic paper regarding the effectiveness of the developed AI, then this qualifies as contributing to generalizable knowledge and, therefore, research. What if the parties develop the AI as part of commercial research and development? What if the parties do not intend to publish the result, but instead intend to sell the AI? This remains a gray area.

Contributing to generalizable knowledge arguably does not necessitate publication. For example, HHS commentary to the Privacy Rule references drug research of pharmaceutical companies—research that is commercial, rather than academic, in nature.[33] Although pharmaceutical companies are acting for commercial purposes, their research and development of new drugs arguably contributes to generalizable knowledge for purposes of HIPAA. Likewise, an AI developer's research and development of AI to improve health care outcomes arguably contributes to generalizable knowledge and, therefore, would be research for HIPAA purposes. Calling the activity "research" is not a free pass under HIPAA. The parties generally would need to obtain an institutional review board or privacy board's waiver of authorization, or limit the information to a limited data set, if they wish to conduct the activity without individuals' authorizations.

Because the line between health care operations and research can get blurry, counsel should consider how the flow of payments between the parties and the ownership rights of intellectual property affect the perception of the primary purpose of each use and disclosure of PHI.

Applying the Minimum Necessary Standard

Generally, a covered entity or business associate must make reasonable efforts to limit the amount of PHI used, disclosed, or requested to the minimum necessary to accomplish the intended permissible purpose of the AI activity.[34] There is a natural friction between the Privacy Rule's minimum necessary standard and AI. Generally, the more data that AI receives, the better the AI will function. Accordingly, data scientists often will seek as much PHI as possible when developing AI algorithms.

> The million-dollar question—or maybe billion-dollar question—is whether development of AI could qualify as "health care operations" or "research."

The parties should consider what PHI really is necessary for the AI development and functioning. For example, assigning each individual a code that is not readily identifiable, rather than using a name or social security number, would promote compliance with the minimum necessary standard and would reduce the risks associated with the data set. Additionally, the parties should consider whether they could remove some fields of PHI as unnecessary.

In the end, the more PHI involved, the greater the risk under the minimum necessary standard. Accordingly, covered entities and business associates should consider documenting justification for why they deemed each data element necessary for the AI activity.

Consideration of Laws Other Than HIPAA

Of course, HIPAA is not the only game in town when it comes to privacy laws. 42 C.F.R. Part 2 governs records of federally-assisted substance use disorder (SUD) treatment programs. Most states have general privacy laws, which may be more stringent than HIPAA and, therefore, applicable. Additionally, states have laws governing certain sensitive conditions and treatments, such as HIV status, substance use disorders, mental health services, and genetic information. These laws may require consents or authorizations when HIPAA does not. Accordingly, parties involved in AI activities should consider whether to exclude certain information, such as SUD information.

As with de-identification, excluding sensitive conditions is sometimes an imperfect science. The parties should consider the level of risk and the allocation of risk (such as through indemnification provisions) should information about sensitive conditions impermissibly slip into the AI activity.

Other Privacy and Reputational Considerations

Finally, when considering applying AI to PHI, parties should weigh risks beyond legal compliance. For example, a partnership related to the use of PHI to develop AI generated widespread headlines and regulatory scrutiny in 2019.[35] The article that first publicized the partnership focused on the lack of patient consent and transparency. Yet the partnership may have fully complied with all applicable privacy laws without consent or transparency.

There is a natural friction between the Privacy Rule's minimum necessary standard and AI.

Sometimes, it is easy to fall into the trap of focusing exclusively on whether an AI activity is legally permissible. The parties also should consider the risks, including litigation and regulatory investigations, should a use or disclosure of health information for AI purposes end up on the front page of the newspaper. Although notice or consent for a specific AI activity may not be legally required, they nevertheless may be prudent means of reducing both legal and reputational risks. Accordingly, the parties should weigh the potential benefits against the burden and feasibility.

In conclusion, AI has great potential to improve health care. Covered entities can navigate privacy laws to use their health information to help achieve AI's potential in the health care sector. Compliance will not happen by accident, however, and will require careful forethought.

Any views and opinions expressed in this article are those of the author alone and should not be attributed to AHLA.

Adam H. Greene is a partner in Davis Wright Tremaine LLP's Washington, DC office, where he primarily counsels health care systems, technology companies, and financial institutions on compliance with federal and state health information privacy, security, and breach notification laws. Adam is a former regulator at the U.S. Department of Health and Human Services, where he was responsible for determining how HIPAA rules apply to new and emerging health information technologies and where he was instrumental in the development of the current enforcement process.

Endnotes

1 For purposes of this article, we use the term "AI" to encompass a suite of related technologies, including artificial intelligence, machine learning, neural networks, and deep learning. Despite the distinctions between these different technologies, we believe they raise similar legal issues with respect to privacy laws.

2 Standards for Privacy of Individually Identifiable Health Information, 45 C.F.R. pts. 160 and pt. 164, subpts. A and E.

3 Ziad Obermeyer et al., *Dissecting racial bias in an algorithm used to manage the health of populations*, 366 Science 447 (2019), https://science.sciencemag.org/content/366/6464/447.

4 Ariel Bleicher, *Demystifying the Black Box That Is AI*, Scientific Am. (2017), https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/.

5 45 C.F.R. § 164.514(b)(2)(ii) (2019); *see also* U.S. Dep't of Health and Human Servs. (HHS) Office for Civil Rights (OCR), *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,* https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html (last visited Dec. 6, 2019) [hereafter *OCR De-identification Guidance*].

6 45 C.F.R. § 164.514(b)(2)(i); *see also OCR De-Identification Guidance*, *supra* note 5.

7 45 C.F.R. § 164.502(b) (2019).

8 Cal. Civ. Code § 1798.140 (2019) (which requires certain safeguards against re-identification and that the data cannot reasonably be "linked" or "associated" with the individual, rather than only focusing on identifiability); Opinion 05/2014 on Anonymisation Techniques, Article 29 Working Party (2014), page 8, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (de-identified information typically qualifies as "pseudonymized data," which is specifically included in the GDPR's definition of personal data).

9 Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82462, 82629 (Dec. 28, 2000) [hereinafter Privacy Rule].

10 OCR, Fact Sheet, *Ransomware and HIPAA*, https://www.hhs.gov/sites/default/files/RansomwareFactSheet.pdf (last visited Dec. 6, 2019).

11 45 C.F.R. § 164.502(a) (2019).

12 45 C.F.R. §§ 164.502(a)(1)(ii) and 164.506(a) and (c)(1) (2019).

13 45 C.F.R. § 164.501 (2019) (definition of "treatment"); Privacy Rule, *supra* note 9 at 82497-98 ("Activities are considered treatment only if delivered by a health care provider or a health care provider working with another party.").

14 Privacy Rule, *supra* note 9 at 82497 ("Treatment refers to activities undertaken on behalf of a single patient, not a population.").

15 *Id.* at 82626 ("While many activities beneficial to patients are offered to entire populations or involve examining health information about entire populations, treatment involves health services provided by a health care provider and tailored to the specific needs of an individual patient. Although a population-wide analysis or intervention may prompt a health care provider to offer specific treatment to an individual, we consider the population-based analyses to improve health care or reduce health care costs to be health care operations (see definition of 'health care operations,' above).").

16 45 C.F.R. §§ 164.308(b), 164.314(a), 164.502(e), and 164.504(e) (2019).

17 45 C.F.R. § 160.103 (2019) (definition of "business associate" at (4)(i)).

18 45 C.F.R. §§ 164.502(a)(1)(ii) and 164.506(a) and (c)(1) (2019).

19 *Id.*

20 45 C.F.R. § 164.501 (2019).

21 *Id.*

22 45 C.F.R. § 164.103 (2019) (definition of "covered functions").

23 45 C.F.R. § 164.501 (2019) (definition of "health care operations" at (6)(v)).

24 45 C.F.R. § 164.502(a)(3).

25 OCR, Frequently Asked Question #544, https://www.hhs.gov/hipaa/for-professionals/faq/544/may-a-health-information-organization-de-identify-information/index.html (Dec. 15, 2008); HHS Nat'l Insts. of Health, Health Services Research and the HIPAA Privacy Rule, https://privacyruleandresearch.nih.gov/healthservicesprivacy.asp (May 20, 2005) ("a covered entity may provide a business associate that is also the de-identified data recipient with PHI, including identifiers, so that the business associate can de-identify the data for the covered entity.").

26 45 C.F.R. § 164.501 (definition of "health care operations").

27 *Id.*

28 45 C.F.R. § 164.501 (definition of "research").

29 Privacy Rule, *supra* note 9 at 82608 ("The distinction between health care operations and research rests on whether the primary purpose of the study is to produce 'generalizable knowledge.'").

30 45 C.F.R. §§ 164.508, 164.512(i), and 164.514(e). HIPAA also includes other permissions for research, such as review preparatory to research or research involving decedent information, but these are likely not applicable.

31 Privacy Rule, *supra* note 9 at 82608.

32 *Id.*

33 *See, e.g.,* Privacy Rule, *supra* note 9 at 82652 ("In some cases, a covered entity could disclose protected health information to a pharmaceutical company for research purposes if the disclosure met the requirements of § 164.512(i).").

34 45 C.F.R. § 164.502(b).

35 Rob Copeland, *Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans*, Wall St. J., Nov. 11, 2019, https://www.hhs.gov/hipaa/for-professionals/faq/544/may-a-health-information-organization-de-identify-information/index.html.