Differential Privacy Enforcement and Anomalous Access Detection: Real-Time AI Systems for Patient Health Data Privacy Management

Dr. Carlos Hernández, Associate Professor of Information Technology, National Autonomous University of Mexico (UNAM)

1. Introduction to Patient Data Privacy in Healthcare

Patient data privacy is an important issue within the medical field, as it deals with the highly sensitive topic of individual health. Making these records available to people without proper authorization can result in serious detrimental effects for the patient, including stigmatization, discrimination in employment and healthcare provision, as well as social isolation. A breach of patient records also demonstrably impacts the relationship between the patient and the healthcare professionals, leading to minimal disclosure and open communication. In light of these considerations, healthcare providers should handle patient records with the utmost care, following ethical guidelines and legal regulations to protect the rights of the patient. Given the increasing integration of digital tools in the healthcare sector, patient privacy is under increasing threat. Prior to the digital revolution, healthcare data was stored in physical files and folders as written or typed notes made by healthcare providers. This made it far less likely that such records would be disclosed to unauthorized individuals. The transfer to digitization introduced vulnerabilities such as records inappropriately left accessible on computer monitors, papers containing sensitive information left unattended, organizational security control failures, and records inappropriately disposed of within the healthcare facility. The digitization of patient data brought forward an increasing number of rules and regulations that governed the privacy and security of these records. More recent introductions, such as national-scale electronic medical record systems and personal health records, have continued to raise concerns regarding patient privacy. The inventory of privacy and data breach laws, as well as the demand to be compliant, creates a compelling case to implement effective strategies to safeguard electronic

This work is licensed under CC BY-NC-SA 4.0.

patient records. Therefore, in 21st-century healthcare, patient privacy is of paramount importance to ensure the broader public trust within the health system. Further, legal provisions require compliance for all healthcare institutions, and those that breach these regulations could face large monetary fines. Techniques such as de-identification and differential privacy are implemented to ensure privacy protection for stakeholders accessing healthcare data and involve – either contextually or explicitly – the collection and analysis of the sensitive clinical features for the purposes of an individual's healthcare requirements. A digitized medical workflow can introduce these downstream use cases and thus increase the potential negative consequences for patients when breaches occur. Therefore, in this technologically advancing world, it is important to adapt innovative sets of technological practices to ensure and safeguard patient privacy when sharing electronic data.

1.1. Importance of Data Privacy in Healthcare

Data privacy is one of the pillars of modern healthcare environments. Patients trust service providers with private matters, which must be maintained as such to allow disclosure to aid clinical activities. This poses an ethical responsibility to every healthcare establishment for the careful and shared handling of personal health data. It also corresponds with laws that mandate consent for not just data collection but also the methods through which data are stored, handled, and shared. Privacy breaches reduce patient trust in service providers, with the most common reason for losing confidence in a company being a lack of data protection. Defective data handling is punishable by reduced fines.

Patients also react relatively poorly to breaches, with a significant percentage boycotting businesses because of breaches. Both confidence loss and backlash bring forth financial loss, alongside damaging a brand. Regulatory and marketing effects are closely related. The GDPR has been implemented to protect personally identifiable information (PII) with privacy by design and privacy by default principles at its core. Compliance with the GDPR involves corporate ethical standards, reduces business risks, removes redundancies, and increases data sharing, which in turn contributes to the improvement of health outcomes. Healthcare data is continuously increasing with the advent of technology. Every device and entity in the healthcare Internet of Things collects data beyond its functional operation. This is of great use to the healthcare industry. If

This work is licensed under CC BY-NC-SA 4.0.

safeguarded, accessibility to patient data in real-time via AI-powered systems may enrich patient treatment with advancements such as predictive analytics, extracorporeal monitoring, centralized medical records, and automated robotic observation and medical activities. However, with every advancement in technology, there are new risks of unauthorized access that grow with the increasing data volume per user. Hence, allies interested in granular patient data privacy provision must increase their rapport, keeping room for lawful special purposes, processing their interference in society. These will enhance societal awareness by ensuring compliance by healthcare industry professionals in private practice, with the aid of the IT community. The general population will need periodic sensitization to recalibrate toward unapologetic disclosure assertion of their data privacy preferences. Publication is vital to maintain responsible privacy and goodwill interests already earned. The core of legitimate interest for study is satisfying oneself that health status can be inferred from the truth-telling using knowledge extraction methods, while withholding general data management, the progress achieved as well as noting the input obtained concerning research. The study, therefore, conforms to the derogatory data processing and as part of a knowledge extraction output, necessary for the performance of an AI-based healthcare system due to uneven societal input, i.e., consent according to the law of confidentiality from duly informed research subjects in the public scientific interest.

2. Overview of AI-Powered Systems in Healthcare

AI-powered systems are transforming today's healthcare. The technologies being used include machine learning, natural language processing, robotics, and others. By employing machine learning, it becomes possible to mine patterns in clinical data and develop predictive models for various diseases' early detection and also to design personalized treatment plans. Various natural language processing-based question answering systems on clinical guidelines and public health protocols have been proposed. A number of AI-powered solutions have been reported for smart hospitals for optimizing staff allocation, bed management, drug inventory, ensuring safe surgery, and disease monitoring. This modern system is anticipated to process and analyze multiple sources of information effectively and at high speed.

Importantly, AI systems can learn to process raw EHRs and help healthcare professionals in decision making. This technology can also help to conduct clinical trials

This work is licensed under CC BY-NC-SA 4.0.

using historic EHRs. AI technologies have been proven to significantly enhance diagnoses and clinical decision making. AI and robotics have helped healthcare management units to automate both real-time and non-real-time applications. One of the most interesting areas of healthcare, where AI and robotics have been playing a vital role, is telemedicine. The robots and AI systems assist in creating more personalized treatment plans. AI has a large number of advanced applications in clinical trials, such as monitoring and reporting; predictive analytics; natural language processing for fraud, waste, and abuse; and the use of EHRs in clinical trials. Nonetheless, the use of AI in healthcare data utilization also raises specific ethical and algorithmic concerns, such as algorithmic bias, which are now of particular interest to scientists. AI systems should be analyzed focusing on the quality of the real-world data as well as the algorithms used and the necessary transparency. AI cannot just be used as a tool for efficiency and practicality; it should aim for a responsive and ethical framework that enhances a patient-centered healthcare environment.

2.1. Applications of AI in Healthcare Data Management

In today's modern world, artificial intelligence is already deeply integrated into the management of healthcare data and is deployed in a range of applications. These include automating the manual administrative data tasks, such as data entry and cleansing; transforming the patient experience with advanced chatbot capabilities; and improving the clinical documentation process for the physician. Additionally, chatbots with natural language understanding and speech-to-text services provide better patient engagement and enhanced clinical documentation. Typically, AI and machine learning technologies are used to analyze large volumes of data and automatically discover insights, alerts, and predictive analytics.

In a governance and compliance scenario, particularly in privacy, the application of AI tools used for compliance training and better incident management capabilities is beginning to bring transformative advancements. Using machine learning, operational data, and accumulated clinical evidence, providers can begin to identify best practices and important care pathways and outcomes. By accessing real-time data to compare with historical data, providers and organizations can identify trends and create alerting and decision support systems. This can significantly help improve patient outcomes, reduce utilization, and manage both population health and disease surveillance. An

This work is licensed under CC BY-NC-SA 4.0.

important function of AI in clinical data management also focuses on data verification or record accuracy. Specifically, more data than necessary or inaccurate data are a direct violation of the regulatory rule of "principle of data minimization." It is significant in the context of healthcare when we consider that a significant majority of data today are more than self-reported: X-rays, blood reports, MRI images, and genomic data are all passively reported without the subject's express consent. Thus, a necessary element of data minimization is to ensure data vitality and accuracy. In such a scenario, an optimal consent-derived platform should focus on real-time data strategies. This ensures record accuracy and truthfulness.

3. Machine Learning Techniques for Data Privacy

Machine learning has the potential to enhance data privacy in healthcare by developing algorithms that offer good utility while providing robust protection of patient data. The diversity and high dimensionality of healthcare data, however, produce challenges that are often not found in other data sets. Differential privacy is meant to improve data privacy by injecting noise into responses so that individual-specific answers are not revealed. Moreover, federated learning is widely used to build models from hospitals by shifting computations from a data center to the hospitals themselves. The learning technique used requires the development of new algorithms capable of complying with patient privacy rules and complying with security needs. Additionally, the methods described must be able to work with EHR systems to develop policies respecting patient confidentiality while offering medical research new sources of analysis.

Data privacy and data utility must be balanced carefully to protect important and sensitive patient data that could be adversely affected when shared, while providing wider analysis of the data. While a large body of research focuses on these and other aims, an important tenet is the continuous development of methods that can adapt, be replaced, or continuously updated. This dynamic feature helps increase the robustness and security of systems and broadens the spectrum of data privacy protection. Current privacy protection techniques include a wide range of methods, such as data encryption and data access control. Data from most epidemiological studies have been shown to be poorly protected after encryption when shared with others. Homogenization is one of the tools data owners use to avoid attacks against the encrypted data. Cognitive and inherent bound attacks against homogeneity encryption, however, have had a potential

This work is licensed under CC BY-NC-SA 4.0.

protective weak response. The real threat of attack may be underestimated. Advances have been stymied in research on sharing tools for politics and privacy where the bystander attack is concerned. Tools of data sharing are often computationally hefty. The problem is that the time of computation is exponentially increased. Moreover, attackers will gain acknowledgment of the patients' identity and history if these tools are used alongside other pre-existing data, such as limited data available to patients. These constraints call for flexible and standardly usable enforcement procedures for the safe medical data exchange. A great deal of research is being done in order to enhance and create new methods. The problem of these procedures is that they are limited to heuristic methods or graph theory.

3.1. Anonymization and Pseudonymization

In general, two complementary methods help to protect sensitive information: anonymization and pseudonymization. Anonymization deletes personal data, i.e., individual identifiers, such as names, identification numbers, and physical characteristics from the available data. This method could bring the protected health information to a point where re-identification is not reasonably expected to occur. Anonymization makes it impossible for anyone who has access to data to link that data back to the individual. In health care, however, while the removal of certain types of information from a dataset can protect an individual's identity, it can also take away the utility of the data. Some anonymization mechanisms include de-identification, purging, complete data suppression, and randomization in order to prevent potential linkages, providing a large enough cell size within the geographic cutoff in the count numerator, and suppression of reporting.

Pseudonymization is not necessarily designed to render individuals anonymous, but to prevent the data from being usable to perform acts of unauthorized re-identification. The way that pseudonymization works is to replace identifying fields within a data record with one or more artificial identifiers, or pseudonyms. The artificial identifiers allow limited re-identification possibilities under set-controlled circumstances, and this might require access to a separate and secured system that stores the original data. The rules require identifying safeguards in order to protect privacy, confidentiality, data availability, and integrity. An accountable entity must protect patient data while ensuring it is available to an individual patient's circle of care as necessary within the

This work is licensed under CC BY-NC-SA 4.0.

health care delivery continuum. In healthcare, several implementations and solutions have been proposed with examples of successful applications.

Choosing between anonymization and pseudonymization is still a difficult choice because while the former offers greater protection for individuals, the latter has the advantage of being reversible. In each use case, a distinction should be made between the two when choosing the method to implement. In any case, it is necessary to correctly choose the method through statistical analysis to verify that the data is not recognizable, thus ensuring utility and data privacy. In addition, an important task in pseudonymization is to introduce regular artificial perturbations to data values and to satisfy data utility constraints simultaneously at the design level. Furthermore, fields that are often insufficient to identify an individual alone, such as zip codes, may be useful with others to compromise the identity of an individual. Proper use and application of these data protection disclosures must be undertaken to ensure no reasonable belief exists that an individual can be identified.

4. Compliance Regulations in Healthcare Data Management

Compliance with privacy and data protection laws and regulations is paramount to protect patient privacy in healthcare. The basis for legal and ethical standards represents the first line of defense in protecting patient privacy. Widespread discussion concerning a variety of complicated and facet-based compliance regulations can be found. Over the last half-century, regulatory compliance has been revised to include concerns of healthcare entities regarding issues of data access, data sharing between healthcare providers, the storage of data, and the handling of data by business associates working with healthcare organizations. Several laws have been passed to create standards for protecting EHRs. In 2018, a regulation was instated, which extensively details appropriate technical and organizational data protection measures expected of data controllers and processors across industries.

The regulatory requirements for handling, storing, sharing, and protecting data are extensive and detailed. Designated staff implement procedures for access, storage, handling, and use of patient and other information to ensure compliance. Audits are regularly performed. Training in data protection for employees is also required. Each of these laws gives patients the right to access and copy their healthcare data at a reasonable cost. The laws also mandate that a patient's medical health information may

This work is licensed under CC BY-NC-SA 4.0.

not be shared with third parties unless certain minimum legal requirements have been met. Penalties, including fines and incarceration, can be issued if organizations do not follow the laws that are enforced by governmental oversight agencies. Non-compliant employees can face reminders and even be dismissed.

4.1. HIPAA Regulations

Developed in 1996, the Health Insurance Portability and Accountability Act (HIPAA) is the most expansive law that strictly deals with the privacy of patient data in the United States. Title II of the law outlines policies, procedures, and implications required to maintain privacy and security in the electronic handling of a patient's Protected Health Information (PHI). Major provisions of the law include the creation of a national set of rules that should be adhered to by every healthcare organization operating in the country, which includes the Privacy Rule, the Security Rule, and the Breach Notification Rule. Protected health information and National Provider Identifier (NPI) have been defined in the act, which places an obligation on all healthcare organizations in the United States, irrespective of their size, to appropriately handle the information.

HIPAA and the HITECH Act require healthcare organizations to implement proper management policies, procedures, and practices to address the integrity of ePHI across the current healthcare ecosystem. The compliance requirements have changed, and consequently, nearly 30% of organizations do not comply with the Security Rule, necessitating a change in the culture of healthcare organizations with respect to privacy practices. This complex and challenging task is further aggravated by technological developments and the increasing integration of patient health data between caregivers and the healthcare ecosystem. The boom in technology brings with it an increasing need and practice for greater interoperability, including better and highly integrated interconnectivity between health databases, even in less traditional areas such as banking and insurance. For example, the increasing intersection of patient medical health data with a newer privacy domain – social media – is mind-boggling. Despite several recent updates, the standards, security requirements, data transmission norms, and the data to be transmitted and maintained have significantly evolved and become complex.

The substantial penalties indicate the seriousness of the United States regarding public service organizations and their transgressions. In the same year, more than 1,700 public

This work is licensed under CC BY-NC-SA 4.0.

services filed to amend their HIPAA compliance measures, and almost 372,000 contractual amendments were filed by technology partners. Moreover, over 2,000 notifications were sent to affected Americans by companies that experienced a data breach. A lack of a strong set of safeguards and controls often preludes a known issue. Similarly, newly recognized public service organizations and their technology partners dominate the march of government contracts. Public service organizations have concluded agreements for diversification through modernization and have diverse vendor contracts, which have led to the highest frequency of breaches.

5. Case Studies and Best Practices

Implementations of real-time AI-powered systems to manage patient data privacy are beginning to take shape. Two examples of companies advancing this implementation are working to ensure full compliance with regulations by developing "privacy-preserving AI" systems. In one application, the AI system analyzes chest X-rays, identifies opportunities for improvements in image quality, and the results are made available to health care professionals for review and acceptance. Of the initially proposed AI-driven computer-generated findings, only a small fraction was accepted; a significant percentage of the computer-generated findings were improved by health care professionals' results. The compliance journey for developing this system involved various lessons learned, including a need to define the incentive for change to move forward and a culture shift to develop the values, ethics, and auditability of AI work. It was found that collaboration and a mutually supportive regulatory environment combined with an enablement risk mindset for privacy-protected data was critical. The partnership between technology developers and health care providers allows for a first checkpoint to be made using state-of-the-art processing of EHR and imaging with state-of-the-art AI. The results allow for benchmarked insights to be realized from EHR reports and AI-refined evaluated chest X-rays so that subsequent predictive AI work about the quality of EHR data can go further in cost-effective and highly efficient direct-to-consumer outreach.

The opportunity to advance privacy protection, patient ownership of their own reported health care data, and empathy-based trust through loyal listening, coaching, and improving the quality of care that individuals need is a no-regret action during a time of increasingly patient-focused future policy and market protection. This made taking on

This work is licensed under CC BY-NC-SA 4.0.

these real-time privacy advances a very natural selection, as validated by health care executives. Many benefits in direct-to-consumer engagement were identified, including acquisition, longevity, increased direct pay, more frequent visits, more services performed, and testimonials that offer profound opportunities in population health. While their business did not yet benefit in a large financial manner from their AI-driven approach to privacy, other examples have occurred in the evolution of the data made available in the world's first workplace of the future focused on wired and non-wired anticipatory employee and patient safety, where there was a large and direct financial benefit. However, privacy protection while providing increasingly individualized AI-coached care was a major differentiator in health care patient engagement. There should be a mutually beneficial partnership opportunity of the same philosophy in digital rights and AI-evaluated cervical, lung, and cardiac screening as well. It is anticipated that more valuable applications will continue to be developed. Best Practices Include: Culture shift to define company values, ethics, and auditability. Continuous monitoring and responding to minimize every risk to data as much as possible. Deep stakeholder/referral partnership for AI deployment. As the data steward, the only privacy protection that stakeholders care about is the one they would want for themselves, their friends, and family, and there are expected to be differing privacy determinations. Convenience Data Ownership partners. Challenges: Corporate conflicting culture between responsible risk management and data stewardship, and a close partnership.

5.1. Real-World Implementation of AI Systems in Healthcare

AI has been implemented in healthcare settings to operationalize patient wait times at imaging centers, guide clinical workflows by organizing and displaying pertinent information about patient care, reduce medical errors, lower hospital readmission rates, and provide personalized feedback to individuals. By describing how AI was integrated into existing clinical care pathways and feature data, we provide proof of concept for newly started AI and data privacy projects. In particular, the AI discussed in this study focuses on maintaining data privacy, pooling, and maintaining patient compliance. Previous studies have emphasized the non-technical aspect of data privacy management; however, the successful implementation of features requires the cooperation and consent of other non-staff stakeholders.

This work is licensed under CC BY-NC-SA 4.0.

The difficulty of maintaining patient compliance while using AI can be seen in existing solutions that expect medical staff to flag 'atypical' recommendations made by AI. For instance, electronic health records list automated recommendations for chemotherapy on the same page as the medical team enters and confirms the patient's final diagnosis. There are two steps to implementing AI technologies: (1) practical aspects of integrating the software into existing clinical routes of care; and (2) the transformation of data privacy management. Key success factors in feature and validation AI feature model prototype—models are tested as 'proof of concept' in silico or with diagnostic cohort. The model algorithm is integrated into the approved clinical care pathway. The model is featured and data integrated into the existing diagnostic pathway; these are validated using patient physiological data. AI used in the current project primarily focused on backend operations management and intelligence, as well as patient privacy, rather than the frontend patient dashboards and patient engagement. Choice of technology: Until now, there has not been one specific AI article; other projects have used rule-based AI. Training and workshops with medical staff as lead innovators, discussing the project and receiving their endorsements. Iterative testing of integrated AI operations and extensive stakeholder buy-in.

6. Future Direction

6.1. Emerging Trends and Technologies

Advances in federated learning and other distributed machine learning approaches may provide improved protection of portions of a model at scale, while advancements in new areas of AI can allow training more effective adversaries in order to stress-test these privacy-enhanced machine learning systems. To accelerate these innovation trends and start adoption, regulatory compliance and systems should be built with an eye toward future-proofing, where proactive privacy-protecting mechanisms are tested with pioneering organizations and their data.

6.2. Adaptive Strategies

Future patient data management efforts will have to be responsive to evolving patient concerns and regulatory changes. However, every healthcare data program will rely on a similar core data protection strategy based on ethical principles that underscore respect for personal privacy. Even as technologies and approaches evolve, healthcare

This work is licensed under CC BY-NC-SA 4.0.

organizations should aim to mature into organizations with a 'culture of privacy' where ethics and privacy by design are core decision-making components.

6.3. Technology Coupling

Coupling standardization and integration of powerful technologies like federated learning with emerging technologies, such as blockchain-based secure data sharing, patient identity management, and secure data usage, has the potential to create powerful analytical platforms alongside secure, privacy-protective data containers. Many of these technologies can be initially developed independently and later integrated. The healthcare ecosystem is distributed, and a 'single solution' is unlikely to appear in isolation. A powerful and comprehensive 'best-of-breed' architecture will require robust relationships and software development across a wide range of healthcare participants. Key stakeholders include government, private industry such as healthcare systems, healthcare IT vendors, patient participants, clinical researchers, life sciences organizations, and digital health companies. Evolution in this area will depend on receptivity to, and potentially motivations from, participants across the healthcare ecosystem. Efforts to move these technologies forward may depend on other developments, such as changes to healthcare regulations. As there will always be a game of 'cat' and 'mouse' between regulations and technological innovation, the future will depend on the compatibility between regulatory change and technology change.

7. Conclusion

As discussed throughout this report, patient data privacy has always been crucial in the healthcare context. However, the growing interest in turning to more digital, remote, and virtual patient care options requires us to insist on dealing with patient health-related data. AI has the potential to change administrative processes and data management by representing an intelligence layer in the decision-making process. Moreover, there are a number of AI-related projects that leverage machine learning to ensure compliance with patient data management and privacy regulatory requirements. In addition, several electronic health records and electronic patient registration systems are already being automatically crawled and managed to implement the requirements. To move forward, healthcare organizations and researchers are recommended to strive for a proactive response to managing patient healthcare data and ensuring data privacy and protection as far as possible. Aim to educate professionals about the choices

This work is licensed under CC BY-NC-SA 4.0.

available and the choices made while managing data privacy. Highlight the gaps where data privacy settings may not be adhered to. State the whole ecosystem involved. It's not always just about technical or process rules; the human angle and the aim of the system should not be missed. Finally, to further accentuate the lines of research for the topics concerned, we also call for potential commitment from policymakers, who have looked at many technology-based solutions as part of the journey. To promote trust and patient engagement, information privacy should be a priority against health risks to ensure new healthcare innovations do not stagnate.

This work is licensed under CC BY-NC-SA 4.0.