Can AI Companies Sell Your Medical Data? A Deep Dive into Privacy, Law, and the De-Identification Loophole

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: February 14, 2023 | AI Diagnostics

DOI: 10.5281/zenodo.17997614

Abstract

Can AI Companies Sell Your Medical Data? A Deep Dive into Privacy, Law, and the De-Identification Loophole The integration of Artificial Intelligence (A...

Can AI Companies Sell Your Medical Data? A Deep Dive into Privacy, Law, and the De-Identification Loophole

The integration of Artificial Intelligence (AI) into healthcare promises transformative advancements, from accelerated drug discovery to personalized diagnostics. However, this progress is fundamentally dependent on vast quantities of patient data, leading to a critical and often-asked question: **Can AI companies legally sell my medical data?** The answer is complex, rooted in the nuances of privacy law, particularly the Health Insurance Portability and Accountability Act (HIPAA) in the United States, and the controversial practice of data de-identification [1].

The Legal Framework: HIPAA and the De-Identification Standard

In the United States, the primary safeguard for patient information is HIPAA. This federal law establishes national standards to protect sensitive patient health information (PHI) from being disclosed without the patient's consent or knowledge. HIPAA applies directly to "Covered Entities" (like hospitals, doctors, and health plans) and their "Business Associates" (third-party vendors that handle PHI).

However, HIPAA contains a critical provision that allows for the commercial use of health data: the **De-Identification Rule**. Under this rule, if health information is stripped of 18 specific identifiers—including names, addresses, social security numbers, and dates of birth—it is no longer considered Protected Health Information (PHI) [2]. Once data is successfully de-identified, it falls outside the scope of HIPAA's privacy protections.

This legal distinction is the primary mechanism through which AI companies and data brokers acquire and sell medical data. Hospitals and other Covered Entities can legally sell these de-identified datasets to third-party companies, often for millions of dollars, to be used for training AI models, pharmaceutical research, and market analysis [3].

The Illusion of Anonymity: The Re-Identification Risk

While de-identification is legally sound under HIPAA, its effectiveness as a privacy protection measure is increasingly questioned by academics and privacy experts. The core issue is that de-identified data is not truly anonymous.

Research has repeatedly demonstrated that seemingly anonymous datasets can be **re-identified** when combined with other publicly available information, such as voter rolls, public records, or even social media data [4]. For instance, a combination of a patient's zip code, birth date, and gender can often uniquely identify an individual, even in a large dataset.

Furthermore, the sophistication of modern AI and machine learning techniques has made re-identification easier than ever. The standards for de-identification were established in 1996, long before the advent of modern AI and the massive, interconnected data ecosystems of today. As Stanford professor Nigam Shah notes, "We still practice de-identification as interpreted in 1996 and then complain that the removal of 18 identifiers selected years ago is not working to maintain privacy in 2021" [4]. The risk is that AI models, trained on this "de-identified" data, could potentially be used to re-identify individuals, creating a significant privacy vulnerability.

Ethical and Policy Implications for Digital Health

The commercialization of health data, even in its de-identified form, raises profound ethical questions. Patients often assume their health information is strictly confidential and are rarely aware that their medical records may be sold for profit. This lack of transparency undermines patient trust, which is the bedrock of the healthcare system.

The current legal landscape creates a regulatory inconsistency where AI developed by a HIPAA-covered entity is protected, but standalone AI applications or those developed by non-covered entities are subject only to general consumer privacy laws, which are often less stringent [5]. This sectoral approach to privacy protection struggles to keep pace with the rapid evolution of AI technology, which can seamlessly move data across different contexts.

To address these challenges, policymakers and industry leaders are exploring alternatives to the current de-identification model. These include stronger consumer consent models, data-sharing agreements that legally prohibit re-identification, and advanced privacy-preserving technologies like federated learning and differential privacy, which allow AI models to be trained without the data ever leaving the hospital's secure environment.

The debate over the sale of medical data is a crucial one for the future of digital health. It forces a necessary confrontation between the immense public good promised by AI-driven medical breakthroughs and the fundamental right to privacy. For more in-depth analysis on this topic, including the global regulatory landscape and emerging technological solutions, the resources at www.rasitdinc.com provide expert commentary.

**

References

[1] Konnoth, C. (2023). AI and data protection law in health. In Research Handbook on Health, AI and the Law*. Edward Elgar Publishing Ltd. [https://www.ncbi.nlm.nih.gov/books/NBK613196/] (https://www.ncbi.nlm.nih.gov/books/NBK613196/) [2] U.S. Department of Health & Human Services. (2024).* De-Identification of PHI*. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [3] Brodwin, E. (2025, January 16). The companies paying hospitals for troves of health data to train AI.* STAT News*. [https://www.statnews.com/2025/01/16/health-ai-electronic-health-records-hipaa-deidentified-data-market/] (https://www.statnews.com/2025/01/16/health-ai-electronic-health-records-hipaa-deidentified-data-market/) [4] Miller, K. (2021, July 19). De-Identifying Medical Patient Data Doesn't Protect Our Privacy.* Stanford Institute for Human-Centered Artificial Intelligence (HAI)*. https://hai.stanford.edu/news/de-identifying-medical-patient-data-doesnt-protect-our-privacy [5] Konnoth, C. (2023). AI and data protection law in health. In* Research Handbook on Health, AI and the Law*. Edward Elgar Publishing Ltd. [https://www.ncbi.nlm.nih.gov/books/NBK613196/] (https://www.ncbi.nlm.nih.gov/books/NBK613196/)

Rasit Dinc Digital Health & AI Research

https://rasitdinc.com