Towards a Universal Privacy Model for Electronic Health Record Systems: An Ontology and Machine Learning Approach

informatics

MDPI

Article

Towards a Universal Privacy Model for Electronic Health Record Systems: An Ontology and Machine Learning Approach

Raza Nowrozy [1,*], Khandakar Ahmed [1], Hua Wang [1] and Timothy Mcintosh [2]

[1] College of Engineering and Science, Victoria University, Melbourne 3000, Australia; khandakar.ahmed@vu.edu.au (K.A.); hua.wang@vu.edu.au (H.W.)
[2] Department of Computer Science and Information Technology, La Trobe University, Bundoora 3086, Australia; t.mcintosh@latrobe.edu.au
* Correspondence: raza.nowrozy@live.vu.edu.au

Abstract: This paper proposed a novel privacy model for Electronic Health Records (EHR) systems utilizing a conceptual privacy ontology and Machine Learning (ML) methodologies. It underscores the challenges currently faced by EHR systems such as balancing privacy and accessibility, user-friendliness, and legal compliance. To address these challenges, the study developed a universal privacy model designed to efficiently manage and share patients' personal and sensitive data across different platforms, such as MHR and NHS systems. The research employed various BERT techniques to differentiate between legitimate and illegitimate privacy policies. Among them, Distil BERT emerged as the most accurate, demonstrating the potential of our ML-based approach to effectively identify inadequate privacy policies. This paper outlines future research directions, emphasizing the need for comprehensive evaluations, testing in real-world case studies, the investigation of adaptive frameworks, ethical implications, and fostering stakeholder collaboration. This research offers a pioneering approach towards enhancing healthcare information privacy, providing an innovative foundation for future work in this field.

Keywords: privacy; privacy policy; ontology; health information privacy; machine learning; natural language processing

Citation: Nowrozy, R.; Ahmed, K.; Wang, H.; Mcintosh, T. Towards a Universal Privacy Model for Electronic Health Record Systems: An Ontology and Machine Learning Approach. *Informatics* **2023**, *10*, 60. https://doi.org/10.3390/informatics10030060

Academic Editor: Jiang Bian

Received: 28 May 2023
Revised: 27 June 2023
Accepted: 5 July 2023
Published: 11 July 2023

1. Introduction

The growing adoption of Electronic Health Records (EHRs) has led to a significant increase in privacy and security concerns [1–5]. Despite the implementation of numerous privacy and security measures, patients' privacy continues to be compromised, often due to unreliable information-sharing methods and inadequate privacy policies [1,6–10]. High-profile data breaches in systems such as Australia's My Health Record (MHR) and the UK's National Health Service (NHS) have exposed millions of records, resulting in substantial financial losses for the healthcare industry [11]. Additionally, the expanding use of Machine Learning in healthcare for diagnostics, drug discovery, and precision medicine intensifies these concerns [12,13]. To achieve high accuracy, ML models often need to rely on analyzing vast amounts of patient data, including sensitive genetic and clinical information [14,15]. The prevalent application of ML in healthcare underscores the need to address the ethical, legal, and privacy challenges associated with implementing artificially intelligent systems (AIS) such as ML, deep learning, and Natural Language Processing (NLP) algorithms.

Context-sensitive privacy policies play a vital role in ensuring that privacy settings and access controls are meticulously adapted to the specific circumstances surrounding data [16–18]. For example, sensitive health information may necessitate more stringent privacy controls compared to less critical data [19]. Numerous privacy policies, such as the Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), and other privacy acts, regulations, and principles [20,21], have been established to address both local and global contexts. However, despite these local and international privacy policies,

existing EHR systems have experienced privacy breaches, diminishing trust in health-related IT systems [18]. Many users have opted out of systems such as Australia's MHR system. These privacy standards tend to be generic, highlighting the need for a novel privacy model that better protects patients' privacy in EHR settings.

Current strategies to safeguard EHRs involve systems that emphasize confidentiality, authentication, integrity, trust, verification, and authorization [22,23]. Intrusion Detection Systems (IDS) have been suggested to detect and categorize suspicious activities and security breaches [22,23]. However, these systems might still be vulnerable due to outdated repositories and potential alterations in patient data caused by malware or unauthorized access [24]. Privacy-preserving ML frameworks have been proposed as potential solutions, using techniques such as homomorphic encryption, secure multiparty computation, and differential privacy to protect sensitive patient information while preserving analytical accuracy [25,26]. Despite these advancements, there remains a need for more robust and comprehensive solutions to safeguard health information. As a result, additional research is necessary to address this gap, focusing on a secure and privacy-preserving health data-sharing framework within the EHR sector that considers all relevant stakeholders and ensures patients' privacy.

To address this gap, we conducted an analysis of electronic health record (EHR) policies, integrating ontologies and machine learning to enhance privacy and security controls over health data. Our focus is specifically on privacy policies with special attention to Personally Identifiable Information (PII). We introduce a machine learning model that not only classifies privacy policies as legitimate or illegitimate but also takes on the crucial role of identifying PII within these policies. This process flags potential privacy risks embedded within an organization's privacy policies. The identification of PII is of critical importance in the context of EHRs, where sensitive data are often intermixed within larger data sets. By pinpointing PII within these data sets, we can apply more precise and targeted privacy measures to the data most vulnerable to breaches. Through the prioritization of PII identification within privacy policies, we propose an additional layer of privacy protection to the existing frameworks. Our research methodology aims to provide an exhaustive exploration of this challenge, striving to contribute significantly to the enhancement of privacy and security measures in health informatics. The major contributions of this study include:

1. We presented a privacy ontology and analyzed EHR use cases to establish a standardized framework for data management, access control, and sensitive health information categorization, promoting interoperability and efficiency for healthcare stakeholders.
2. We proposed an ML-based model to identify PII from privacy policies, integrated it with the ontology for a robust medical records protection framework, and demonstrated its effectiveness in distinguishing valid and illegitimate EHR privacy policies, enhancing patient care and privacy.
3. For future research directions, we recommended conducting thorough assessments, concentrating on adaptive frameworks, ethical considerations, and implementation strategies to create a widely embraced solution for healthcare information privacy.

Our research makes significant strides in addressing the privacy issues in EHRs by innovating at the intersection of privacy ontology, machine learning, and electronic health records. The technical proposition of this paper lies in our unique approach of integrating a privacy ontology model with machine learning techniques to enhance the security and privacy of health information. This integration is manifested in the development of a new machine learning-based model that leverages the systematic organization provided by the privacy ontology to categorize sensitive health data automatically. The model was designed to efficiently identify and categorize PII from privacy policies, distinguishing between valid and illegitimate ones, thereby enhancing the privacy and security of EHRs. This fusion of machine learning with privacy ontology offers a modern strategy that extends beyond traditional privacy protection measures by providing a targeted and effective solution to the privacy concerns in EHRs.

The rest of the article is organized as follows. In Section 2, we discuss related studies and how our study addresses some of the issues that have not addressed by those related studies. In Section 3, we briefly present several application scenarios along with the research challenges. In Section 4, we propose a conceptual privacy model for EHR platforms. In realising the privacy model, we introduce a privacy ontology and its associated core and domain specific concepts in Section 4. An ML-based model is proposed in Section 5 to categorize valid versus illegitimate privacy policies. The related research is discussed in Section 6. Finally, Section 7 concludes the paper and identifies future research issues.

2. Related Work

This section discusses the associated privacy related research issues. The existing literature can be grouped into three areas: personally-controlled EHR systems, blockchain-based EHR systems, and context-sensitive privacy policies.

2.1. Personally-Controlled EHR Systems

Personal Electronic Health Record (PCEHR) systems enable individuals to manage their health information and control access. However, this also requires individuals to safeguard their data. Privacy is a crucial factor in sensitive sectors such as healthcare, and non-compliance can lead to substantial penalties. Regrettably, many large health information systems still display privacy issues and identification risks for users due to inadequate implementation of legal requirements [27,28]. Various proposals (e.g., [29,30]) have been presented to address privacy concerns in personal health records, but they frequently lack empirical evidence and real-world testing, and failed to address potential ethical and legal concerns of implementing such systems [31]. Likewise, a proposed privacy-preserving personal health record (P3HR) system lacked a comprehensive evaluation of security and performance [32], while a proposed Hippocratic database approach did not furnish empirical evidence or case studies to support its efficacy [33]. One essential aspect to consider when developing personally controlled EHR systems is striking a balance between privacy and accessibility [34]. It is critical to safeguard patients' health information privacy while ensuring that authorized healthcare providers can access the information they need to deliver effective care. Another important factor when developing personally controlled EHR systems is ensuring they are user-friendly and accessible to all patients [35,36], regardless of age, education, or technological literacy. This is challenging due to the complex nature of health information and the variety of devices and platforms used to access EHR systems. Mamum et al. [37] proposed a homomorphic encryption approach to encrypt patients' information. The decryption key will be used by the patient, ensuring no other person can access their information without prior authorization. To enhance reliability and privacy, a cryptographic verification technique is introduced to ensure that only the authorized person has access to corresponding records [38].

Privacy is a vital factor in sectors such as healthcare, banking, and defense, where confidential and sensitive data must be protected from unauthorized parties [39]. Numerous legislative rules and regulations have been introduced in European countries to ensure citizens' privacy [39,40]. Global data protection standards have been established, which outline specific data protection requirements and non-compliance penalties. According to Baker [27], patient care involves providing relevant care for individual patients based on their preferences, needs, and values, and ensuring good clinical decisions are made. This patient care includes involving, informing, and listening to patients. Due to recent digital transformations in healthcare sectors and associated data and privacy breaches, rebuilding trust in health-related IT systems has become an urgent challenge.

While personally controlled EHR systems have the potential to enhance privacy and patient empowerment in healthcare, several challenges must be addressed to ensure their effectiveness and acceptability. These challenges include balancing privacy and accessibility, making EHR systems more user-friendly and accessible, and acknowledging the cultural and social context of EHR system development and implementation. Overcoming these

challenges will require further research, collaboration, and innovation among healthcare providers, researchers, and technology developers.

2.2. Ensuring Privacy through Smart Contract—Healthcare Blockchain Systems

Blockchain-based EHR systems are increasingly gaining recognition for their potential to enhance security and privacy in managing health data. By leveraging distributed ledger technology, these systems can effectively prevent unauthorized access and data breaches. However, challenges still need to be addressed when implementing blockchain systems in healthcare, particularly when sharing patient information with multiple stakeholders.

Recent studies have explored the use of blockchain technology to improve security and privacy in healthcare IT systems. In [41], the authors proposed a consortium blockchain for secure and privacy-preserving data sharing in e-health systems. Although their study provided an in-depth description of the proposed architecture and its benefits, it lacked empirical evidence and real-world evaluations, and did not discuss potential limitations or challenges associated with implementing such a system. In [42], the study examined the applications of blockchain distributed ledger technologies in biomedical and healthcare settings [42]. While the authors thoroughly reviewed existing literature and proposed various use cases, the study was published in 2017, and blockchain technology has evolved significantly since then. Moreover, the authors did not address potential drawbacks or limitations of using blockchain in healthcare settings. In [43], the authors focused on the potential of blockchain technology for improving security and privacy of healthcare data stored in the cloud. The authors provided a comprehensive overview of the challenges and explained how blockchain could address them. However, the article did not critically evaluate the technology's limitations and challenges, such as scalability and interoperability issues. In [44], the authors proposed a blockchain-based incentive mechanism for privacy-preserving crowd-sensing applications. Despite presenting an interesting idea, the paper lacked sufficient detail on technical implementation and evaluation and did not compare the proposed mechanism to existing solutions or discuss limitations or future work. In [45], the authors introduced a blockchain-based solution called Medblock for efficient and secure sharing of medical data. The authors claimed that their system could overcome traditional centralized data storage limitations but they did not provide a comprehensive evaluation of the proposed system's scalability and efficiency or detailed information about its implementation. Finally, in [46], the authors proposed a healthcare blockchain system using smart contracts for secure automated remote patient monitoring. While the authors presented a detailed description of the proposed system and a theoretical analysis of its security and privacy features, they lacked empirical evidence to support the system's feasibility and effectiveness and did not address potential challenges in implementing the system in a real-world healthcare setting.

While blockchain-based EHR systems can offer significant benefits in terms of security and privacy, there are still challenges and limitations to be addressed, especially when sharing patient data with multiple stakeholders. Further research, validation, and critical analysis are needed to ensure the practicality, scalability, and effectiveness of these systems in real-world healthcare scenarios.

2.3. Context-Sensitive Privacy Policies

In recent years, there has been a growing interest in context-sensitive approaches within the EHR domain. In [47], the paper presented a context-aware access control model for cloud-based data resources, incorporating imprecise context information. The authors utilized fuzzy logic to model the uncertainty in context information and developed a context-aware access control framework. However, the paper did not comprehensively evaluate of the proposed model, including comparative analysis with other state-of-the-art approaches, scalability, and performance testing. Additionally, there was no mention of any practical implementation of the proposed framework in real-world settings. While the proposed approach seemed promising, the lack of evaluation and practical implementation

made it difficult to assess its effectiveness and feasibility. In [48], the article introduced a policy model and framework for context-aware access control to information resources. Their model integrated contextual factors such as user identity, location, and time to determine access privileges. However, it lacked empirical validation of the proposed framework, leaving its effectiveness in real-world scenarios uncertain. Moreover, the article did not address potential ethical implications of context-aware access control, such as privacy and discrimination concerns. Further research and analysis are required to address these issues. In [49], the article proposed a fog-based context-aware access control (CAC) system to achieve security scalability and flexibility. The authors argued that their system could enhance security in fog computing environments by providing dynamic and context-aware access control. The article offered a comprehensive overview of the proposed CAC system and discussed its implementation details. However, the article lacked empirical evaluation of the proposed system's performance and scalability. Additionally, it did not address the potential challenges and limitations of implementing such a system in real-world scenarios. Overall, the proposed system appeared promising, but further research is necessary to validate its effectiveness and practicality. In [50], the paper suggested an ontology-based approach for dynamic contextual role-based access control in pervasive computing environments. The authors described the architecture of the proposed system and evaluated its effectiveness through simulations. Nevertheless, the evaluation of the system was limited to simulations, and a real-world implementation and evaluation of the approach would be advantageous. Additionally, the paper could benefit from a more in-depth discussion of related work in the field of contextual role-based access control.

To summarize, while these context-sensitive approaches have made strides in proposing enhanced protection for EHRs, they have proven insufficient for accurately modeling relevant stakeholders and health information.

2.4. Homomorphic Encryption in EHR Systems

The role of homomorphic encryption in preserving the privacy of EHRs has been explored in various studies, which have claimed that the approach offers computation on encrypted data without necessitating decryption, effectively facilitating secure data sharing and collaboration. Paul et al. [51] constructed a privacy-preserving framework, leveraging homomorphic encryption for protecting EHRs during collaborative machine learning processes. Although the proposed framework held potential, the study did not sufficiently address the framework's limitations, including potential vulnerabilities of the encryption scheme, scalability, and maintaining confidentiality during collective learning. Ikuomola et al. [52] addressed privacy concerns in e-health clouds using homomorphic encryption and access control. However, the research was marked by the absence of a detailed analysis of the solution's effectiveness. Furthermore, potential vulnerabilities or attacks that could undermine the security of the proposed system, and scalability issues related to large-scale e-health cloud environments were not adequately addressed. Vengadapurvaja et al. [53] developed an efficient homomorphic medical image encryption algorithm for secure medical image storage in the cloud. Despite its focus on medical images, the approach did not extend to the encryption of other types of EHR data. This narrow scope limited its comprehensive application to broader EHR privacy concerns. Alzubi et al. [54] integrated homomorphic encryption with deep neural networks to secure the transmission and diagnosis of medical data. However, unspecified inadequacies were identified in preserving the privacy of EHR. A thorough examination of the study would provide a better understanding of these limitations. Subramaniyaswamy et al. [55] implemented a somewhat homomorphic encryption scheme for IoT sensor signal-based edge devices. However, without detailed insights from the paper, it is difficult to identify specific inadequacies in preserving EHR privacy. Potential challenges could include scalability, performance, or vulnerability of the implemented scheme when applied to real-world EHR systems. Finally, Vamsi et al. [56] investigated various homomorphic encryption schemes for securing EHR in the cloud environment. Despite potential benefits, several inadequacies were noted in the application of homomorphic encryption for

preserving EHR privacy. Challenges, such as the performance overhead of homomorphic encryption, integration difficulties with existing healthcare systems, and the need for efficient key management strategies, were some identified concerns.

While various studies have explored the role of homomorphic encryption in preserving the privacy of EHR, each presents certain inadequacies. Key among these are the vulnerability of the encryption schemes employed, limitations in scalability, difficulties in maintaining the confidentiality of sensitive data, and the substantial computational overhead that their encryption techniques have introduced. Furthermore, a narrow focus on specific data types, such as medical images, excludes comprehensive coverage of EHR privacy concerns. Challenges in integrating homomorphic encryption schemes into existing healthcare systems, including issues of interoperability, data access control, and key management strategies, further compound the problem. This study aimed to address these shortcomings by proposing a novel privacy-preserving approach for EHRs, leveraging the benefits of homomorphic encryption while addressing its limitations. We sought to develop a robust, scalable, and versatile homomorphic encryption scheme that can safeguard various types of EHR data. Our methodology focused on ensuring efficient performance, facilitating secure data sharing, and improving integration with existing healthcare systems. Furthermore, we will offer solutions for effective key management, ensuring a holistic and comprehensive approach to preserving EHR privacy.

2.5. Comparison with Our Study

Our study acknowledged that several attempts have been made to tackle privacy and security issues within the realm of EHRs. These efforts, while crucial, often exhibit certain shortcomings. For instance, they frequently fail to consider concerns raised by previous research and lack comprehensive and robust evaluations of their effectiveness, scalability, and process efficiency. Additionally, these studies sometimes focus too narrowly on specific solutions such as homomorphic encryption or context-sensitive privacy policies, overlooking the need for more holistic and comprehensive strategies that can navigate the complexities of modern healthcare systems. Furthermore, the scalability of these solutions—especially when implemented in larger and more diverse healthcare systems—often remains inadequately explored. Another under-addressed concern pertains to the process efficiency of these proposed solutions. In the high-stakes, fast-paced environment of healthcare, solutions that are computationally demanding or overly complex may not be feasible, despite their theoretical advantages. Our added perspective does not devalue the existing body of work. Instead, it illuminates the multifaceted nature of privacy and security in healthcare data management. Our research aspired to address these challenges through an approach that balances privacy protection, process efficiency, scalability, and real-world applicability. We aimed to build on these prior efforts, incorporating their strengths while also striving to rectify their shortcomings.

In aiming to apply an ontology and ML-based approach to protect health information, our study sought to explore new solutions to the challenges in this domain. The distinguishing aspects of our research are:

- ✓ An Attempt at Comprehensive Privacy Protection: Our approach endeavored to create a privacy protection solution that is more robust and specific to healthcare information systems by combining ontology and ML. While we believe it can offer improved protection, further studies and real-world testing are necessary to validate this claim.
- ✓ Consideration of Legal Requirements: We made an effort to incorporate the breadth of GDPR and other privacy regulations into our proposed solution, aiming to ensure compliance with the necessary legal requirements. However, adapting to evolving legal landscapes will require continuous updates and adjustments.
- ✓ Exploration of Balancing Privacy and Accessibility: Our proposed solution attempts to balance the preservation of patients' health information privacy with the necessity of access for authorized healthcare providers. Future studies should focus on how well we have achieved this balance in various real-world scenarios.

- ✓ Aim for User-friendly and Accessible Systems: We acknowledged the complex nature of health information and the diverse range of devices and platforms used to access EHR systems. Our study aimed for a more inclusive approach to healthcare information management, although the user-friendliness of our solution remains to be evaluated in diverse patient groups.
- ✓ Emphasis on Real-world Implementation and Evaluation: We strove to provide solutions that can be implemented and evaluated in real-world healthcare settings. Our approach takes a practical perspective, though extensive empirical evidence to support its effectiveness is yet to be collected.
- ✓ Acknowledgment of Cultural and Social Context: We considered the cultural and social context of EHR system development and implementation in our study, aiming for a solution that can cater to diverse needs. However, further research is required to confirm the adaptability of our solution to different cultural and social contexts.

Our study offers an exploration into combining ontology and ML for safeguarding privacy in healthcare information systems. By acknowledging the importance of factors such as legal compliance, balancing privacy and accessibility, creating user-friendly systems, real-world implementation, and considering cultural and societal aspects, we strove to extend the knowledge in the field. However, it is important to note that our proposed solution is a preliminary attempt and further validation through future research is needed. The potential impact of our study is in providing new perspectives and suggesting areas of focus for ongoing exploration in the field of healthcare information management.

3. Research Motivation

In this section, we delve into a variety of application scenarios, examine them, and highlight research challenges that remain unaddressed in the current EHR privacy literature. Today, privacy is a vital concern in cybersecurity, and safeguarding patient data is essential by implementing robust EHR privacy and security policies on both national and international levels [57].

To further strengthen our motivation for this research, we reflected upon real-world examples that underline the privacy and security challenges currently plaguing the realm of health information sharing. For instance, the considerable data breach at Anthem Inc. in 2015, where hackers gained unauthorized access to the personal information of nearly 78.8 million individuals [58], showcases the vulnerabilities of large-scale health data systems. Despite the robust security measures in place, the attackers were able to exploit weak points in Anthem's system, leading to a catastrophic loss of privacy for millions of people. Such breaches elucidate the crucial need for improved privacy protection mechanisms, specifically those that are capable of safeguarding Personally Identifiable Information (PII) against increasingly sophisticated forms of cyberattacks.

Moreover, real-life examples can also provide insight into the efficacy of existing privacy regulations in the face of evolving technological landscapes. For example, consider the case of the UK's National Health Service (NHS) in 2018, when it was discovered that third-party organizations were purchasing anonymized patient data for market research [59]. Despite adhering to existing privacy regulations, the anonymization techniques employed failed to prevent the re-identification of individual patients from the purchased data, leading to serious privacy concerns. These incidents not only demonstrate the importance of our research but also highlight the urgent need for an integrative approach that combines machine learning with ontologies to secure EHRs effectively. Our proposal aimed to identify and protect PII within privacy policies, adding an extra layer of security to existing frameworks. Such a solution can potentially prevent future privacy violations, particularly those related to the re-identification of anonymized data, thereby ensuring the integrity of patient information in digital health platforms.

An appropriate privacy model is needed to allow patients control over their own data, and it could also facilitate the tracking of who has accessed their health information and the parties to whom it has been sent [60]. As mandated by the Australian Privacy Principles

(APP) 1988 (https://www.oaic.gov.au/privacy/australian-privacy-principles (accessed on 6 July 2023)), patients should be informed about the data being collected and the manner in which their personal health information is being utilized. While visiting hospitals or clinics, patients also need to be notified about the reasons behind collecting and using their data, the duration of data retention, and the parties with whom it will be shared. According to NHS England, centralizing health information at a national level is crucial. When a General Practitioner updates a patient's registration information in their clinical system, the Primary Care Support England (PCSE) system leverages this information to update the National Health Application and Infrastructure Services (NHAIS), responsible for maintaining the National Patient Register. The Royal Australian College of General Practitioners (RACGP) has also introduced a sample registration form for new patients [61]. To comply with federal and state privacy laws, this form aligns with the RACGP Standards for general practices. If patients harbor privacy concerns, they can discuss them with their GP and opt to leave the form blank. However, it is not considered best practice to let patients leave the form blank, as the information may be crucial at any stage of their treatment, and a lack of data could result in improper treatments.

Various individuals or groups, including health-related and non-health stakeholders, may engage with EHR systems. Stakeholders can be categorized and arranged differently based on their roles in managing EHR records (Table 1). In the subsequent paragraphs, we will explore several such scenarios.

Table 1. List of role-based stakeholders and privacy rules.

Stakeholders' Category	Stakeholders' Example	Role-Based (Senior/Junior)	Privacy Roles
Support Professionals	Receptionist, Pharmacist	Junior	Policy 1: Support professionals can only deal with personal health information
Nursing Professionals	Nurse Manager, Nurse	Junior	Policy 2: Nursing professionals can deal with both personal and private health information
Medical Practitioners	General Practitioner, Specialist	Senior	Policy 3: Medical practitioners can deal with personal, private, sensitive, and historical health information
Diagnosis Professionals	Radiologist, Medical Technician	Senior	Policy 4: Diagnosis professionals can deal with personal and historical health information
Medical Scientists	Researcher, Junior Researcher	Junior	Policy 5: Medical scientists can deal all types of health information with the approval of relevant stakeholders

3.1. Use Case Scenarios

In order to establish a comprehensive privacy ontology, the current scenarios must be enhanced and diversified, taking into account the multifaceted reality of healthcare organizations. The following use case scenarios span a range of situations, each with differing stakeholders, types of health information, and privacy concerns. These cases, while varied, represent a snapshot of the highly complex landscape of privacy preservation in the context of EHRs.

3.1.1. Scenario 1: Primary Care Physician

An elderly woman, living with her eldest son, is struggling with her mental health after witnessing the sudden death of her youngest grandson. As she displays signs of distress and paranoia, her son seeks the help of a primary care physician (PCP). After an examination, the PCP suggests consultation with a mental health professional. In this

scenario, the privacy concerns pertain to the sensitive nature of the woman's mental health status and her medical history.

- Stakeholders: PCP, patient, patient's son, and mental health professional.
- Health Information: Private (address, location), Personal (demographic details), Sensitive (medical and psychological history), Historical (previous health assessments).

3.1.2. Scenario 2: Emergency Care

Following a serious car accident, an unconscious patient is rushed to the Emergency Department (ED) where a nurse assesses him. The patient's Medicare details are used to gain access to his medical history to determine the best course of urgent care. The privacy concern here relates to the patient's inability to provide consent for access to his medical records due to his unconscious state.

- Stakeholders: Nurse, patient, medical team, and Medicare.
- Health Information: Private (address, location), Personal (demographic details, Medicare details), Sensitive (medical history, accident details).

3.1.3. Scenario 3: Clinical Research

A breast cancer patient undergoing radiotherapy expresses concerns to her GP about her family history, particularly since her mother died of a brain haemorrhage. Her GP consults with a research team to access clinical trials data and explore the prevalence of similar cases. The patient's personal information, medical history, and family history need to be handled discreetly due to the sensitive nature of her condition and personal fears.

- Stakeholders: Researcher, GP, patient, clinical trials team.
- Health Information: Private (address, location), Personal (demographic details), Sensitive (medical history, family history), Historical (previous treatment records).

3.1.4. Scenario 4: Multi-Disciplinary Consultation

A patient with a rare genetic disorder requires consultation with a multi-disciplinary team, involving primary care physicians, specialists, therapists, and social workers. The complexity of the case necessitates sharing extensive patient data across the team while ensuring the patient's privacy.

- Stakeholders: Primary care physicians, specialists, therapists, social workers, patient.
- Health Information: Private (address, location), Personal (demographic details), Sensitive (medical and genetic history), Historical (treatment and therapy records).

3.1.5. Scenario 5: Telehealth

A remote patient is receiving care via a telehealth platform. The patient's electronic health records need to be accessed and updated by healthcare providers during virtual consultations. The privacy concerns here arise due to the potential vulnerabilities associated with the transfer of sensitive health data over digital channels.

- Stakeholders: Patient, healthcare providers, telehealth platform provider.
- Health Information: Private (address, location), Personal (demographic details), Sensitive (medical history), Historical (previous consultation records).

3.1.6. Scenario 6: Data Breach

A healthcare organization experiences a data breach and the EHRs of multiple patients are potentially compromised. This scenario raises significant privacy concerns related to the unauthorized access and potential misuse of health data.

- Stakeholders: Patients, healthcare organization, IT department, potentially unauthorized third parties.
- Health Information: Private (address, location), Personal (demographic details), Sensitive (medical history), Historical (previous treatment records).

These enhanced scenarios provide a broader understanding of the intricate landscape of privacy preservation in EHRs and should provide a strong foundation for the development of a comprehensive privacy ontology. The complexity and diversity of these scenarios reflect the dynamic nature of healthcare delivery and the myriad of privacy concerns that arise in real-world healthcare environments.

3.2. Research Challenges

In the context of the GP and researcher scenarios previously discussed, we have recognized several challenges that must be tackled when developing a universal privacy ontology for various EHR platforms, such as MHR and NHS systems. Addressing these challenges is crucial to ensure the privacy and security of sensitive patient information while enabling seamless data exchange across different EHR platforms:

1. Sharing personal information with relevant stakeholders: According to APP 6 [21], personal information should only be shared with pertinent stakeholders for a specified purpose, such as treatment or daily care with GPS or nurses. Personal health information may also be shared for secondary purposes under certain conditions. However, discerning these secondary purposes within the existing EHR literature proves to be a substantial research challenge. This difficulty extends to sensitive health information as well. While the APP principles are generic and can be applied to any domain, implementing these principles within the health information domain is particularly challenging. As a result, the development of a privacy ontology for EHR systems is necessary.
2. Identifying relevant stakeholders: One of the detailed challenges involved in building the privacy ontology includes identifying the relevant stakeholders associated with different EHR systems. These stakeholders can range from healthcare providers, insurance companies, and government agencies to patients themselves. A privacy ontology should be able to accommodate these various stakeholder groups and their respective access levels, ensuring that sensitive patient information is only accessible to those with appropriate authorization.
3. Categorizing different levels of health-related patient information: Another challenge is to categorize different levels of health-related patient information, which may range from general health indicators to highly sensitive data, such as genetic test results or mental health records. Creating a privacy ontology that can effectively classify this information is essential for implementing appropriate access controls and maintaining patient confidentiality.
4. Defining privacy rules and policies: The privacy ontology should also define privacy rules and policies for relevant health-related stakeholders, allowing them to share patient health records across different EHR platforms securely. These rules and policies should be robust, flexible, and adaptable to accommodate the diverse and evolving needs of different healthcare systems and their stakeholders.

By addressing these challenges, a universal privacy ontology for EHR platforms can be developed, providing a framework for ensuring the privacy and security of patient information while facilitating interoperability and collaboration among healthcare stakeholders. Such an ontology will ultimately enhance the efficiency and effectiveness of healthcare delivery, benefiting both patients and providers.

4. A Privacy Model for EHR Systems

This section introduces a privacy model for EHR systems to address the research challenges identified in Section 2.

4.1. Leveraging Ontology and ML for Enhanced e-Healthcare Privacy

Recent years have witnessed a growing emphasis on the connection between privacy ontology and ML in the context of e-Healthcare systems. These approaches have been employed to improve various aspects of healthcare systems, such as intrusion detec-

tion, confidentiality, and privacy of EHR. In this section, we will discuss the key themes surrounding ontology and ML in e-Healthcare systems.

4.1.1. Intrusion Detection and Prevention

Sreejith and Senthil's research [62] proposed a model for detecting intrusion attacks based on a NoSQL database and semantic features. This model highlights the role of ML in detecting and preventing real-time intrusion attacks in healthcare systems.

4.1.2. Confidentiality and Privacy of EHR

In [63], their research focused on an ontological framework designed to enhance the confidentiality and privacy of EHR . Their framework aimed to detect anomalies in abnormal patterns of healthcare record access, predict vulnerable healthcare records for prioritized security efforts, and analyze stakeholders' behavior to detect suspicious activity.

4.1.3. Improved Indexing and Retrieval Performance

ML and ontology-based techniques have been shown to enhance the effectiveness of indexing processes and retrieval performance in various studies [64,65]. For instance, a framework for smart e-healthcare systems employs IoT technology while maintaining privacy and authentication through a combination of encryption, secure authentication protocols, and Blockchain technology.

4.1.4. Secure Data Access and Privacy Preservation

Sun et al. [66] explored a bilateral fine-grained access control mechanism in cloud-enabled industrial IoT for healthcare, using Blockchain-based frameworks for granular access control, secure data access, and privacy preservation.

4.1.5. Privacy Disclosure Measurement

Research on ontology-based approaches for protecting personal information in online privacy policies suggests that these models offer a standardized and objective way to measure privacy disclosure [67]. Privacy-preserving ontology is analyzed through various stages, including data collection, data publication, and output regarding modeling and training.

4.1.6. Efficiency and Accuracy in Clinical Information Extraction

Studies such as Yehia et al.'s [68] have demonstrated increased efficiency and accuracy in extracting clinical information from free-text notes written by physicians using ML-based approaches.

4.1.7. Structured Approach to Organizing Clinical Data

Bosco et al. [69] illustrated that ontologies can provide a structured and standardized approach to organizing clinical data, supporting interoperability between EHR systems and ultimately improving patient care and facilitating clinical research.

To summarize, incorporating ontology-driven strategies alongside machine learning in EHRs yields considerable benefits in terms of privacy, security, and efficacy. The utilization of ontology-focused techniques allows for the creation of a standardized and structured framework to organize and handle health data, leading to enhanced interoperability, information exchange, and decision-making. Our research distinguishes itself by tackling various obstacles such as intrusion detection, EHR confidentiality, optimized indexing and retrieval performance, secure data accessibility, privacy conservation, and clinical data extraction. This all-encompassing approach guarantees the protection of sensitive health data and empowers healthcare practitioners to deliver superior patient care.

By persistently examining and refining these methodologies, we can further progress electronic healthcare systems and contribute to the creation of more secure, privacy-focused, and effective solutions for managing delicate health information. Consequently, our

ontology-centered method, in conjunction with machine learning tactics, possesses immense potential in securing health data and boosting overall patient care results.

4.2. Conceptual Privacy Models

We identified the concepts behind the privacy model for different EHR systems. A brief description of the concepts is presented in Figure 1, which can be titled the identification layer. It shows the different types of stakeholders (e.g., GP, nurse, policy maker), technologies (e.g., the digital platforms to interact with relevant stakeholders) and health information (e.g., personal, sensitive) involved in the EHR scenarios.

- Identify stakeholders: the relevant stakeholders need to be identified from the application scenarios.
- Identify technologies: the relevant health technologies and platforms need to be identified to gather health related information.
- Identify health information: the health information and records need to be identified and labelled in different categories, such as private record, sensitive record, and so on.

4.3. Identifying Stakeholders

We analyzed various EHR scenarios, including those for receptionists, nurses, policy makers, and media personnel, as well as those outlined in Section 4.2. Based on our findings, we have identified several stakeholders associated with EHR platforms, which we have listed in Table 1.

To simplify the privacy model and utilize the inheritance concept of context-aware role-based access control systems (as described in [49,70]), we classified EHR stakeholders into two categories: primary and secondary stakeholders. Primary stakeholders, such as GPS and nurses, are directly involved with EHR platforms and have access to patients' health information. Secondary stakeholders, such as media personnel, use health information without being directly involved with EHR systems. We also defined two types of roles for different stakeholders to create a privacy model applicable across different stakeholder types. These roles include senior roles and junior roles. Table 1 shows the role inheritance relationship among stakeholders, where top-level stakeholders, such as medical practitioners and medical scientists, are classified as senior roles, while bottom-level stakeholders, such as GPs and researchers, are classified as junior roles. This approach helps ensure that access to sensitive patient information is appropriately controlled and that privacy is maintained for all stakeholders involved with EHR systems.

4.4. Redefining Health Information and Privacy Rules

In response to valuable feedback and mindful of evolving privacy legislation such as the GDPR and several national laws, we revisited our initial classification of health information and redefined them as follows:

- Identifiable Health Information: This includes any data that can be used to identify an individual, such as name, address, and location. This is akin to what was previously described as 'private health information' and 'personal information'. Under the GDPR and similar laws, all health information is considered sensitive, hence our shift towards a unified category.
- Health Record Information: This encompasses the clinical details of a patient's health condition and medical history. It includes past diagnoses, treatment records, and other medical reports. This category is more compatible with current legislation and consolidates what we previously classified as 'sensitive information' and 'historical information'. The important distinction here is that this information is sensitive by its very nature and is treated as such under GDPR and similar laws.

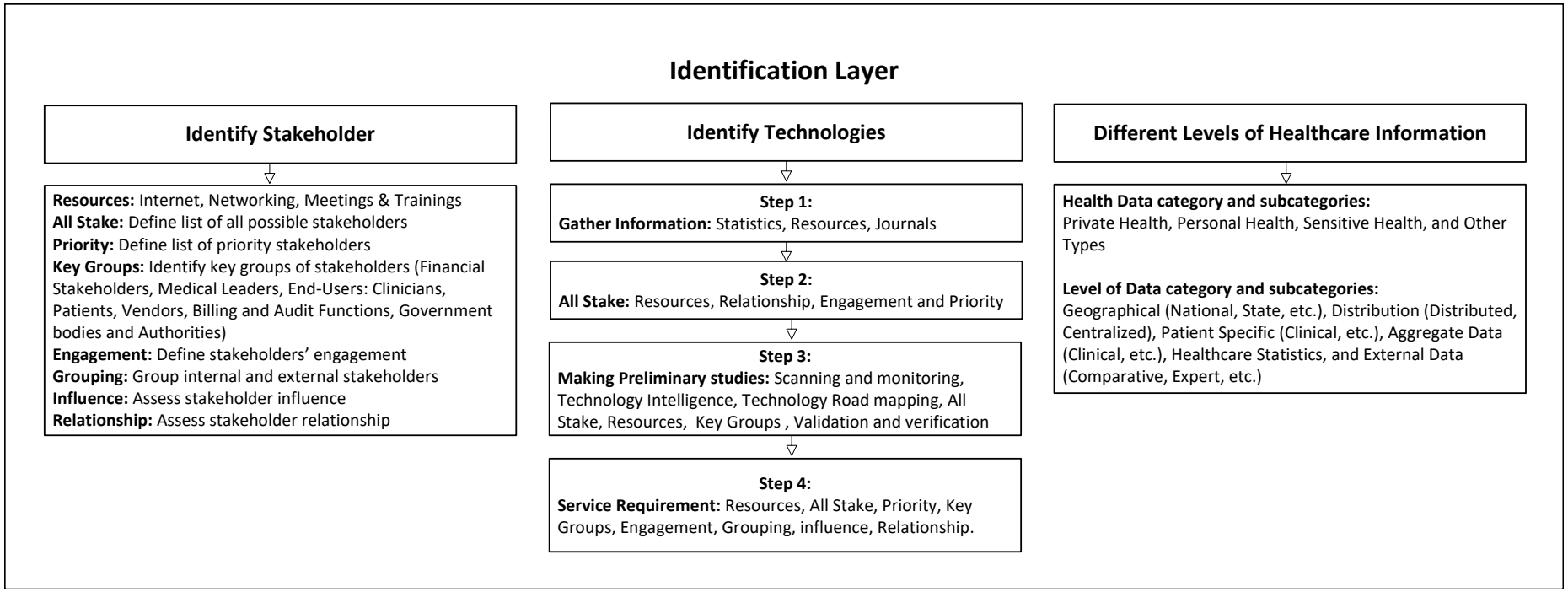

Figure 1. The relevant concepts to build the privacy model.

With this redefinition, we present an updated privacy ontology and its associated concepts, including a role ontology, health information ontology, and privacy policy ontology. These concepts, which are grounded in the revised privacy model introduced earlier, can be viewed in Figures 2–5. These modifications ensure that our ontology aligns with current legislation, making it more applicable to the complex healthcare and health-related information landscapes.

Our ontology and knowledge bases were defined using the widely used Web Ontology Language (OWL) [71], specifically Prot'eg'e OWL 5.5 (https://protege.stanford.edu (accessed on 6 July 2023)). We employed an object-oriented approach to model diverse stakeholders, health information, and privacy rules. It includes classes, subclasses, datatype, and object type properties. The health-related stakeholders were defined as classes and subclasses (i.e., primary and secondary stakeholders) and their relevant properties were defined using class-to-class object type and datatype properties.

Table 2 provides a technical description of our updated privacy ontology. It comprises three core concepts: role ontology, health information ontology, and privacy policy ontology. This revised framework reflects a more accurate portrayal of the current healthcare landscape and the information contained within EHRs, making it more suitable for privacy preservation.

Table 2. The modeling criteria of the ontology.

Basic Modeling Criteria	Privacy Ontology Elements
Classes	All primary stakeholders are represented as senior roles.
Sub classes	All secondary stakeholders are represented as junior roles.
Object-type Properties	The relationships between classes to classes are represented as object type properties.
Datatype Properties	The relationships between classes to their features are represented as datatype properties.

Our updated ontology provides a better foundation for dealing with the intricate healthcare and health-related information environments. The identifiable health information and health record information better adhere to the GDPR and other national laws, providing a robust framework for data protection. With this new approach, the ontology can better reflect the complex and diverse processes and results found in healthcare systems, ensuring that EHRs remain secure and patients' privacy is respected.

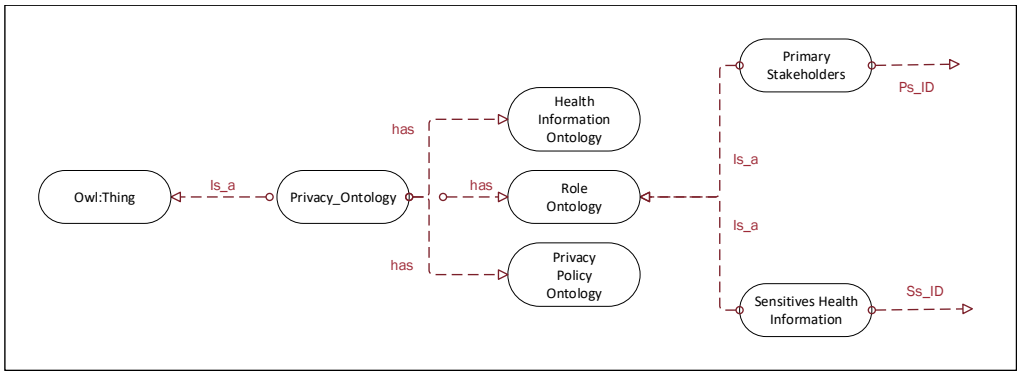

Figure 2. The core concepts of privacy ontology.

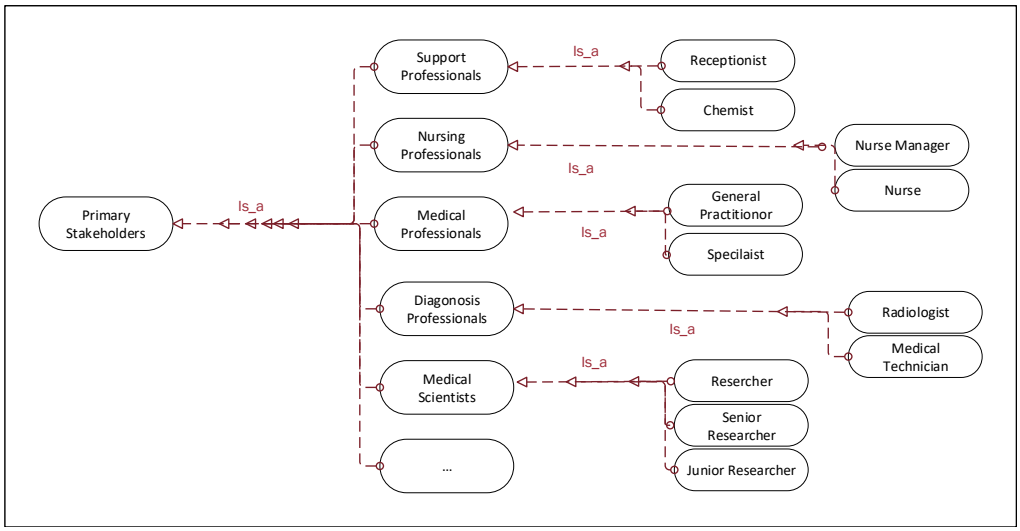

Figure 3. Role ontology.

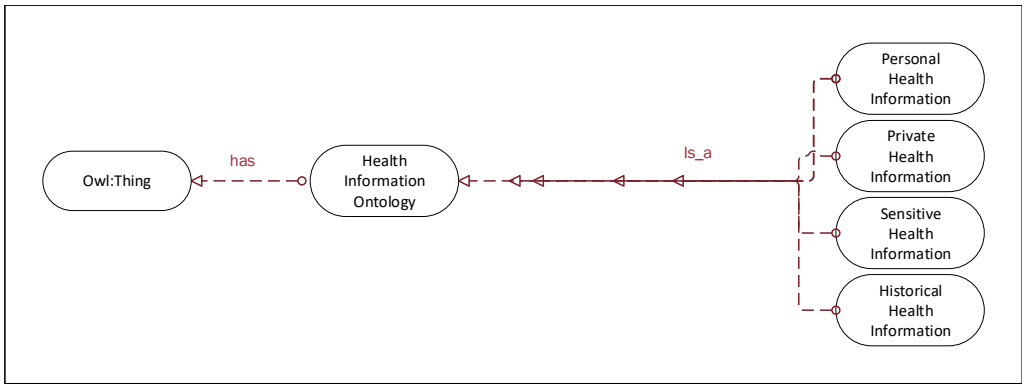

Figure 4. Health information ontology.

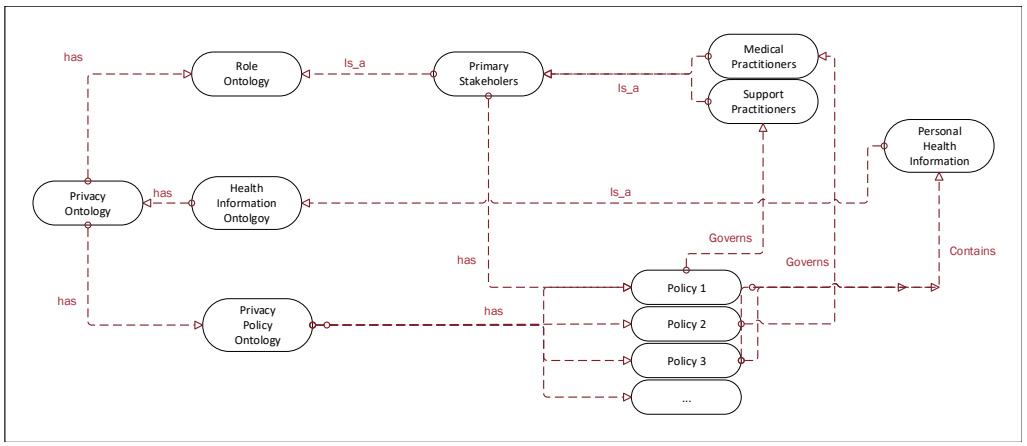

Figure 5. Privacy policy ontology.

4.5. Role Ontology

In this section, we will introduce a snapshot of the role ontology, which is based on the different stakeholders associated with EHR environments. Figure 2 shows the core concepts of the privacy ontology, which consists of three parts: Role Ontology, Health Information Ontology, and Privacy Policy Ontology. The classes Primary Stakeholders and Secondary Stakeholders are both subclasses of the Role Ontology class. The "Is _*a*" property indicates the relationship between the Role Ontology class and its subclasses. Each class

and subclass is defined by its datatype properties, such as the Primary Stakeholders class having a datatype property "Ps_*ID*".

Primary stakeholders, such as healthcare professionals, can directly access patients' health information through EHR systems, whereas secondary stakeholders, such as media professionals, can use some health information without being directly involved with EHR systems. Figure 3 provides a snapshot of the primary stakeholders. For the purpose of this study, the Role Ontology (see Figure 3) includes five key domain-specific classes: Support Professionals, Nursing Professionals, Medical Practitioners, Diagnosis Professionals, and Medical Scientists. Each class has its own set of subclasses; for example, the Nursing Professionals class has Nurse Manager and Nurse subclasses. The proposed privacy ontology can be expanded to include new classes and subclasses.

Understanding the integral role of ontology in enhancing data privacy in EHRs is paramount. In our proposed Role Ontology, classes such as 'Primary Stakeholders' and 'Secondary Stakeholders' elucidate who has authorized access to the EHRs and the extent of that access. In scenarios where a healthcare professional (primary stakeholder) accesses patient data directly from EHRs, or a media professional (secondary stakeholder) utilizes some health information without direct EHR involvement, it is the defined datatype properties such as 'Ps_*ID*' that regulate this access. Such a structured hierarchy of data access, based on well-defined classes and subclasses, ensures only authorized access to sensitive patient data, thereby enhancing data privacy. Additionally, our ontology model provides flexibility in expanding to include new classes and subclasses as required, ensuring the model's scalability and adaptability to growing and diversifying healthcare data needs. Through our role ontology, we aimed to build a robust privacy-preserving framework where the right to access and the extent of that access is pre-defined based on the role of the stakeholder. This approach considerably mitigates unauthorized access, reduces privacy breaches, and promotes data confidentiality in EHR environments.

4.6. Health Information Ontology

In this section, we present an overview of the health information ontology, which is based on various medical information types that are found in EHR systems. Our privacy ontology proposal includes classes and subclasses of relevant health information. The core concept of the Health Information Ontology consists of domain-specific concepts such as Personal Health Information, Private Health Information, Sensitive Health Information, and Historical Health Information. The different types of health information are represented as sub-classes, and are linked to the core concept of the Health Information Ontology using an "Is_*a*" relationship, which is an object type property. You can see a snapshot of the health information ontology in Figure 4.

4.7. Privacy Policy Ontology

In this section, we take a look at the privacy policy ontology, which is based on the privacy rules we identified in the previous section. Figure 5 presents a snapshot of the privacy policy ontology, which also includes domain-specific concepts from EHR environments. To illustrate, based on Policy 1, Policy 2, and Policy 3 (as shown in Table 1), all medical practitioners are authorized to access patients' health information. Another example is that Policy 1 allows support professionals to access patients' personal health information but not their entire medical records.

4.8. Disclosing Emergency Health Information for Patients in Car Accident Case Study

Case Study Overview: In the healthcare scenario we previously examined, nurses have the authority to access a patient's personal health data, which they share with the main parties involved in the patient's care. Nonetheless, private information, such as the precise home address of the patient, remains confidential unless an urgent situation necessitates disclosure.

- Patient Location: Somewhere, a suburb of Melbourne.

- Primary Stakeholders: Patient, GP, Paramedics, Emergency Room Nurse.
- Secondary Stakeholders: Family members, Insurance provider.
- Patient Health Situation: A man with Type 2 diabetes living at 5 Somewhere Street in a Melbourne suburb regularly sees his GP for health monitoring and treatment. His insurance provider and primary stakeholders have access to his health records, but not his specific home address. One night, he experiences a car accident and suffers from severe chest pain, difficulty breathing, and dizziness. A witness calls 000 for emergency help and the operator dispatches paramedics after learning the accident occurred in a Melbourne suburb.

Due to the urgency, the operator shares the patient's exact address with the paramedics, enabling them to reach him promptly and provide life-saving treatment. In emergencies, every second counts, and sharing a patient's information can be vital for paramedics to act quickly, administer treatment, and transport the patient to a hospital if needed. Knowledge of the patient's home address helps plan the most efficient route to the hospital, which is crucial in some cases. Therefore, the emergency operator shares the address with the responding paramedics. Upon arrival, the paramedics evaluate the patient's condition and suspect a heart attack. They provide oxygen and aspirin to stabilize him before transporting him to the closest hospital for further care. At the hospital, the emergency room nurse is briefed on the patient's condition and gains access to his personal health data to aid in his treatment. In light of the emergency, the nurse also obtains the patient's private information, including his precise home address, to facilitate any required follow-up care or communication with family members.

Throughout the entire process, the primary stakeholders, including the patient's GP, paramedics, and the nurse, stay informed about the patient's condition and treatment. They work together to ensure the patient receives the best possible care and support during this emergency. In this case, the patient's private health information is disclosed only to the necessary primary stakeholders in response to a critical emergency. This disclosure enables a fast and effective response that ultimately saves the patient's life.

A detailed health scenario (an emergency case is given in Table 3).

Table 3. A case study of a health emergency.

Personal Health Information	Privacy Policy	Emergency Situation
A primary address, such as "suburb is a suburban area in Melbourne", will only be released to all stakeholders. However, the patient's actual address, which is "5 Somewhere Street, a suburban area in Melbourne", is only released in the event of any emergency.	All primary stakeholders have access to patients' personal health information. However, they do not have access to private health information.	In an emergency, some primary stakeholders can have access to patients' private health information. For example, in cases where the patient considered to be in a very critical condition and needs to be admitted to hospital, medical practitioners have access to patients' exact home address for an emergency treatment/situation.

5. Evaluation of Privacy Ontology and Experiments

We evaluated the privacy ontology by categorizing health-related privacy policies into two types: valid and illegitimate privacy policies. To differentiate between valid and illegitimate policies, we employed NLP-based ML models, specifically Bidirectional Encoder Representations from Transformers (BERT) [72]. Based on our proposed privacy ontology, we identified the following five steps to conduct the experiments:

✓ Data Collection and Preprocessing: We collected a large dataset of health-related privacy policies from various online sources, such as hospitals, clinics, health insurance providers, and health-related mobile applications. These policies were then preprocessed to remove any irrelevant information, formatting inconsistencies, and to convert them into a machine-readable format.

- ✓ Annotation and Labeling: After preprocessing the data, we manually annotated and labeled the privacy policies based on their adherence to our privacy ontology. The annotation process involved experts in privacy and health domains, who categorized the policies into two classes: valid (policies that comply with the privacy ontology) and illegitimate (policies that do not comply with the privacy ontology).
- ✓ Feature Extraction: We extracted relevant features from the preprocessed privacy policies, such as the presence of specific keywords or phrases, using NLP techniques. These features were critical for training the BERT model, which was then used to classify the policies into valid and illegitimate categories.
- ✓ Training and Evaluation of the BERT Model: We used the annotated and labeled dataset to train the BERT model. During the training process, the model learns to identify patterns and relationships between the extracted features and the corresponding labels (valid or illegitimate). After training, we evaluated the model's performance using standard evaluation metrics such as precision, recall, and F1-score [73].
- ✓ Analyzing and Interpreting Results: We analyzed the results obtained from the BERT model to identify common patterns and trends in the classification of health-related privacy policies. This analysis provides valuable insights into the effectiveness of the privacy ontology and helps identify potential areas for improvement

Through these five steps, we aimed to demonstrate the effectiveness of our privacy ontology in distinguishing between valid and illegitimate health-related privacy policies. By leveraging state-of-the-art NLP techniques and ML models, such as BERT, we can automate the evaluation process, making it easier to ensure that privacy policies adhere to established privacy principles and guidelines.

5.1. Experiment Setup and Dataset Preparation

This section delineates the processes involved in dataset preparation and experimental setup to construct and evaluate our proposed privacy mechanism. To overcome previous limitations, our enhanced experiment design incorporated a broader range of healthcare privacy policies and adopts cross-validation techniques, ensuring a robust and comprehensive evaluation of the machine learning models.

Our dataset is an amalgamation of numerous health-related privacy policies, collected from various healthcare organizations worldwide. It includes 100 manually labeled policies categorized as either valid privacy policies (labelled '1') or illegitimate privacy policies (labelled '0'). Additionally, we incorporated 69 pre-labelled illegitimate privacy policies (labelled '0') from the "SpywareGuide" online archive. The total dataset comprises 169 labelled privacy policies, providing a diverse and substantial foundation for training our privacy model. The labeling was performed based on specific health privacy and security terms, in accordance with the APPs. A custom Python script implemented on Google Colab facilitated the process, scanning for 13 specific phrases within each policy before labeling them as '1' or '0'. This comprehensive 13-phase word analysis provided a rigorous method for categorization, ensuring accurate labeling.

Our research method approach involved using this dataset in combination with four BERT-NLP techniques: BERT, Distil BERT, Albert Tokenizer, and Roberta Tokenizer. The objective was to develop a machine learning-based privacy model specific to EHR environments, capable of accurately distinguishing between valid and illegitimate privacy policies. The model was built upon the principles derived from our updated privacy ontology, identifying a valid privacy policy as one that is either identical to or similar to a known policy previously labeled as '1'. Conversely, new policies that do not conform to these principles were considered illegitimate and labeled as '0'.

5.2. Clarifying Privacy Policy Classification

In our study, we used machine learning (ML) models to classify privacy policies as 'valid' or 'illegitimate'. This raises the question: why is there a need to classify a policy as legitimate or illegitimate, especially in a context where, as an example, healthcare providers

in the U.S. must abide by the regulations established by the Health Insurance Portability and Accountability Act (HIPAA) [74], regardless of what the ML model dictates?

1. First, it is essential to note that, while healthcare institutions in the U.S. must follow HIPAA guidelines, not all institutions worldwide must adhere to the same guidelines. Different countries have different privacy regulations, and the level of privacy protection can vary significantly across countries and even across different institutions within the same country. This variability increases the importance of having a tool that can automatically assess the validity of privacy policies across a broad range of contexts.
2. Second, even within the U.S., not all healthcare entities are required to follow HIPAA regulations. HIPAA applies primarily to healthcare providers, health plans, and healthcare clearinghouses, but it does not extend to entities such as life insurers, employers, or schools. Some of these entities might have access to sensitive health-related information and might formulate their privacy policies, necessitating a mechanism to evaluate their policies' validity.
3. Third, even when healthcare entities are required to follow HIPAA guidelines, there may still be differences in how these guidelines are interpreted and implemented. Privacy policies can be complex and nuanced, and different entities might have different interpretations of what constitutes a 'valid' policy under HIPAA guidelines. An ML model that classifies privacy policies can serve as an additional check on these entities' interpretations, highlighting potential areas of concern.
4. Finally, while we used the example of HIPAA in our study, the principles of our privacy ontology and our ML-based privacy model are not limited to HIPAA. Our approach was designed to be adaptable to various privacy regulations, not just HIPAA. This adaptability makes our approach potentially useful in a wide range of contexts, even in situations where the applicable privacy regulations differ from HIPAA.

Given these considerations, it becomes clear that there is a need to classify privacy policies as 'valid' or 'illegitimate' and that our ML-based privacy model can provide a valuable tool in this regard. It enables a more nuanced understanding of privacy policies across different contexts, aiding in the ongoing effort to ensure that sensitive health-related information is appropriately protected.

5.3. Development of ML-Based Privacy Model

In this section, we present an improved ML-based privacy model for classifying valid and illegitimate privacy policies in EHR environments. We have opted for four transformer-based text classification models, specifically BERT, DistilBERT, ALBERT, and RoBERTa, to categorize privacy policies. The rationale for choosing these models lies in their state-of-the-art performance in various natural language processing tasks, including text classification, sentiment analysis, and named entity recognition. Moreover, these models have demonstrated strong generalization capabilities in different domains and languages.

The primary goal of our privacy model was to determine whether a privacy policy is 'valid' or 'illegitimate' based on the type and extent of personal, private, sensitive, and historical information that the relevant organization usually collects from its clients (e.g., healthcare stakeholders requesting patient information). Our proposed privacy model employs an automated text classification mechanism to effectively distinguish between valid and illegitimate privacy policies.

To further enhance the model, we provide a detailed explanation of the four transformer-based text classification models:

- BERT: BERT is a powerful pre-trained language model that has achieved state-of-the-art results in various NLP tasks. Its bidirectional nature enables it to understand the context from both left and right sides of a word, resulting in improved understanding and classification accuracy.
- DistilBERT: DistilBERT is a distilled version of BERT, offering a smaller and faster alternative while maintaining most of the original model's performance. This model is particularly useful in cases where computational resources or model size is a concern.

- ALBERT (A Lite BERT): ALBERT is another variant of BERT that reduces model size and improves training efficiency by sharing parameters across layers. Despite its reduced size, ALBERT maintains competitive performance, making it a suitable choice for our privacy model.
- RoBERTa (Robustly Optimized BERT Pretraining Approach): RoBERTa is a modified version of BERT that has been optimized for increased training efficiency and performance. It uses dynamic masking, larger batch sizes, and other training optimizations to achieve superior results in text classification tasks.

By leveraging these advanced transformer-based models, we have developed a machine learning-based privacy model derived from the concepts of our proposed privacy ontology. The objective of the model is to accurately and efficiently classify privacy policies in EHR environments as either valid or illegitimate, ensuring the protection of sensitive patient information. We first manually annotated a training dataset of 169 policies, labeling each policy as either valid (1) or illegitimate (0). A privacy policy was considered valid if it is either identical or closely comparable to an already labeled "1" policy within the training dataset. Policies that did not meet these criteria were labeled as "0" and deemed illegitimate. An example of an illegitimate privacy policy sourced from the "SpywareGuide.com" online dataset is shown in Table 4.

To enhance the model's performance, we explored various BERT-based techniques for processing the training data. These methods included transforming the data into tensors, creating data batches, fine-tuning the model, and finally validating its accuracy. We trained multiple BERT variants and systematically compared their results to determine the most accurate model for our privacy policy classification task.

Table 4. An example privacy policy.

Policy Type	Policy Statement
0]	. . . Alexa Internet Privacy Policy, based on the last updated version on 7 April 2011. What Personal Information About Users Does Alexa Gather? Information You Give Us—We receive and store any information you enter on our Web site or give us in any other way. Automatic Information We Collect From the Toolbar Service—When you use the Toolbar Service, we collect information about the websites you visit and the advertisements that you see on those websites, the searches you perform using search engines. . .

5.4. Application of Privacy Ontology in the Machine Learning Experiment

In this section, we elaborate on how the developed privacy ontology model is integrated into the machine learning experiment and how it enhances the privacy of patient data. Our privacy ontology model is crucial in setting up the ground rules for classifying privacy policies. It defines the valid and illegitimate privacy policy categories, shaping the direction of our machine learning model training. The privacy ontology model, constructed based on in-depth analysis and understanding of privacy laws and principles, ensures a high standard in distinguishing valid and illegitimate privacy policies. The application of privacy ontology in our machine learning experiment can be elaborated in three distinct steps:

1. Annotation and Labeling: The privacy ontology serves as a guidance for annotating and labeling the collected privacy policies. The experts involved in the annotation process utilize the ontology's principles to identify the class of each policy. As such, the ontology enables a more reliable and consistent labeling process.
2. Feature Extraction: Privacy ontology plays a crucial role in determining the relevant features from the privacy policies. It aids in identifying the specific keywords or phrases that signify valid or illegitimate policies, thereby helping in effective feature extraction.
3. Model Training: During the model training phase, the privacy ontology plays a pivotal role in guiding the learning process of the BERT model. The model learns to identify

patterns that align with the principles of our privacy ontology, which helps to classify new privacy policies accurately.

The integration of our privacy ontology model into the machine learning experiment significantly enhances the privacy of patient data. It does so by ensuring that any privacy policy, whether from healthcare providers, health insurance providers, or health-related mobile applications, complies with established privacy principles and guidelines before being classified as valid. This procedure guarantees that only policies that uphold high privacy standards will be deemed valid, thereby providing a robust safeguard for patient data privacy. It is worth noting that our ML model's potential to accurately classify privacy policies does not negate the human role in this process. Our model acts as an assisting tool that automates and speeds up the classification process, but the initial rules and principles (as defined by the privacy ontology) are still set by human experts. Thus, while our model provides an additional layer of protection for patient data privacy, it does not eliminate the need for human oversight, particularly in complex cases where a human might be more adept at identifying the appropriate policy.

5.5. Evaluation Results

The main objective of this section is to identify the legitimate privacy policies that are associated with the health and personal information in different EHR environments. The primary purpose was to ensure the EHR remains secure from any unauthorized attempts to access it. Towards this goal, we conducted a set of experiments to evaluate the efficiency of the proposed ML based privacy model. The technical details of the experiments and the relevant results have been presented as follows.

5.5.1. Technical Details

A valid privacy policy is represented as 2-tuple relation, including privacy policies from our proposed privacy ontology and legitimate actions (e.g., collecting, storing and/or disseminating health information where personal, private or sensitive information is involved with authorized parties). On the contrary, an illegitimate privacy policy is one where illegitimate actions (e.g., using health information for marketing purposes) are involved with unauthorized parties (e.g., who are not primary or secondary health related stakeholders).

We calculated accuracy, precision, recall, and f1 score using different BERT techniques to classify the privacy policies (the dataset of 169 policies). These BERT models were used to model our ML-based privacy approach. We use automated text categorisation mechanism to classify a privacy policy is valid or illegitimate.

5.5.2. Dataset and Results

The proposed model was evaluated on a dataset containing valid and illegitimate privacy policies generated by health related organisations. We collected those privacy policies from multiple health related organisations. We also used some illegitimate privacy policies from the online SpywareGuide archive.

Table 5 shows the experiment's results using the different NLP based BERT techniques, such as BERT, Distil BERT, Albert Tokenizer, and Roberta Tokenizer. In these experiments, we used the concept of automated text categorisation mechanism in our ML-based privacy approach through our introduced privacy classification technique—illegitimate versus valid. We achieved 94% accuracy using the Distil BERT technique, which is better than the other BERT techniques which achieved 76%, 90%, and 92% accuracy using Albert Tokenizer, Roberta Tokenizer, and BERT techniques, respectively.

In the aforementioned experiments conducted on Google Colab, a cloud-based machine learning platform, we utilized a custom dataset comprising privacy policies obtained from a variety of health-related organizations, both legitimate and illegitimate. The illegitimate policies were primarily sourced from the SpywareGuide archive, an online resource providing information about privacy risks. Each privacy policy in our dataset was transformed into a vector representation using the BERT tokenizer before being processed by

our machine learning model. For the experimental setup, we used the BERT, DistilBERT, Albert, and RoBERTa models as implemented in the Google Colab Transformers library. The models were fine-tuned on our dataset using a learning rate of 2×10^{-5}, batch size of 16, and for a total of four epochs, leveraging the high-performance computing power of Google Colab. The choice of these parameter settings was informed by preliminary experiments and the recommended settings from the original papers introducing these models. To validate the practicality of the proposed model, we performed an additional set of experiments using privacy policies from a diverse range of healthcare sectors, including hospitals, insurance providers, and digital health applications. We found that our model, trained and fine-tuned on Google Colab, consistently achieved high accuracy in identifying legitimate and illegitimate privacy policies across these sectors, reinforcing its practical applicability in a real-world context.

Table 5. Experiment results using NLP-based BERT techniques.

BERT Techniques	Accuracy	Precision	F1 Score
BERT	0.92	0.86	0.90
Distil BERT	0.94	0.94	0.94
Albert Tokenizer	0.76	0.87	0.76
Roberta Tokenizer	0.90	0.82	0.92

5.6. Summary of the Findings

In our experimental setup, we employed various BERT techniques to distinguish between legitimate and illegitimate privacy policies, with the primary aim of establishing privacy measures that effectively protect EHR from unauthorized access. To conduct this experiment, we utilized a dataset consisting of 169 illegitimate and valid privacy policies collected from the online SpywareGuide archive and relevant health-related organizations. Our experiments demonstrated that the proposed ML-based privacy approach can reliably recognize illegitimate policies, achieving an F1 score exceeding 0.94 when using a testing dataset comprising 20% of the data and a training dataset of 80%, indicating that our approach has been effective in identifying privacy policies that may not provide adequate protection for sensitive health information. The results of our analysis showed that Distil BERT outperformed the other techniques in terms of accuracy, precision, and F1 score, achieving a balanced performance in identifying both legitimate and illegitimate privacy policies. This finding suggested that Distil BERT can be an effective tool for analyzing privacy policies in e-Healthcare systems, providing valuable insights for enhancing privacy protection measures.

We anticipate that the accuracy and precision of our approach could be further improved by using a larger dataset containing a more diverse range of health-related privacy policies. In future experiments, we plan to expand our dataset and refine our proposed ML-based approach to better determine the validity of these privacy policies. By doing so, we hope to develop a more robust and accurate system for identifying and protecting against potential risks associated with inadequate privacy policies in e-Healthcare systems.

6. Discussion

This section will focus on discussing the key insights gathered during the research.

6.1. The Relationship between the Ontology and the ML Model

Integrating an ontology-driven approach with ML could enhance the confidentiality of MHRs and refine data categorization processes [75–80]. Primarily, an ontology-driven methodology offers a structured classification of medical terms and concepts. This structure eases the accurate identification and categorization of sensitive health data [79]. An organized and standardized approach to health information ensures interoperability and streamlines data management. Additionally, the ontology can link various data compo-

nents, delivering contextual information that could refine ML algorithm performance when detecting sensitive information [77].

Training ML algorithms with a dataset containing pre-identified sensitive information can further enhance data categorization accuracy [76]. Here, the algorithm learns to recognize patterns and characteristics associated with sensitive data, ensuring the effective detection and protection of such data. In conclusion, combining an ontology-driven approach with ML offers a robust platform for protecting medical records. This combination substantially improves health data confidentiality and security, contributing to superior patient care and privacy.

6.2. The Complementarity of Ontology and ML Model

Our research illustrates that the interplay between ontology and ML can significantly enhance healthcare data management. The ontology's role is multifold: it serves as a semantic framework that provides context and meaning to raw data, it defines the scope of data to be processed, and it structures the data in a manner that can be effectively utilized by ML models [81,82]. The ontology model plays an instrumental role in the preprocessing stage of ML by identifying relevant data sources and features [83]. This identification ensures that ML models are trained on pertinent and meaningful data, thus enhancing the models' capability to accurately identify sensitive information while maintaining patient confidentiality. Additionally, ML models assist in the continuous refinement of the ontology [83,84]. The models identify patterns and relationships within data, which provide insights into potential improvements to the ontology. These insights help refine the ontology's structure and contribute to a more accurate representation of the health information domain.

In essence, the interplay between ontology and ML in our work exhibits a synergistic relationship, where the strengths of one approach are leveraged to complement the other [82]. The ontology model provides a meaningful and context-rich foundation for ML models, whereas the ML models contribute to the iterative refinement and validation of the ontology. This complementary relationship culminates in a system that is not only secure and privacy-preserving, but also efficient in managing sensitive health information [81].

6.3. Adoption of Privacy-Preserving Technologies for Health Information Security

Although our focus has largely been on ontology-driven methodologies and ML techniques, it is also essential to acknowledge the role of privacy-preserving technologies in ensuring the security of health information. The privacy ontology model we propose in this paper provides a structured framework for understanding and managing health-related information but it needs to be complemented with various privacy-preserving technologies to fully realize its potential. These technologies include, but are not limited to, data encryption, differential privacy, and secure multi-party computation, which provide the technical means to protect sensitive data while still enabling valuable insights to be gleaned [85,86]. The integration of such technologies with our proposed privacy ontology model can ensure that the privacy rules and regulations, as well as the rights and privileges of stakeholders, are effectively enforced in real-world applications. This integrated approach can also address potential vulnerabilities, such as data breaches and unauthorized access, thus further enhancing the confidentiality and security of health data. Consequently, while the privacy ontology model contributes significantly to conceptualizing and organizing privacy in healthcare, the adoption of privacy-preserving technologies is integral to operationalizing these concepts and effectively safeguarding health information.

6.4. Section Summary

The relationship between privacy ontology and ML significantly enhances e-Healthcare systems' security, privacy, and interoperability. By offering structured and standardized frameworks, these techniques improve data management, access control, and overall sys-

tem efficiency. As a result, they support the secure and confidential exchange of health information in an increasingly digital landscape.

7. Conclusions and Future Research

This study has explored the potential of a universal privacy model in the realm of EHR systems and context-sensitive privacy policies. Challenges such as the trade-off between privacy and accessibility, user-friendliness, and legal compliance persist, and our work aimed to contribute to these ongoing discussions. We proposed a conceptual privacy model, employing a novel privacy ontology and an ML-based mechanism, which sought to discern between legitimate and illegitimate privacy policies while factoring in patients' PII.

We leveraged various BERT techniques in our endeavor to pinpoint illegitimate privacy policies, indicating that our proposed ML-based approach has the potential to effectively discern such policies. Distil BERT was particularly adept at identifying both legitimate and illegitimate policies. The research suggests that refining the ML-based approach and expanding the dataset could result in a more resilient system to combat potential risks linked to inadequate privacy policies in e-Healthcare systems.

7.1. Limitation

Our study, while pioneering, is not without limitations. Indeed, these limitations underscore the need for further research and validation of the proposed privacy model in the context of evolving technology and privacy regulations. Our study lacks empirical evidence to fully support the effectiveness and reliability of our approach, suggesting the necessity of thorough evaluations and real-world testing. Moreover, our solution's scalability and interoperability with existing healthcare IT systems and EHR platforms remain largely untested. As such, our model may require continuous updates and adjustments to align with technological advancements and emerging privacy-enhancing techniques. Ethical considerations, such as potential bias in ML algorithms, data ownership, and consent management, have yet to be explored within the scope of this study. Additionally, the adaptability of our solution in the face of changing legal landscapes and ongoing compliance with shifting privacy regulations warrants further scrutiny. Lastly, the practicalities of implementing our proposed solution in real-world healthcare settings, including overcoming resource constraints and resistance to change, as well as addressing the need for user training and support, are areas that require future exploration and validation.

7.2. Future Research Directions

To address the limitations elucidated above, future research should focus on comprehensive evaluations of the proposed ontology and ML-based approach in terms of performance, scalability, and interoperability. Future studies should also investigate strategies for integrating novel technological advancements and changes in privacy regulations to ensure the maintenance of a relevant and effective solution. Key areas of future research to advance this field include thorough evaluations through real-world case studies and pilot implementations, exploring frameworks to adapt to advancements and changes, examining ethical implications and fostering collaboration among stakeholders. In addition, future research should investigate user-centered design principles to create a solution that is both user-friendly and accessible, alongside the development of practical implementation strategies to seamlessly integrate the proposed solution into existing healthcare settings. The proposed solution's resilience against various security threats and attack scenarios, along with strategies to mitigate potential vulnerabilities, remains an essential focus area for future research. By following these research directions, we hope to contribute to the ongoing evolution of secure healthcare information systems, aiming to enhance both privacy and accessibility in the world of e-Healthcare.

Author Contributions: R.N.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing of Original Draft, Writing Revisions Review & Editing, Visualization, Project administration. K.A.: Formal analysis, Investigation, Resources, Review & Editing, Visualization, Supervision. H.W.: Supervision, Formal analysis, Investigation, Review & Editing, Visualization. T.M.: Supervision, Revision, Validation, Formal analysis, Investigation, Resources, Data Curation, Review & Editing, Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: The corresponding author graciously confirmed that data pertinent to this study will be made accessible, should a formal request be submitted.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, H.; Song, Y. Secure cloud-based ehr system using attribute- based cryptosystem and blockchain. *J. Med. Syst.* **2018**, *42*, 152. [CrossRef]
2. Dagher, G.G.; Mohler, J.; Milojkovic, M.; Marella, P.B. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustain. Cities Soc.* **2018**, *39*, 283–297. [CrossRef]
3. Bani Issa, W.; Al Akour, I.; Ibrahim, A.; Almarzouqi, A.; Abbas, S.; Hisham, F.; Griffiths, J. Privacy, confidentiality, security and patient safety concerns about electronic health records. *Int. Nurs. Rev.* **2020**, *67*, 218–230. [CrossRef]
4. Keshta, I.; Odeh, A. Security and privacy of electronic health records: Concerns and challenges. *Egypt. Inform. J.* **2021**, *22*, 177–183. [CrossRef]
5. Ozair, F.F.; Jamshed, N.; Sharma, A.; Aggarwal, P. Ethical issues in electronic health records: A general overview. *Perspect. Clin. Res.* **2015**, *6*, 73. [PubMed]
6. Zaghloul, E.; Li, T.; Ren, J. Security and privacy of electronic health records: Decentralized and hierarchical data sharing using smart contracts. In Proceedings of the 2019 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 18–21 February 2019; pp. 375–379. [CrossRef]
7. Akarca, D.; Xiu, P.; Ebbitt, D.; Mustafa, B.; Al-Ramadhani, H.; Albey-Atti, A. Blockchain secured electronic health records: Patient rights, privacy and cybersecurity. In Proceedings of the 2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT), Leeds, UK, 5–7 June 2019; pp. 108–111. [CrossRef]
8. Omar, A.H.A. *The Effect of Electronic Health Records on Undergraduate and Postgraduate Medical Education: A Scoping Review*; University of Toronto: Toronto, ON, Canada, 2019.
9. Rezaeibagha, F.; Mu, Y. Distributed clinical data sharing via dynamic access-control policy transformation. *Int. J. Med. Inform.* **2016**, *89*, 25–31. [CrossRef]
10. Farhadi, M.; Haddad, H.; Shahriar, H. Static analysis of hippa security requirements in electronic health record applications. In Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 June 2018; Volume 2, pp. 474–479. [CrossRef]
11. Vimalachandran, P.; Zhang, Y.; Cao, J.; Sun, L.; Yong, J. Preserving data privacy and security in australian my health record system: A quality health care implication. In Proceedings of the International Conference on Web Information Systems Engineering, Dubai, United Arab Emirates, 12–15 November 2018; pp. 111–120. [CrossRef]
12. Budd, J.; Miller, B.S.; Manning, E.M.; Lampos, V.; Zhuang, M.; Edelstein, M.; Rees, G.; Emery, V.C.; Stevens, M.M.; Keegan, N.; et al. Digital technologies in the public-health response to COVID-19. *Nat. Med.* **2020**, *26*, 1183–1192. [CrossRef] [PubMed]
13. Mooney, S.J.; Pejaver, V. Big data in public health: Terminology, machine learning, and privacy. *Annu. Rev. Public Health* **2018**, *39*, 95–112. [CrossRef] [PubMed]
14. Ahmed, Z. Practicing precision medicine with intelligently integrative clinical and multi-omics Data Analysis. *Hum. Genom.* **2020**, *14*, 35. [CrossRef]
15. Kumaar, M.; Samiayya, D.; Vincent, P.M.; Srinivasan, K.; Chang, C.Y.; Ganesh, H. A hybrid framework for intrusion detection in healthcare systems using Deep Learning. *Front. Public Health* **2021**, *9*, 824898. [CrossRef]
16. Alagar, V.; Alsaig, A.; Ormandjiva, O.; Wan, K. Context-based security and privacy for healthcare IoT. In Proceedings of the 2018 IEEE International Conference on Smart Internet of Things (SmartIoT), Xi'an, China, 17–19 August 2018; pp. 122–128. [CrossRef]
17. Demuro, P.R.; Petersen, C. Managing privacy and data sharing through the use of health care information fiduciaries. In *Context Sensitive Health Informatics: Sustainability in Dynamic Ecosystems*; IOS Press: Amsterdam, The Netherlands, 2019; pp. 157–162. [CrossRef]
18. Kisekka, V.; Giboney, J.S. The effectiveness of health care information technologies: Evaluation of trust, security beliefs, and privacy as determinants of health care outcomes. *J. Med. Internet Res.* **2018**, *20*, e9014. [CrossRef]

19. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The future of digital health with federated learning. *NPJ Digit. Med.* **2020**, *3*, 119. [CrossRef] [PubMed]
20. Kruse, C.S.; Smith, B.; Vanderlinden, H.; Nealand, A. Security techniques for the electronic health records. *J. Med. Syst.* **2017**, *41*, 127. [CrossRef]
21. Otlowski, M.F.A. Disclosing genetic information to at-risk relatives: New Australian privacy principles, but uniformity still elusive. *Med. J. Aust.* **2015**, *202*, 335–337. [CrossRef] [PubMed]
22. Ahmed, Z.; Mohamed, K.; Zeeshan, S.; Dong, X.Q. Artificial Intelligence with multi-functional machine learning platform development for better healthcare and Precision Medicine. *Database* **2020**, *2020*, baaa010. [CrossRef]
23. Chen, M.; Decary, M. Artificial Intelligence in healthcare: An essential guide for health leaders. *Healthc. Manag. Forum* **2019**, *33*, 10–18. [CrossRef] [PubMed]
24. Koczkodaj, W.W.; Mazurek, M.; Strzałka, D.; Wolny-Dominiak, A.; Woodbury-Smith, M. Electronic health record breaches as social indicators. *Soc. Indic. Res.* **2019**, *141*, 861–871. [CrossRef]
25. Abramson, W.; Hall, A.J.; Papadopoulos, P.; Pitropakis, N.; Buchanan, W.J. A distributed trust framework for privacy-preserving machine learning. In *Trust, Privacy and Security in Digital Business*; Springer: Cham, Switzerland, 2020; pp. 205–220. [CrossRef]
26. Islam, T.U.; Ghasemi, R.; Mohammed, N. Privacy-preserving federated learning model for healthcare data. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 26–29 January 2022. [CrossRef]
27. Baker, A. Crossing the quality chasm: A new health system for the 21st century. *BMJ Clin. Res.* **2001**, *323*, 1192. [CrossRef]
28. Olive, M.; Rahmouni, H.B.; Solomonides, T.; Breton, V.; Legré, Y.; Blanquer, I.; Hernández, V.; Andoulsi, I.; Herveg, J.A.M.; Wilson, P. Share roadmap 1: Towards a debate. *Stud. Health Technol. Inform.* **2007**, *126*, 164–173.
29. Li, M.; Yu, S.; Ren, K.; Lou, W. Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings. In Proceedings of the International Conference on Security and Privacy in Communication Systems, Singapore, 7–9 September 2010; pp. 89–106. [CrossRef]
30. Caine, K.; Hanania, R. Patients want granular privacy control over health information in electronic medical records. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 7–15. [CrossRef]
31. Abouelmehdi, K.; Beni-Hessane, A.; Khaloufi, H. Big healthcare data: Preserving security and privacy. *J. Big Data* **2018**, *5*, 1. [CrossRef]
32. Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S.; Wang, G. Security and privacy in the medical internet of things: A review. *Secur. Commun. Netw.* **2018**, *2018*, 5978636. [CrossRef]
33. Johnson, C.M.; Grandison, T. Compliance with data protection laws using hippocratic database active enforcement and auditing. *IBM Syst. J.* **2007**, *46*, 255–264. [CrossRef]
34. Fernández-Alemán, J.L.; Señor, I.C.; Lozoya, P.Á.O.; Toval, A. Security and privacy in electronic health records: A systematic literature review. *J. Biomed. Inform.* **2013**, *46*, 541–562. [CrossRef]
35. Eze, E.; Gleasure, R.; Heavin, C. Mobile health solutions in developing countries: A stakeholder perspective. *Health Syst.* **2020**, *9*, 179–201. [CrossRef]
36. Peute, L.W.; Wildenbos, G.A.; Engelsma, T.; Lesselroth, B.J.; Lichtner, V.; Monkman, H.; Neal, D.; Van Velsen, L.; Jaspers, M.W.; Marcilly, R. Overcoming Challenges to Inclusive User-based Testing of Health Information Technology with Vulnerable Older Adults: Recommendations from a Human Factors Engineering Expert Inquiry. *Yearb. Med. Inform.* **2022**, *31*, 74–81. [CrossRef]
37. Mamun, Q. A conceptual framework of personally controlled electronic health record (pcehr) system to enhance security and privacy. In Proceedings of the International Conference on Applications and Techniques in Cyber Security and Intelligence, Barcelona, Spain, 12–16 November 2017; pp. 304–314. [CrossRef]
38. Samet, S.; Ishraque, M.T.; Sharma, A. Privacy-preserving personal health record (p3hr) a secure android application. In Proceedings of the 7th International Conference on Software and Information Engineering, Cairo, Egypt, 2–4 May 2018. [CrossRef]
39. Wachter, S. The GDPR and the Internet of Things: A three-step transparency model. *Law, Innov. Technol.* **2018**, *10*, 266–294. [CrossRef]
40. Cavoukian, A. Privacy by design: The definitive workshop. A foreword by Ann Cavoukian, Ph. D. *Identity Inf. Soc.* **2010**, *3*, 247–251. [CrossRef]
41. Zhang, A.; Lin, X. Towards secure and privacy-preserving data shar- ing in e-health systems via consortium blockchain. *J. Med. Syst.* **2018**, *42*, 140. [CrossRef]
42. Kuo, T.T.; Kim, H.E.; Ohno-Machado, L. Blockchain distributed ledger technologies for biomedical and health care applications. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 1211–1220. [CrossRef]
43. Esposito, C.; Santis, A.; Tortora, G.; Chang, H.; Choo, K.K.R. Blockchain: A panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Comput.* **2018**, *5*, 31–37. [CrossRef]
44. Wang, J.; Li, M.; He, Y.; Li, H.; Xiao, K.; Wang, C. A blockchain based privacy-preserving incentive mechanism in crowdsensing applications. *IEEE Access* **2018**, *6*, 17545–17556. [CrossRef]
45. Fan, K.; Wang, S.; Ren, Y.; Li, H.; Yang, Y. Medblock: Efficient and secure medical data sharing via blockchain. *J. Med. Syst.* **2018**, *42*, 136. [CrossRef] [PubMed]
46. Griggs, K.N.; Ossipova, O.; Kohlios, C.P.; Baccarini, A.N.; Howson, E.A.; Hayajneh, T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *J. Med. Syst.* **2018**, *42*, 130. [CrossRef] [PubMed]

47. Kayes, A.; Rahayu, W.; Dillon, T.; Chang, E.; Han, J. Context-aware access control with imprecise context characterization for cloud-based data resources. *Future Gener. Comput. Syst.* **2019**, *93*, 237–255. [CrossRef]
48. Kayes, A.; Han, J.; Rahayu, W.; Dillon, T.; Islam, M.S.; Colman, A. A policy model and framework for context-aware access control to information resources. *Comput. J.* **2019**, *62*, 670–705. [CrossRef]
49. Kayes, A.; Rahayu, W.; Watters, P.; Alazab, M.; Dillon, T.; Chang, E. Achieving security scalability and flexibility using fog-based context-aware access control. *Future Gener. Comput. Syst.* **2020**, *107*, 307–323. [CrossRef]
50. Kayes, A.; Rahayu, W.; Dillon, T. An ontology-based approach to dynamic contextual role for pervasive access control. In Proceedings of the 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Krakow, Poland, 16–18 May 2018; pp. 601–608. [CrossRef]
51. Paul, J.; Annamalai, M.S.M.S.; Ming, W.; Al Badawi, A.; Veeravalli, B.; Aung, K.M.M. Privacy-preserving collective learning with homomorphic encryption. *IEEE Access* **2021**, *9*, 132084–132096. [CrossRef]
52. Ikuomola, A.J.; Arowolo, O.O. Securing patient privacy in e-health cloud using homomorphic encryption and access control. *Int. J. Comput. Netw. Commun. Secur.* **2014**, *2*, 15–21.
53. Vengadapurvaja, A.; Nisha, G.; Aarthy, R.; Sasikaladevi, N. An efficient homomorphic medical image encryption algorithm for cloud storage security. *Procedia Comput. Sci.* **2017**, *115*, 643–650. [CrossRef]
54. Alzubi, J.A.; Alzubi, O.A.; Beseiso, M.; Budati, A.K.; Shankar, K. Optimal multiple key-based homomorphic encryption with deep neural networks to secure medical data transmission and diagnosis. *Expert Syst.* **2022**, *39*, e12879. [CrossRef]
55. Subramaniyaswamy, V.; Jagadeeswari, V.; Indragandhi, V.; Jhaveri, R.H.; Vijayakumar, V.; Kotecha, K.; Ravi, L. Somewhat homomorphic encryption: Ring learning with error algorithm for faster encryption of iot sensor signal-based edge devices. *Secur. Commun. Netw.* **2022**, *2022*, 2793998. [CrossRef]
56. Vamsi, D.; Reddy, P. Electronic health record security in cloud: Medical data protection using homomorphic encryption schemes. In *Research Anthology on Securing Medical Systems and Records*; IGI Global: Hershey, PA, USA, 2022; pp. 853–877.
57. Spencer, A.; Patel, S. Applying the data protection act 2018 and general data protection regulation principles in healthcare settings. *Nurs. Manag.* **2019**, *26*, 34–40. [CrossRef] [PubMed]
58. Mills, J.L.; Harclerode, K. Privacy, mass intrusion, and the modern data breach. *Fla. Law Rev.* **2017**, *69*, 771.
59. Alghrani, A.; Brazier, M.; Farrell, A.M.; Griffiths, D.; Allen, N. Healthcare scandals in the NHS: Crime and punishment. *J. Med. Ethics* **2011**, *37*, 230–232. [CrossRef]
60. Schaar, P. Privacy by design. *Identity Inf. Soc.* **2010**, *3*, 267–274. [CrossRef]
61. Malamed, S.F. *Sedation-e-Book: A Guide to Patient Management*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2017.
62. Sreejith, R.; Senthil, S. Dynamic Data Infrastructure Security for interoperable e-healthcare systems: A semantic feature-driven NoSQL intrusion attack detection model. *BioMed Res. Int.* **2022**, *2022*, 4080199. [CrossRef]
63. Deebak, B.D.; Memon, F.H.; Cheng, X.; Dev, K.; Hu, J.; Khowaja, S.A.; Qureshi, N.M.; Choi, K.H. Seamless privacy-preservation and Authentication Framework for IOT-enabled Smart eHealth Systems. *Sustain. Cities Soc.* **2022**, *80*, 103661. [CrossRef]
64. Sharma, A.; Kumar, S. Machine learning and ontology-based novel semantic document indexing for information retrieval. *Comput. Ind. Eng.* **2023**, *176*, 108940. [CrossRef]
65. Fries, J.A.; Steinberg, E.; Khattar, S.; Fleming, S.L.; Posada, J.; Callahan, A.; Shah, N.H. Ontology-driven weak supervision for clinical entity classification in Electronic Health Records. *Nat. Commun.* **2021**, *12*, 2017. [CrossRef]
66. Sahoo, S.S.; Kobow, K.; Zhang, J.; Buchhalter, J.; Dayyani, M.; Upadhyaya, D.P.; Prantzalos, K.; Bhattacharjee, M.; Blumcke, I.; Wiebe, S.; et al. Ontology-based feature engineering in machine learning workflows for Heterogeneous Epilepsy Patient Records. *Sci. Rep.* **2022**, *12*, 19430. [CrossRef] [PubMed]
67. Zhu, N.; Chen, B.; Wang, S.; Teng, D.; He, J. Ontology-based approach for the measurement of privacy disclosure. *Inf. Syst. Front.* **2021**, *24*, 1689–1707. [CrossRef]
68. Yehia, E.; Boshnak, H.; Abdelgaber, S.; Abdo, A.; Elzanfaly, D.S. Ontology-based clinical information extraction from physician's free-text notes. *J. Biomed. Inform.* **2019**, *98*, 103276. [CrossRef] [PubMed]
69. Bosco, A.D.; Vieira, R.; Zanotto, B.; Etges, A.P.D.S. Ontology based classification of electronic health records to support value-based health care. In Proceedings of the Brazilian Conference on Intelligent Systems, Virtual, 29 November–3 December 2021; pp. 359–371. [CrossRef]
70. Kayes, A.S.M.; Han, J.; Colman, A. An ontology-based approach to context-aware access control for software services. In Proceedings of the International Conference on Web Information Systems Engineering, Nanjing, China, 13–15 October 2013; pp. 410–420. [CrossRef]
71. McGuinness, D.L.; Van Harmelen, F. OWL web ontology language overview. *W3C Recomm.* **2004**, *10*, 2004.
72. Wang, A.; Cho, K. BERT has a mouth, and it must speak: BERT as a Markov random field language model. *arXiv* **2019**, arXiv:1902.04094. https://doi.org/10.48550/arXiv.1902.04094.
73. Bisong, E. Google colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*; Springer: Berlin, Germany, 2019; pp. 59–64. [CrossRef]
74. Schweitzer, E.J. Reconciliation of the cloud computing model with US federal electronic health record regulations. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 161–165. [CrossRef]
75. Adel, E.; El-Sappagh, S.; Barakat, S.; Hu, J.W.; Elmogy, M. An extended semantic interoperability model for distributed electronic health record based on fuzzy ontology semantics. *Electronics* **2021**, *10*, 1733. [CrossRef]

76. Afzal, Z.; Schuemie, M.J.; van Blijderveen, J.C.; Sen, E.F.; Sturkenboom, M.C.; Kors, J.A. Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. *BMC Med. Inform. Decis. Mak.* **2013**, *13*, 30. [CrossRef]
77. Gu, T.; Wang, X.H.; Pung, H.K.; Zhang, D.Q. An ontology-based context model in intelligent environments. *arXiv* **2020**, arXiv:2003.05055. https://doi.org/10.48550/arXiv.2003.05055
78. Puri, C.A.; Gomadam, K.; Jain, P.; Yeh, P.Z.; Verma, K. Multiple Ontologies in Healthcare Information Technology: Motivations and Recommendation for Ontology Mapping and Alignment. In Proceedings of the ICBO, Buffalo, NY, USA, 26–30 July 2011.
79. Quamar, A.; Özcan, F.; Miller, D.; Moore, R.J.; Niehus, R.; Kreulen, J. Conversational BI: An ontology-driven conversation system for business intelligence applications. *Proc. VLDB Endowment* **2020**, *13*, 3369–3381. [CrossRef]
80. Tsymbal, A.; Zillner, S.; Huber, M. Ontology–supported machine learning and decision support in biomedicine. In Proceedings of the Data Integration in the Life Sciences: 4th International Workshop, DILS 2007, Philadelphia, PA, USA, 27–29 June 2007; pp. 156–171. [CrossRef]
81. Zhang, L.; Qi, F.; Wang, Z.; Wang, E.; Liu, Z. Integrating Semantic Knowledge to Tackle Zero-shot Text Classification. *arXiv* **2019**, arXiv:1911.04841. https://doi.org/10.48550/arXiv.1903.12626.
82. Hitzler, P.; Krisnadhi, A.A.; Janowicz, K. *Ontology Engineering with Ontology Design Patterns: Foundations and Applications*; IOS Press: Amsterdam, The Netherlands, 2016.
83. Sharma, S.; Alebouyeh, A.R.; Perera, S.; Barreto, M.D.; Udgata, S.K. The role of ontologies for sustainable, semantically interoperable and trustworthy EHR solutions. *Healthc. Technol. Lett.* **2020**, *7*, 14–22.
84. Stevens, A.; Dehghan, A. Using machine learning to identify disease-relevant genes. *Curr. Opin. Genet. Dev.* **2018**, *50*, 48–53.
85. Sánchez, D.; Batet, M. C-sanitized: A privacy model for document redaction and sanitization. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 148–163. [CrossRef]
86. Kanaan, H.; Mahmood, K.; Sathyan, V. An ontological model for privacy in emerging decentralized healthcare systems. In Proceedings of the 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS), Bangkok, Thailand, 22–24 March 2017; IEEE: Piscataway, NJ, USA; pp. 107–113. [CrossRef]