CONTEXT-AWARE PRIVACY

A concept of contextual Integrity application and approach

by

Manogna Sunkara

Student ID: 100788950

manogna.sunkara@ontariotechu.net

A capstone project submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Masters in IT Security – Artificial Intelligence Specialization

Faculty of Business and Information Technology

University of Ontario Institute of Technology (Ontario Tech University)

Oshawa, Ontario, Canada

FALL 2022

CAPSTONE PROJECT REVIEW INFORMATION

Submitted by: **Manogna Sunkara (100788950)**

Degree Name in **Program Name**

Project/Major Paper title: CONTEXT-AWARE PRIVACY - A concept of contextual Integrity application and approach

The capstone project was approved in April 2022 by the following review committee:

Review Committee:

Research Supervisor	Stephen Marsh

The above review committee determined that the capstone project is acceptable in form and content and that a satisfactory knowledge of the field was covered by the work submitted. A copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

CONTEXT-AWARE PRIVACY

A concept of contextual Integrity application and approach

Manogna Sunkara
Department of Business and Information Technology
Ontario Tech University
Oshawa, ON, Canada
manogna.sunkara@ontariotechu.net

ABSTRACT

The epoch of data exchange has arrived and led to big data evolution. Petabytes of transaction information, weblogs, cookie logs, file types, online databases, and records from social networks, mobile devices, and wearable technology are all now accessible. Both these data and the technology used to glean insights, discoveries, and meaning from them are expanding at an exponential rate. This new data science has been predominantly controlled by professionals working in information-rich businesses. A significantly wider range of industries, including health care, business, access to education, and government, are now pursuing the value of data-driven decision-making that big data provides as the tools for working with big data are improving quickly. They aimed to improve our comprehension of a phenomenon by analyzing—and frequently sharing publicly—large data sets of user information they regarded as freely available for research, training, and improving the business models. However, in every instance, concerns about the ethicality of these big data efforts surfaced right away. Several of the core rules of research ethics, such as protecting user privacy, obtaining valid consent, preserving the confidentiality of any data collected, and minimizing harm, appeared to be broken by the professionals' methodological practices. Additionally, knowledge of the specific ethical issues of Internet-based research has expanded even during the previous 20 years. In the age of information technologies, the practice of public surveillance and sensitive data exchange is one of the least understood and most divisive dangers to privacy. The fragmented nature of privacy regulations also reflects the ambiguity of unresolved intuitions regarding commonplace occurrences like using data outside of the context for which it was gathered, in addition to the competing interests of various parties involved. We outline a technique for evaluating privacy regulations using the contextual integrity framework. This technique enables the systematic identification of privacy policy statement flaws that prevent users from comprehending and assessing business data-gathering activities. These problems include the absence of context information, ambiguous wording, and an abundance of potential interpretations of the specified information transfers from the legal-ethical dimension as well as from real-world application. Finally, we will conclude with a model to protect personal information using cryptographic encryption and decryption techniques.

Keywords: Contextual Integrity (CI), privacy in Network contexts, Hash function, Symmetric key Encryption, Access controls and regulations

1. INTRODUCTION

A rising trend is disclosing information in an easily accessible format, particularly for public organizations that are required by law to do so. We examine the privacy effects of publishing open data, focusing on how organizations might decide on privacy concerns associated with open data publishing before dissemination. Given the world of information technologies, the practice of public surveillance and sensitive data exchange is one of the least understood and most divisive dangers to privacy. The fragmented nature of privacy regulations also reflects the ambiguity of unresolved intuitions about commonplace occurrences like exploiting data outside of the context for which it was gathered, in addition to the competing interests of various parties involved. For developing theoretical and model solutions to the aforementioned issues, the contextual integrity heuristic can be an intriguing research area for information transfers both from the legal–ethical dimension and also real-time practice. Protection of security and privacy has been given top priority by public policy for many years. The public and governments have been increasingly concerned about privacy as a result of rapid technological breakthroughs, the rapid development of the Internet and online commercial channels, and the advancement of better ways to gather, examine and exploit personal data. Due to how simple it is to gather and store vast volumes of data using a computer system, data mining has become increasingly important.

A substantial amount of personal information was recently made public because of the massive amount of data amassed through multiple media. It is important to consider if the release or analysis of sensitive personal data will violate the privacy of the person whose data it is, Cambridge Analytica obtained the private and sensitive information of millions of Facebook subscribers. Since none of Facebook's systems were breached or passwords stolen, the mining of millions of records was not caused by a data breach but by using the data (that was collected with consent) out of the context of the purpose of collection. As a result of Facebook's systems operating as intended and Cambridge Analytics employing a third-party application to collect, extract, and use the data for political campaigns, there was a violation of people's privacy (Chan, Rosalie,2020). Another example of such data misusage is “In re GOOGLE INC. COOKIE PLACEMENT CONSUMER PRIVACY LITIGATION”, where Google violated Rule 23(b) of the Federal regulations by exploiting the cookies for advertising purposes and endangering the context in which they were acquired (In re GOOGLE INC. COOKIE PLACEMENT CONSUMER PRIVACY LITIGATION., *2015)*.

The privacy of users must be protected for a variety of reasons, including the abovementioned data breaches and countless actual use cases. These topics that get a lot of attention include healthcare and banking. In the health industry, the use of infrastructure that is supported by information technology has grown significantly, enabling the storage and analysis of patient data that contains highly confidential and secret medical files. Because the obtained data is used by both the primary service provider and various third parties. Widely employed in the healthcare industry, the Health

Insurance Portability and Accountability Act (HIPAA) was primarily created to safeguard patient data privacy.

Central location data is considered personal under the General Data Protection Regulation (GDPR), among the most important laws. This suggests that users should directly and voluntarily consent to location tracking rather than choose to withdraw consent. Due to this, several recent research publications have discussed privacy-preserving techniques and proposed novel approaches that allow users' privacy to be safeguarded while information typically extracted is taking place. Some of these policies are aimed at protecting personal privacy, while others are aimed at protecting business privacy.

There are significant discrepancies among privacy experts over how to evaluate, enhance, and regulate current industrial practices for better protection of personal information, even though few academics would contest the significance of information privacy. The interdependence of privacy and information technology necessitates a highly interdisciplinary approach to investigating information privacy concerns from several angles. The information scientific establishment, in my opinion, is particularly well-positioned to contribute to the present privacy debate and to reshape the problem space with fresh concepts. In fact, a cursory check of publications from the past ten years reveals that they have addressed privacy issues in a range of statistical contexts, such as health apps, social networking sites, and platforms (Squicciarini, Xu, & Zhang, 2011; Stern & Kumar, 2014), as well as new approaches to model and measure privacy in scholastic research, these works represent a wide range of intellectual traditions in the society and show varied perceptions of how Information and communication technologies and privacy interact.

However, there are still gaps in the studies. Particularly, despite the variety of intangible assets used in privacy research, there has been the little synthesis of these resources in the formulation of workable and novel privacy-enhancing solutions. For instance, it is widely acknowledged that social networking sites' privacy settings fall short of users' expectations for privacy (Wu, 2019, p.206-216), but few studies have yet proposed and empirically tested alternative designs for better control of privacy parameters (with Stern & Kumar, 2014, as a notable exception). Similarly, to this, sociopsychological researchers have discovered a number of variables that influence how people perceive and behave concerning their privacy, but it can be challenging to turn these findings into specific policy recommendations (Acquisti & Grossklags, 2005, p. 24-33). We demonstrate how this can be converted into a useful concept that can help public bodies or business corporations assess what privacy implications or risks might be associated with making a specific data file available as public information by using the well-known theoretical framework for privacy evaluation, Contextual Integrity.

We will give a thorough literature study of established protection measures, security problems involving personal information, and contextual integrity heuristic approaches for the proposed model in this report. The key ideas utilized throughout the paper are defined broadly in Section 2.

In Section 3, we go into further detail about the literature review methodologies that are outlined by the CI theory and topics for the proposed model. Furthermore, we will look into the related works that help in building the model idea with a conceptual meta-model. A brief summary of some of the most important personal data privacy issues and decision-making ambiguities is provided in Section 6. Section 7 concludes the notion of the suggested model and outputs.

2. BACKGROUND

Several of the least comprehended and contentious threats to privacy in the age of information technologies is the practice of mass monitoring and data exchange. In addition to the conflicting interests of several entrenched entities, the fragmented character of privacy laws also reflects the ambiguity of unresolved intuitions towards everyday occurrences. This section shows why some of the well-known theoretical methods of privacy, which were developed over time to address traditional privacy concerns, generate unsatisfying findings in the case of public surveillance. It builds on past work on the issue of expectations of privacy. Before going any further, let's look into some example cases that highlighted the importance of privacy in the context of regulations and ethics.

2.1 Facebook – Cambridge Analytica Case:

The method by which Cambridge Analytica obtained its data from Facebook is the more crucial aspect of loopholes in privacy regulations. Additionally, researcher Aleksandr Kogan, a Russian American who worked at the University of Cambridge, sent it to the company, claims a former Cambridge Analytica employee. Kogan created a quiz-based Facebook application.

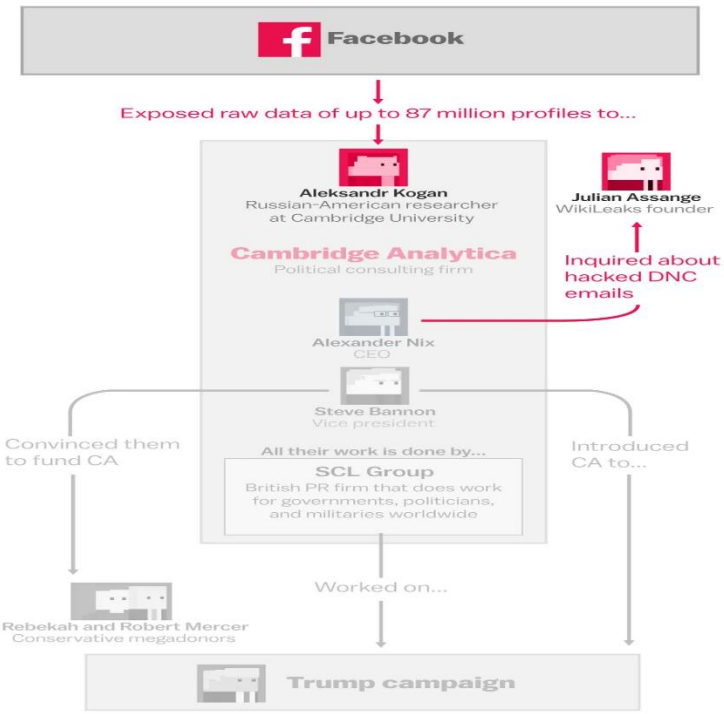

Figure 2.1.1: Using a straightforward flowchart, the Facebook and Cambridge Analytica controversy is described (Chang, 2018)

In addition to gathering information from quiz takers, it also discovered a flaw in the Facebook API that allowed it to gather information from the quiz takers' Facebook friends. Facebook forbade the monetizing of data gathered using this technique, yet Cambridge Analytica nonetheless did so. Up to 87 million Facebook accounts' raw data were collected by Cambridge Analytica, a political consulting business that worked for the Trump campaign (Chang, 2018).
The important thing to note from this case study is that users were asked for consent before taking the quiz. Irrespective of the consent it is unethical to use that data out of the context for which it was collected. Facebook made a rational decision where the loophole was exploited by selling its data to a political campaign by Cambridge Analytica.

2.2 In re GOOGLE INC. COOKIE PLACEMENT CONSUMER PRIVACY LITIGATION case:

Google, a division of Alphabet Inc., was charged with using security flaws in the Safari and Internet Explorer browsers from Apple Inc. and Microsoft Corp. to assist marketers to get around cookie blocks. Tracking cookies were found to be immediately deleted when a user visited the domains, in violation of the country's Data Protection Act, after the regulator investigated the websites over the course of the previous year. Users of Google anticipate that we will protect their privacy, which was not clearly the case. As in the context of Google, the Commission also discovered that the technique only partially worked when a user chose to deactivate personalized advertising — via an option that Google's cookie notice supplied them with — as one advertising cookie remained retained on their computer and continued to process data in blatant contravention of the consent requirement (Lomas, 2020).

From the above cases, we can observe the “respect for privacy” is violated and the need for context-aware privacy has become an important ethical and practical way to preserve privacy.

2.3 CONCEPTS:

Let’s look at the concepts that help us better understand the need for this approach.

2.3.1 Personal Data:

Any information that can be used to identify or contact an individual is considered personal data. Personal data also refers to many pieces of information that, when put together, can identify a specific person. Regulatory bodies such as GDPR, PIPEDA, or others, apply to personal data that has been de-identified, encrypted, or pseudonymized but can still be used to re-identify a person. Individual personal information is no longer regarded as personal information when it has been made anonymous in a way that makes it impossible or impossible to identify the individual. Data must be fully anonymized for anonymization to be effective.

2.3.2 Privacy:

The area of data management known as data privacy is responsible for processing personal data in accordance with all applicable laws, rules, and practice guidelines for privacy. Setting access controls to guard against unwanted access, obtaining agreement from data subjects where required, and preserving data integrity are all part of ensuring data privacy. Certain data, including retail business, project roadmaps, and financial information about a company, could be considered sensitive. Information about persons, specifically confidential details about any recognized or recognizable human, is among the most sensitive data. Nearly everything can be personally identifiable information. Unlike a name or Social Security number, it isn't always immediately apparent. Sometimes it's a different identifier, like a cookie or an IP address. Data is considered personal if it may be used to identify a specific user from a field or data.

It is impossible to exaggerate the significance of data privacy in the commercial world of today. Personal information, including credit card numbers and medical histories, is governed by data privacy rules in the majority of the world.

2.3.3 Contextual Integrity:

This specific conceptual chasm can be bridged by adopting Carpenter and Dittrich's proposal to reframe research protection standards from "human subjects" to "human harming." There may be ways to balance the ethical challenge with the objectives of the research initiatives in each of the conceptual gaps mentioned above. Nissenbaum's (2004) thesis of "privacy as contextual integrity" is a helpful heuristic to help researchers make ethical decisions in big data research initiatives. A conceptual framework that connects the protection of personal information to the rules of personal information flow within particular settings is known as contextual integrity. It is a benchmark theory of privacy.

Nissenbaum seeks to explain why these normative constraints and the public/private dichotomy with respect to information privacy left out significant interests in information flows that are present in all social relations and are not constrained by the public/private dichotomy. A theory of information privacy known as contextual integrity contends that social relationships are governed by norms in information flows and that privacy is endangered whenever these norms are broken. The norm of appropriateness and the norm of flow or dissemination are the informational norms, whose integrity is necessary for privacy (Joshi, 2020).

1. Appropriateness: Specifies what information is appropriate to share in the given situation and illustrates how people differ in the informational types they share depending on the relationship.

2. The norm of flow distribution: This theory contends that information sharing should and must preserve context and clearly express valid consent.

2.3.4 **Privacy in Networked Contexts**:

When seen from a contextual angle, privacy can be defined as the process of enforcing boundaries within different social contexts. As the social environment changes, the boundaries may move, disappear, or reappear (Wu et al., 2019). Network privacy is the idea that people don't have complete control over how and what information about them is shared online, and that users of the platform work together to regulate privacy.
This demonstrates how different platforms' capabilities enable varying results, with some sites enabling significant visibility of material and others enabling a greater degree of invisibility. It is anticipated that users participate in a range of privacy and security tactics, constant curating of networks and material, and use more private platforms for sensitive disclosures due to the unique partnership of privacy in these places.

2.3.5 **Privacy Compliance:**

Context is important for resolving regulatory issues in a worldwide environment as well as for comprehending people's privacy demands and behaviors. To preserve people’s privacy, governments have debated whether and how to control information flows across international platforms and services. A complex terrain of international laws and policies for the protection of privacy and the movement of personal data across networks has developed as a result of the diversity of interests, histories, and cultural contexts (Greenleaf, 2017). Some nations, including Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) and the European Union's General Data Protection

Regulation, have chosen to adopt comprehensive and relatively rigorous rules governing the collection, use, and sharing of personal information (GDPR) (Wu et al., 2019).

Ambiguity in transnational regulatory laws:

Given the growing flows of personal information between transnational networks and across borders, the disparities in regulatory approaches to privacy—and the underlying contradictions between different jurisdictions' views toward the protection of the personal data of individuals are even more problematic. Users of the world wide web and its services, and corporations can access their goods and services from all over the world, and the data processing and storage infrastructure is also dispersed.

2.4 Architecture for searching encryption

In order for the server to recover the encrypted data including the searched term, data encryption that can search enables users to create a search token from the searched keyword. The search token essentially represents an encrypted query over the encrypted material and can only be produced by those who have the required secret key. The architecture primarily consists of four entities: the secret key, the cloud service provider, the content owner, and the data consumer.

- Key generator: The development and management of the encryption and decryption keys are handled by this organization, which is regarded as a trusted third party. During the system setup, user-specific keys are generated and disseminated.
- Cloud service provider/ cloud server: This organization offers subscribers access to cloud-based data storage and retrieval services. Cloud service manager and cloud data server make up the cloud service provider. The first entity is used to store encrypted data that has been outsourced, while the latter one is utilized to maintain data on the cloud. The cloud service provider runs tests on the encrypted meta information in the cloud storage after receiving the encrypted search queries from the data users. Upon successful completion of the test, the encrypted data is retrieved and sent back to the original data owner. The vendor of cloud services shouldn't gather any knowledge from the operation.
- Data user: The mentioned entity uses encrypted queries to search for specific encrypted data. It is a subscriber to the cloud storage system. There may be multiple data users in a system, and in certain cases, the data owner and the data user may be the same person or company.
- Data owner: The entity that creates, encrypts, and uploads the data to the cloud server is the data owner. Either a person or an organization could be the culprit. The owner of the data utilizes its application, which includes a data processor for uploading fresh content to the cloud, in order to access the service. It uses a cryptographic algorithm to encrypt the data and metadata, enabling searching.

2.5 **Security Requirements for encryption**

• The server wouldn't be able to tell which documents were which or identify the contents of a search.

The server was unable to find any information on the term being searched for. The server can only return pointers to the encrypted content that contains the keyword when given a token.

• A coded query should not be able to be generated by the server. Just those users who have the necessary secret key can create the query.

• Search query result- The server shouldn't be made aware of the search result's contents.

• The server shouldn't be made aware of the order or frequency in which a user accesses a particular document.

• The server was unable to identify whether two tokens were part of the same query.

2.6 Design approaches for Encryption and decryption

Either a non-quasi-based technique or an index/keyword-based approach can be used to create an accessible cryptographic protocol. This approach involves word-by-word scanning of the entire document to identify the relevant word. This makes it possible to search the document for any word. However, searching through a huge document set takes a while. The index approach, on the other hand, creates an index [x] for each key term of interest and lists the relevant texts that contain that word. When the document set is huge, this makes searching faster. The index's storage and upkeep can, however, be a burden. This scheme can primarily be depicted using either an asymmetric or symmetric setup from the perspective of choosing a cryptographic technique. The differences between these two settings are mentioned briefly in the sections that follow.

- **Asymmetric Searchable Encryption (ASE)** is a configuration where users encrypt data using an asymmetric/public key cryptography algorithm (such as RSA) before offloading it to cloud servers. This configuration works well in situations where the data is being viewed by someone other than the one who created it. For example, many users can use a particular user's public key to encrypt and upload data, but only that user's owner with access to the associated private key can create a search token, resulting in a cryptographic You can search for formatted data. In contrast to its drawbacks, ASE's main advantage is its functionality. The ASE scheme is more versatile, as readers and writers can be different in this scenario. However, unlike hash functions and block ciphers, all known ASE schemes require evaluating pairings on elliptic curves, a time-consuming and costly process (Kamara,2010). Public-key cryptography has been used in several studies to create SE schemes.
- SSE, or symmetric searchable encryption - In this configuration, the user encrypts the data before dumping it to the cloud server using a symmetric/private key encryption strategy (like AES). When the user accessing the data is also the user generating the data, this setting is advantageous. This setting's primary benefit is efficiency, but because it can only be applied in single-user situations, it lacks functionality. The

majority of SSE schemes also expose access patterns. Because most SSE techniques are built on symmetric primitives like block ciphers and pseudorandom functions and require little computational work, cryptography is efficient. The SSE system, which was initially put forth by, offers methods for remotely decrypting data using symmetric key primitives. Later, a finer definition of security was added to the SSE schema security concept.

3. LITERATURE REVIEW

In the current digital-centered learning, all of these Westin's theories can be disputed. Humans end up leaving information traces that are captured, analyzed, and exchanged with or without our knowledge as our routine actions are made easier like purchasing, and occasionally deeply implanted like social networking. As a result, people rarely have a full understanding of the knowledge about themselves that is available. Additionally, privacy policy execution and innovation have been backward tech innovations. For instance, despite the US federal trade commission's 2013 suggestion for smartphone owners to have a single confidentiality dashboard to evaluate information accessed all over mobile applications, the industry has not yet broadly accepted this idea. A cross-application platform and a strict confidential panel are very difficult to materialize as digital enterprises build walled gardens to trap customers and keep their edge. It's also crucial to keep in mind that people have much less control over the details regarding themselves in today's hyper-connected world (Floridi, 2015), as data is comanaged with colleagues, relatives, and other people who can share your sensitive information to various of digital portals.

For instance, Richter Lipford discovered that when photographs are shared throughout their numerous overriding societal boundaries, picture labeling on SNSs diminishes individuals' authority over their information disclosures. Information privacy still has several important components, including possession, accessibility to, and transmission of confidential information. But in the modern world, privacy protection encompasses more than just who gets access to what information. The technical ability to analyze massive amounts of data from various sources to find trends in purchasing, lifestyles, sexual orientation, political preferences, and more is a significant breakthrough in recent times (for example, Ohm, 2009). A user's anonymity is in danger not just because data about them could be conveyed to others without their permission, but also because existing dots can now be joined quickly to reveal personal information concerning them.

Last but not least, Westin's concept makes the assumption that an educated and logical human being is able of choosing the optimal course of action for their security under various circumstances, but research shows that this is rarely the case (Acquisti, Taylor, & Wagman, 2016). Visibility and information asymmetries can make it difficult for people to receive accurate and

complete data for decision-making. Furthermore, cognitive biases and shifting tastes are known to lead people to make bad decisions. For instance, when weighing the advantages and disadvantages of disclosing personal information, people usually choose to prioritize short-term rewards above long-term, including implications. Numerous investigations have shown no consistent challenges in selecting the optimal security trade-off in different situations.

An increasing number of confidentiality academics are pushing for a more situational method of approaching information confidentiality, stressing the significance of comprehending, and honoring the circumstances and context that influence people's decision to reveal confidential material. This is because they recognize that the disclosure situation in Westin's conceptualization of confidentiality doesn't really fit with the digital world of confidentiality presently. Helen's thesis of privacy as contextual integrity, which connects the safeguarding of sensitive data to the rules of information goes through certain settings, serves as one of the theoretical underpinnings for this method.

Contextual integrity has been studied in relation to a number of confidentiality situations, including browsers, social networking sites, area-based innovations, e-health records, student learning analytics, and smart home appliances. This research suggests that infringements in contextual integrity can tell us why customers would be worried about the usage of data that extends beyond the initial intent or background in which it was at first reported. This research has also discovered more nuanced interpretations for perception inconsistencies or paradoxes in confidentiality behaviors.

We support a more comprehensive contextual understanding of security at all levels of analysis individual, social, and societal given the fundamental role of contextual integrity in the research of security. Security in connected environments, security for disadvantaged people, and privacy in a worldwide regulatory environment are three particular situational factors that are expected to influence future directions of security study.

The quest for a normative framework entail looking for concepts and standards that provide justifications for supporting or opposing particular broad policies and for resolving specific issues. Understanding why prevailing ideas, which have shaped most of the modern privacy policies provide little advice in many complex situations.

Helen discovers that three principles dominate public discussion of privacy when we examine the realms of public policy development, regulation and statute legislation, court decisions, and social and economic behaviors during the twentieth century.

The three guiding principles focus on:

(1) limiting public monitoring and the use of personal data about individuals by government officials.

(2) limiting access to sensitive, private, or confidential information; and

(3) limiting incursions into areas considered to be private or personal (Nissenbaum, 2004).

Researchers want to enhance privacy engineering, which is described by Kenny and Borking as "a systematic attempt to embed privacy-related legal primitives [concepts] into technical and governance design." (Kenny & Borking, 2002). We should contend that in addition to the incorporation of legal fundamentals, privacy engineering must also take into account philosophical, sociological, pedagogical, and other viewpoints. It is understandable why Lahlou, Langheinrich, and Rucker discovered that engineers were extremely hesitant to embrace privacy: "Privacy was either an esoteric problem, not a problem yet (as prototypes), not a problem at all (access controls, firewalls, and cryptography will indeed keep hold of it), not their problem (but one for lawmakers, legislators, or, more obliquely, society), or simply not part of the project deliverables." (Lahlou and colleagues, 2005, p. 60) Once it comes to human-computer interaction, the term "privacy" is frequently interpreted incorrectly and used inappropriately (Barkhuus, 2012). As a result, it is necessary to agree on a small number of core privacy theories and frameworks to direct privacy research and design (Badillo-Urquiola et al., 2018).

3.1 Formalizing and normative strengthening Contextual Integrity

In their initial attempt to formalize a portion of CI, Barth, Datta, Mitchell, and Nissenbaum (2006) concentrated on "communicating agents who take on different roles in contexts and send each other messages including attributes of other agents" (Barth et al. (2006), p. 4-6). Nissenbaum offered a nine-step decision heuristic in 2010 to analyze new information flows and ascertain whether a new practice would constitute a prospective invasion of privacy. For the first time, she specifically stated in this heuristic which concepts needed to be described to satisfy a CI evaluation (Nissenbaum, 2010, pp. 180-182). In circumstances when there is no established definition of privacy, it is difficult to translate existing norms and values. A good example is using digital platforms. Understanding personal privacy concerns necessitate a contextually rooted grasp of the circumstance and society, not just a known collection of contextual factors (Barkhuus, 2012).

There are a number of research gaps that have been shown by this analysis of how context has risen to prominence in privacy architecture. We have discussed how the advent of artificial intelligence and unrivaled access to processing power has created new opportunities for synchronized and continuous privacy decision-making.

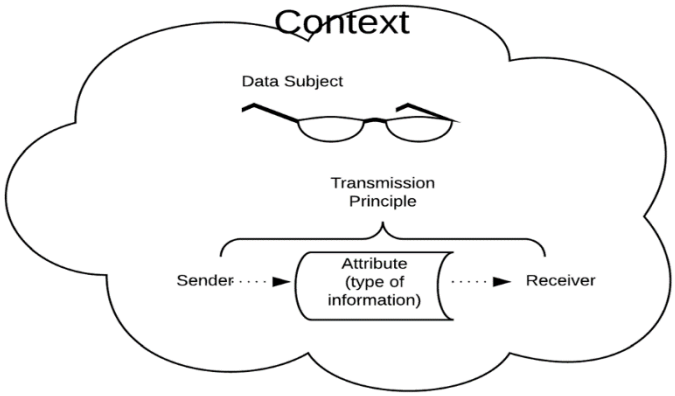

Figure 3.1.1: The theory of contextual integrity (Nissenbaum, 2010)

To make this possible, we must expand the definition of "context" as a notion that encompasses more than merely framing present preference settings. Giving context a crucial role in IT systems that satisfy user expectations, means advancing the concept of contextual integrity in the direction of a prescriptive theory. As a result, we have determined design difficulties at the theoretical, procedural, and functional levels.

3.2 The objective of ML – Technology's role in privacy:

A contextual and dynamic strategy employing ML was adopted in place of a method based on pre-programmed rules. By learning from data, this system lets computers carry out particular jobs appropriately, and the system keeps becoming better at producing accurate results over time (Shalev-Shwartz & Ben-David, 2014). A drawback of ML is that it might be challenging to create systems that have a contextual perception of a subject.

Attempting to make contexts the very focus of machine learning and emphasizing context cues as the fundamental idea of supervised machine learning. This indicates that in order to train the system, we will require a specific volume of labeled data, and the system will need to be adaptive—that is, able to receive further training—once it is put into use. According to the user's online activity, the system needs to be able to determine when to bring up a context trigger. As a result, it must first create a library of possible events before learning what triggers them and how each user would react to them (Hoel et al., 2020).
The administration of the data-sharing policies will also involve ML. To enable acceptable data-sharing streams, it is beyond the scope of this work to investigate how all the various data-sharing policies a user will encounter could be condensed into a structured collection of policies that the user could add to and edit. In the suggested system, machine learning is

utilized to generate training data, which is then used to generate new data for the system's ongoing improvement (see Figure 3.2.1).

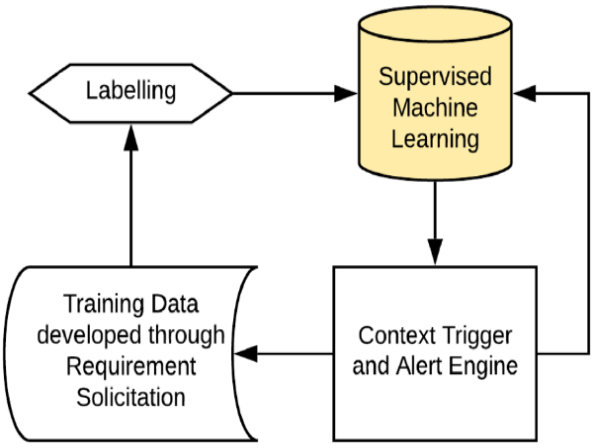

Figure 3.2.1: The contribution of machine learning to the contextually negotiated policy system (Hoel et al., 2020).

Figure 3.2.2 shows how policy documents are translated into a conceptual model that can be customized to reflect each person's chosen data-sharing arrangement. This structure is then utilized to create a smart contract that will regulate data sharing and ultimately affect service providers' practices (Lyons, Courcelas, & Timsit, 2018).

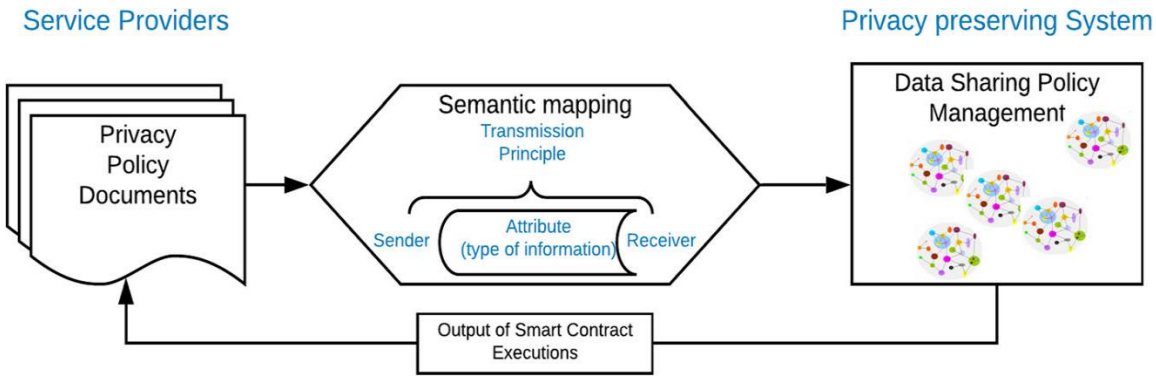

Figure 3.2.2: mapping personal data sharing agreements to privacy policy documents (Hoel et al., 2020).

The user-subscribed data-sharing policies are carried out through smart contracts. For instance, these agreements may permit the sharing of one's data with third-party businesses that possibly execute specialized analyses for the service provider to employ. Yet, if a warning goes off about one of these organizations being associated with a data breach issue, a tweak in the smart contract might be made to prevent a further data transmission.

4. RELATED WORK

Contextual integrity (CI) theories and models are designed to give governmental organization professionals a quick-view visual representation of CI. It will show how a tiered method of identifying the extent to which a specific file includes possibly privacy-sensitive information and/or may pose a threat of anonymity getting breached may help with generating knowledgeable privacy choices. The creation of prospective privacy risk judgment solutions that could simplify and simplify portions of this procedure for professionals in the future can also be informed by CI analysis.

The Unified Modelling Framework is the design approach that will be employed to graphically depict how CI can be implemented in practice (UML). This approach was chosen to visually display and demonstrate each process of the contextual model because UML flowcharting is a global visual medium that is employed to encapsulate and portray notions and the connections among each other (Rumbaugh et al., 2004). This offers users an incredibly simple source of reference and a greater understanding of how the components connect inside every step (Fowler, 2004). Two sets of governing concepts from Nissenbaum's architecture that were determined to be essential for putting CI into practice were used to develop the meta-model:

Key elements: To better align with definitions that professionals will connect to and support the stream of the program leading that the metamodel will be requesting professionals to pursue, the metamodel substitutes Clarification, Risk Analysis, and Decision for Nissenbaum's Justification, Assessment, and Prescribing these three components have been employed as the meta-overarching model's concepts to divide the model's conceptual growth into manageable sections.

Nissenbaum offers nine decision heuristics (DH) that need to be considered while analyzing both current and emerging data flows.

First phase: explanation

The practice or approach that will be evaluated is mentioned in the description section. These should be evaluated in light of any potential violations of "context-related information norms." This evaluation must take into account the main "actors," or the individuals who are/could be impacted, as well as their roles as data subjects, information senders, or data receivers. The open dataset, which refers to the information itself (also known as "the data") and how this data is

transferred (also known as "transmitting principles"), should also be taken into account. Any adjustments to these components may potentially violate the current or proposed new flow of information.

It was established that the initial four DHs pertain to getting a somewhat more thorough understanding of the data, the persons (actors), the current informative standards, and transmitting guidelines in order to include the DHs into the meta-model. The very first decision heuristic (DH1) is focused on the information itself and the suggested transmission method. The second demand we consider is the current scenario, atmosphere, and individuals around the data (DH2). The third (DH3) focuses on the individuals who interact with the information, and the fourth (DH4) aims to determine how the information is currently transferred (Nissenbaum, 2010). In order to illustrate the explanation aspects and how they interact, these four DHs were utilized. In order to provide a more comprehensive understanding, specific information regarding the data itself, the participants and their functions, the data transmission method, and the backdrop must be gathered. Consequently, the superclass and each of the items below is shown as subclasses in the explanation (see Fig. 1). The link between the subclasses is depicted in Fig. 1 and can be described as follows:

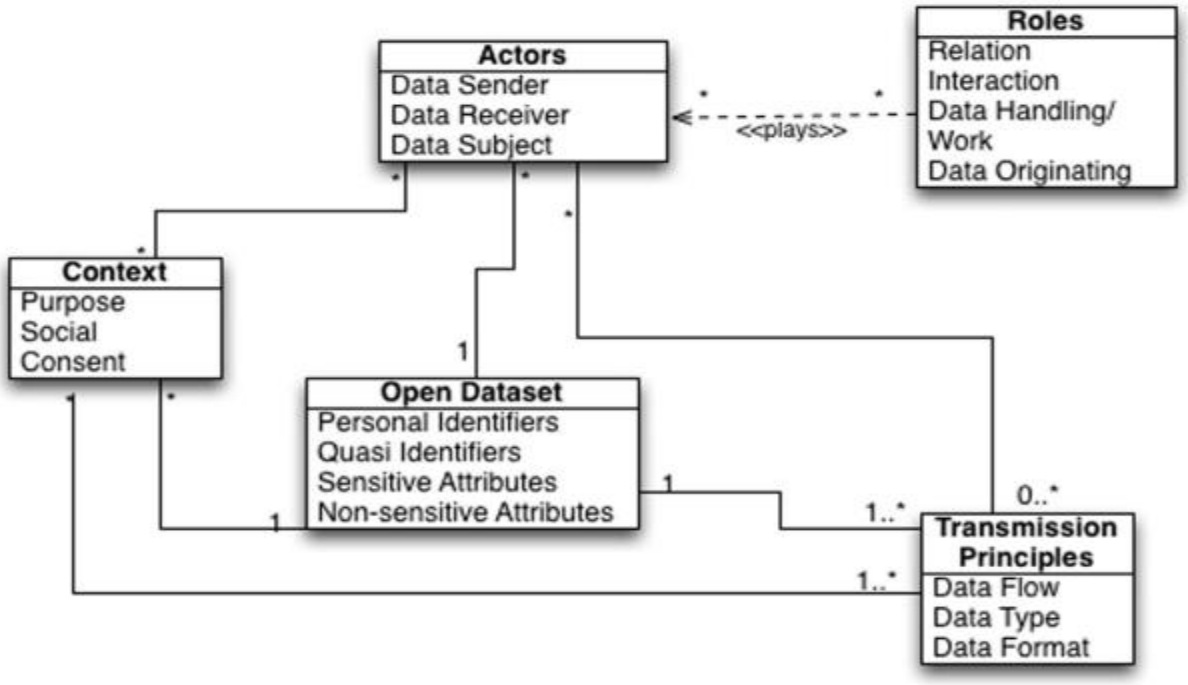

Fig.1- explanation (Henriksen-Bulmer et al., 2019).

Open dataset: To prevent convoluting the judgment procedure, make sure all factors are carefully considered, and keep CI all through the evaluation of each set of data, each set of data must be

considered into account individually. The characteristics for each unique data element in the dataset will be sorted by attribute type: Identifiers, or information that can be used to specifically identify an individual such as their identity, social security number, or date of birth; Information that does not explicitly identify a person but is probable to do so if connected, such as age or gender (Thomson et al., 2005) vulnerable features, or person-specific information that might be used to recognize a user, such as income, ethnicity, religion, or a medical condition (Fung et al., 2010); Non-identifying, non-sensitive features, even though connected.

Actors: Each actor will take on a variety of roles. The actor will play one of three data transfer roles at the data level: transmitter, recipient, or subject of the communicated data. Another possibility is for an actor to play many roles. For instance, the information received may also be the data provider or the data subject if they download the information. The actor will also play numerous parts in relationships and/or at the workplace in addition to the information transfer duty. The responsibilities have been divided into different categories to allow for the consideration of these complexities.

Roles: Based on the statistics, each actor plays one or more responsibilities. As a result, each actor may play many roles, that could be determined by their connections, the situation, or their job descriptions, such as Connections, or specifics of the ties between both the actors, such as those that are either personal or professional; Data on how the actor(s) communicate, such as data about exchanges between friends or citizens and professionals; Details regarding the actor's inputs and outputs during information management; Work, or the actor's profession or job description, is what they do for a living. Data Originating, or information on the creator of the information, which may be a third party; as a result, the role of that third party must also be considered.

Context: Gathering details regarding the circumstances around the initial data collection, including how and why it was done (the prevailing context). Such contexts include the following: Intent, or the initial reason the information was obtained; Social, that is, the social setting in which the information was gathered (for instance, the welfare agency is a setting for collecting taxes and payments); Approval, or if it has been acquired, and if so, its legality, must also be taken into account.

Transmission principles: describing how information moves among actors.

Phase II: risk analysis

The examination of any security hazards connected to a specific activity or communication inside a specific context, taking into consideration how the data is shared or transmitted, and the actors connected with that practice or communication, constitutes the second major component,

risk analysis. In actuality, the goal of the risk assessment is to determine the hazards related to any suggested modifications or adjustments to the transmitted data.

Norms are the guidelines or norms that guide how we lead our lives, whether formally or informally. An instructor may disclose a student's performance record to the pupil or their families, but they wouldn't be anticipated to disclose the same information to other family members within the school. Information-based norms, or the social rules that the actors will be obligated to follow by the ability of their position, are what needs to be taken into account in aspects of CI.

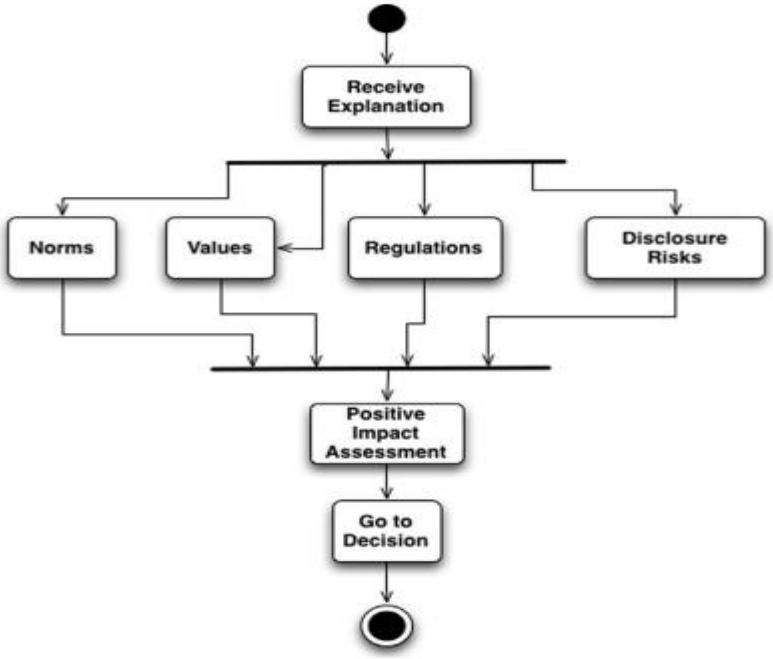

Fig.2 - Activity diagram (Henriksen-Bulmer et al., 2019)

Values: These are the things that a context is focused on. It is conceivable that the newly proposed data transmission may create any kind of unbalance and so violate one or more of these ideals. These morals may be societal, economic, or moral and might be influenced or transformed as a result of the suggested new data flow. obviously, such principles will have to be considered in relation to the standards and any duties imposed by laws or regulations.

Regulations: Any requirements that the government entity may be subject to in connection to the information. The re-use of Government Sector Data Laws of 2015, which mandate that government agencies disseminate data in open platforms, may impose restrictions like DPA guidelines or make it mandatory.

Risk of disclosure: The data from the explanation phase is examined, and it is determined what the disclosure danger will be in the context of the following factors: the engrained information-based norms; any coordinates of defection from the engrained information - based standards recognized; the effect of any possible violation; any suggested prevention measures put in place and any suggested controls in place or to be put in place.

Third phase: decision

The judgment made regarding if a practice infringes confidentiality is the third crucial component. In order to determine whether an activity or method could potentially represent a threat to anonymity, the results must be presented. This, it is argued, entails deciding whether the data is compatible or incompatible in order to let those modifications or changes in the data flow. The final heuristic has been employed in this step to determine if contextual integrity advises for or against the newly proposed practices based on the results from the preceding analyses. There are just two phases in the decision-making process. The activity flow for this stage shows the choice on its own and documents the result.

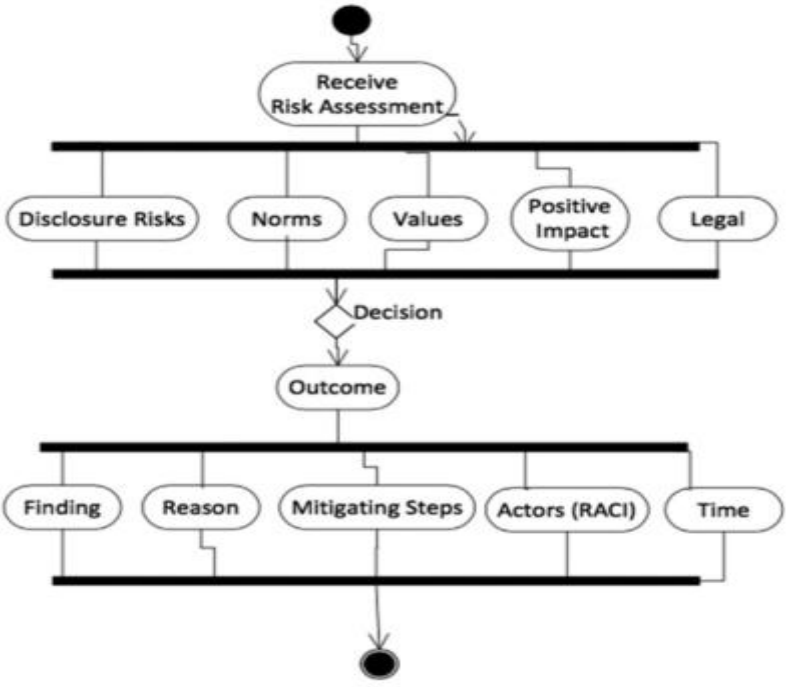

Fig.3-Decision diagram (Henriksen-Bulmer et al., 2019).

Risk assessment: The Phase II Risk Analysis will be continued. This will influence the decision-making process and result in a decision. This result will include the choice made for each characteristic category, giving the professional a rating to help them decide.

Result: Once a choice has been made, the result must be documented and any subsequent actions must be listed as follows: Determining the final choice, such as whether to broadcast or not to disclose; and Explanation, or the justification and justification for the choice taken, may refer, for instance, to compliance with the law, such as Data Protection, or it may refer to a finding that there are no privacy concerns.

mitigation measures This section will describe any preventive actions that must be taken prior to publishing might occur; these actions might include deletion or anonymization, for example.

Actors: In order to establish openness and give assurance that the right procedure is used when formulating, adopting, and executing choices made, the procedure should include a track of who is liable, accountable, contacted, and instructed (RACI) moving ahead (Henriksen-Bulmer et al., 2019).

5. PERSONAL DATA PRIVACY ISSUES

When it comes to guiding your business toward growth and innovation or governing legislation, data privacy should be at the tip of your focus ranking. Making sure that third parties cannot access, utilize, or spread your private user data, safeguarding your business operations or services, and maintaining the reputation as well as respect to the privacy of the data collected is essential components of both of these. To strengthen the data protection methods, we'll analyze which data are vulnerable to attacks, explain major data privacy problems, and go over how to fix them.

Data that are most vulnerable to privacy breaches are the sensitive information that can cause harm when exposed. This information can be used to identify an individual who is at risk of data loss, financial loss, reputational damage, or physical or mental harm.

5.1 Data Trading

Personal information may be accessed and stolen by third parties, sold to other third parties, or continued to be sold and resold until pertinent exposures are fixed as part of data trading.

One of the cornerstones of your data privacy strategy should be preventing unwanted access to your sensitive data as well as its potential sale to other parties. Once data brokers acquire access to your business's or customers' personal information, they can carry out a number of potentially detrimental tasks, including:

- **Identity theft** - If hackers have access to enough private customer information, they may pose online as your company or your clients for their gain. They might make illicit purchases, apply for loans using your federal tax ID number, or make electronic transfers from your bank accounts.
- **Data detainee/hostage** - In severe situations, information brokers will keep your data in captivity in exchange for a ransom, inviting you to the negotiating table.

This can also lead to accepting bids from other competitors as they wait for your response.

- Data traders can sell your personal information to marketing firms so they can build adverts specifically catered to your purchasing behaviors, digitally saved searches, and web query engines.

5.2 Lacking Information Privacy Protocols

The business should undertake privacy risk assessments to understand the existing and potential privacy hazards from those practices to both the organization and the individual consumers. This is done once an organization has a basic grasp of its data collecting, usage, and sharing. Organizations can conduct as many individuals or group assessments as necessary to assess the privacy policies of their business operations. It is advised to conduct a privacy risk assessment to identify privacy issues early, improve internal data that is available to support informed decision-making, prevent complexities and costs or humiliating privacy compliance errors, and show that it is trying to reduce privacy risks and issues. Proper IT privacy policies and protocols should be in place and must be updated with every assessment of privacy compliance.

5.3 Inadequate SOPS

Humans are nevertheless prone to error, even with the strongest data privacy technology at their discretion. As a result, companies shouldn't rely just on technology to safeguard their data. Additionally, businesses need to create and improve standard operating procedures (SOPs) for data privacy. SOPs must outline protocols for new device setup and privacy protection, personnel equipment protocol, record name and filing norms, and more.

5.4 Data Retention and Redundance

More information increases the likelihood of unauthorized access. If your business is storing digital records on servers, on the cloud, or individual devices without a need, you should perform some spring cleaning to get rid of any unnecessary or out-of-date files to avoid privacy concerns.

The essential items that should be destroyed in order of priority when trimming your data collection are the redundant files, records for programs that are stale or not being utilized, old non-financial papers, etc.

6. REGULATION AND ETHICAL DIMENSION:

In this chapter, we'll examine the regulatory organizations that work to safeguard privacy using real-world examples, as well as the ethical issues that businesses should consider when developing technology and business models to secure user privacy.

Technologies ascended in importance to adhere out of the indescribable chaos of web services. Their developers were motivated by the freedom it promised, but they also wished to create spaces for the greatest and most social features of the web. But when these venues expanded, the pandemonium returned. Websites and services that host keynote addresses, save them in the cloud and servers, structure accessibility to them through search and recommendation, or download it to mobile devices. This covers web browsers, online platforms, social networking websites, etc. Their common offer, which is to store and compile user content for public consumption without creating or commissioning it themselves, is what ties them all together. Although they don't create the content, they do make significant decisions about it, including what they will distribute and to whom, how they will connect users and mediate their activities, and what they will reject, this is an expanding and more potent group of digital intermediaries. It's time to rethink platform responsibilities. This should include Safe Harbor reviews tailored to social media platforms rather than whole laws designed for ISPs and search engines (Gillespie,2017). This includes articulating realistic expectations of what the platform can and should achieve from a legal, cultural, and ethical perspective. These are clearly the rules of the game, and we have the right to enforce them. We also provide details on the inner workings of the moderation process, detailed information about who reports complaints and how they are resolved, and how they do it, including greater transparency about the employees involved. We need to include new standards of openness and accountability for what we do. Be more open about how and why rules are created.

First, let’s take a look at some of the regulatory bodies (focusing on Canadian laws) that strive to preserve privacy:

6.1 PIPEDA: Personal Information Protection and Electronic Documents Act

The PIPEDA law first became effective in 2000. It has undergone various revisions and since, the most substantial of which was made in 2015. The law, called the Digital Privacy Act, strengthened the authority of Canada's Privacy Commissioner, and imposed obligatory incident reporting obligations.

Rules in PIPEDA:

- One should designate a representative who will be in command of your firm's adherence because you are accountable for the personal details in your care. This individual is referred to as the chief privacy officer under PIPEDA (CPO).
- One must only gather the minimum amount of user information required to fulfill the objectives put forth by your business (such as analytics, remarketing, or A/B testing). You must make sure that the personal details you collect from your users are true, complete, and current.
- Without permission, you are not permitted to collect, use, or disclose of personal data.

- The reason for gathering the data and the methods you intend to use to analyze it ought to be made explicit.
- The sorts of information you gather, the parties with whom you share it, and any possible risks to the person who is involved must all be made clear. Both when seeking user consent and also in your privacy statement, you can include this data.
- You ought to keep open guidelines and procedures for the handling of personal data. Keep these records accessible to the general public.
- Every concerned participant must be made aware of the gathering, use, and dissemination of their personal information and made available to it. Additionally, they must have the opportunity to verify the precision and thoroughness of their information. Lastly, they ought to be allowed to contest the PIPEDA compliance of your company and present their argument to your CPO.

6.2 CPPA: Consumer Privacy Protection Act

The CPPA's standards are applicable to any company that:

- gathers, utilizes, and disseminates personal data from Canadians for business gain.
- gathers, utilizes, and disseminates personal details about workers and job applicants

CPPA is not applicable to:

- agencies of the state protected by the Privacy Rule
- Use of private data in journalism, the arts, and literature
- Personal data used for private objectives
- Individuals' personal data utilized in connection with work, company, or occupation

Rules in CCPA

1)Control and accountability - Individual data that is gathered, utilized, or released by you or another individual acting on your behalf is solely the responsibility of your organization under the CPPA. Additionally, it requires you to designate a single individual to be responsible for adhering to security responsibilities and to provide their personal information when necessary, such as in your privacy statement or in response to a user's request.

2) Consent- For the purpose of acquiring, using, and revealing users' personal details, you must get valid consent. Similar to how you must phrase your request in understandable terms so that users are aware of their possibilities.

Unless the organization determines that it is permissible to rely on a user's implicit consent, taking into consideration the user's reasonable expectations and the delicacy of the personal details that will be acquired, utilized, or published, consent should be explicitly acquired.

3)The ability to transfer and delete data - In addition to the PIPEDA-guaranteed rights of access and alteration to personal details and the ability to dispute conformity, people will also be granted the following rights:

- Data mobility - the ability to move personal details among institutions like banks or insurance companies.
- Personal data disposal - to ask for their data to be deleted. This holds true for all personally identifiable data that is managed by an organization.

4)Programs for privacy management and openness- Businesses will be required to set up open procedures for managing personal details in accordance with the new law. Every company should document and fully explain:

- How it will safeguard personal data
- How it will handle inquiries for data and grievances
- How it will create documentation outlining the institution's rules and regulations
- What staff training and data will be provided?

5)Keeping consent records -Your company must maintain track of consent and the reasons for data collection, usage, and disclosure. You must seek additional consent, document it, and add it to those documents if you choose to use the information for a new purpose.

This information should be kept in an easily accessible format. You will need to give the privacy commissioner access to your files in the event that data security officials conduct an audit.

6)Working with anonymized data - CPPA now makes two key data-related concepts clearer when working with anonymized or de-identified data.

In order to guarantee that no one can be directly or indirectly recognized from the data, anonymized data refers to information that has been completely and irrevocably changed in accordance with widely established best practices. We also discover that this law does not apply to anonymized data.

De-identified personal data refers to information that has been altered so that it can't be used to directly identify a specific user, yet there is still a chance that it could be. De-identified personal details are nonetheless regarded as personal information and are therefore covered by the law.

7)Protection of minor's personal information - Personal information on children is regarded by law as sensitive. Children may object to their parents' or guardians' use of their rights (including consent) on their behalf. Children also have broader rights to request the deletion of their private information.

Last but not least, the CPPA grants users the ability to bring private action against businesses that misuse personal data in a way that is against the law.

6.3 Ethical Dimensions:

Computer ethics is an analysis of the natural and social impact of computer technology and its corresponding language and justification of guidelines for the ethical use of such technology. For example, concerns about software as well as hardware, and concerns about networks that connect computers as well as the computers themselves. A typical problem with computer ethics arises because of guidelines. A vacuum about how computer technology should be used. Computers give us new skills and these open up new possibilities for action. Often there are no guidelines on how to act in these situations, or the existing guidelines seem inadequate. The central task of computer ethics is to decide. To formulate a policy to guide our actions of course, as individuals and as a society we face some ethical situations. Computer ethics involves considering both personal and societal guidelines for the ethical use of computer technology. In terms of their ability to accomplish any task defined in terms of the connection of inputs, outputs, and logical processes, computers are conceptually malleable (Moor,1985). An organized process is a logical operation. a computer transitioning between states. Hardware and software modifications have unlimited potential to manipulate and modify computer logic. The raw material for the industrial revolution was the steam engine's power, and the raw material for the computer revolution is computer logic. Since the reasoning is universal, new computer technologies in order to utilize computers ethically must consider both social and personal norms. This section deals with impediments to the ethical decision-making process, the EDM model, and Ethical decision-making in technologies with respect to context-aware privacy.

- Impediments to the Ethical decision-making process:

  First, moral awareness is tied to the outset. This is due to the possibility that we misunderstand the problem's focus to be on legal or economic issues rather than ethical repercussions. We frequently encounter moral conundrums when doing this and fail to recognize that we are not acting ethically for moral reasons. We all have cognitive and psychological biases, which is the second barrier. Each of these biases and tendencies has the potential to impair our ability to think clearly and influence immoral judgments. These

include deference to superiors, such as following a manager's directives to engage in an unethical action, prejudice conforming to what is viewed as socially acceptable behavior or starting with tiny transgressions. There is gradualism, which leads to the larger violation that comes next. Moral justification can skew moral reasoning, allowing us to commit unethical acts while maintaining moral satisfaction. can result in incorrect behavior and frequently serves as a basis for unsuitable framing, prejudice, or moral justification.

Improper framing:
Improper framing is a barrier to making ethical decisions. When we just consider the situation's economic and/or legal ramifications without considering its ethical ramifications, we are improperly framing the issue (Schwartz, 2017). We can remain in a state of lacking moral awareness, which can raise our odds of participating in the unethical activity unless we define the challenge we are facing as ethical in nature.

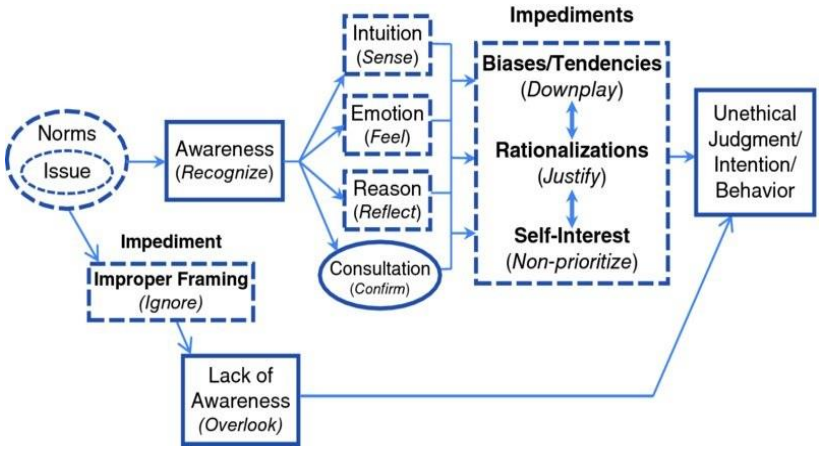

Figure 6.3.1: Barriers to making moral decisions. Source: (Schwartz, M.S. 2016).

We fail to recognize when we are in the midst of an ethical conundrum for a variety of reasons, including ethical fading, ethical blindness, and moral astigmatism. In other scenarios, we merely take certain roles as workers and managers, which can cause us to frame events without regard for ethical considerations. 2 It's not because we meant to act in an ethical or unethical way that we wind up acting in that way. Nevertheless, if we are morally alert, morally conscious, or morally creative, we raise the possibility that we will become aware that we are confronted with an ethical issue that has a number of viable solutions and that we will then engage in an appropriate moral reasoning process.

- Ethical decision-making process for organizations:
  From the perspective of computer professionals who need help dealing with ethical issues, the ACM Code of Ethics is beneficial in three of the six phases of ethical decision-making. Recognize, resolve, and decide. No information is available for the other three stages.

At the cognitive stage, norms are helpful, especially if the person is familiar with the various norms of the Code of Ethical Conduct and already has a basic knowledge of what is morally good. Some analytical thinking skills are essential during the dissolution phase. By enumerating and detailing the general, professional, and administrative responsibilities of computer professionals, ACM Codes are useful during the decision-making stage. At the same time, code alone does not help balance the various obligations that result in conflicting requirements in certain situations. At this stage of decision-making, it helps to have some understanding of the basic elements of justice.

Computer professionals learn along the moral spectrum by doing, rather than through formal training that would teach them how to apply code in various contexts. For both people and organizations, this may have negative and permanent effects. In contrast to the more prevalent use of hypothetical case scenarios, training programs designed as seminars using feature films like the one above assist individuals in classifying, identifying, analyzing, and resolving ethical dilemmas in more difficult circumstances. In this regard, the Association for Computing Machinery has discovered that two of the six stages of the ethical decision-making process - judgment and judgment - are where its code of ethics is ineffectual. It advises that sponsored or sponsored training to be used to make up for these shortcomings.

The Code's implementation and day-to-day operation are ongoing challenges for any organization. We make two recommendations. First, adding an individual ethical audit to the employee's recurring performance evaluation could help advance the application of the ACM Code. The person can next be requested to sign a personal declaration endorsing the Code. Second, when top officials do their business in a way that is above reproach, it improves the moral climate of the organization. This is true because people in higher-ranking positions are strongly impacted by the behavior of those in lower-ranking positions. Hypocrisy in the top echelons is the single biggest threat to a company's moral environment (O'Boyle, 2002).

An organization that takes ethics seriously will look for candidates who have some knowledge of ethical behavior, train them to identify it in their professional capacity, and expect that they behave ethically, especially as they are promoted to positions of more responsibility. Because of serious shortcomings in these areas, any code of ethics is unlikely to help computing professionals conduct themselves in a morally appropriate way and thereby grow more fully as human beings. Instead, it will become meaningless as a piece of paper to hang on the wall or stick in the bottom drawer.

- Ethics in technology for preserving privacy:

  The development and ethical application of technologies such as AI, Machine learning, cloud, and big data are guided by a set of moral guidelines and procedures. Organizations are beginning to create codes of ethics as technology has become ingrained in goods and

services. A policy statement that formally defines the function of intelligent machines as it relates to the advancement of the human race is known as a code of ethics, also known as a technology value platform. This code of ethics is intended to offer stakeholders direction when making moral choices involving the use of machine intelligence. Technology for extracting data, storing it, and using the data for training to provide better cost-effective services and performance. Here, our focus is on how we can implement moral values while using this technology to preserve privacy.

First comes human augmentation, where a human evaluation should be involved in every decision made. Various techniques such as web scraping, collecting data, extracting, and storing data are used to gather data from users. Human involvement is essential to decide what to be stored, what to be destroyed and what functions to be carried out. Similarly, the company should also focus on bias evaluation. Although they cannot be eliminated, major sociocultural and computational biases exist in data and can be reduced or documented. Technologists should identify the underlying bias in the data while developing the process and methods to identify the bias in the data, inference outcomes, and ramifications of the bias instead of immediately integrating ethics into algorithms.

An ethical system must be inclusive, understandable, serve a worthwhile purpose, and ethically handle data. An inclusive system is impartial and effective in all facets of society. To ensure there is no inherent bias in the data collection, this necessitates complete knowledge of each data source utilized to train the models. To remove any undesirable attributes that were picked up throughout the training process, the trained model must also undergo a comprehensive assessment. Additionally, the models must be continuously watched to make sure there is no corruption in the future.
Data privacy rights are respected by systems that use data responsibly. Data is essential to a system, and more data often yields better algorithms. The right to privacy and openness must not be compromised in the quest to amass ever more data, though. To build a trustworthy Information system, responsible data collection, administration, and use are crucial. The granularity of data should be as little as feasible, and it should only be collected, when necessary, not continuously.

A system with good intentions attempts to do things like lessen fraud, get rid of the trash, reward individuals, slow down climate change, treat disease, etc. Any technology has the potential to cause harm, so it is crucial that we consider how to protect systems from being misused. The risk of not addressing this challenge and misusing technology is much higher than it has ever been, despite the fact that this will be a difficult challenge.

7. PROPOSED MODEL IDEA

The purpose of this project is to create and build a data storage and retrieval system in cloud servers that protect user privacy. The scopes entail using encryption algorithms to look for certain keywords across encrypted content without first requiring the user to download the database and decrypt its contents. The suggested remedy delegated searching on encrypted data while maintaining privacy. We have decided to use the adaptively safe Searchable Symmetric encryption (also known as the SSE-2 scheme in the original study) method suggested by Curtmola et al. as the searchable encryption algorithm.

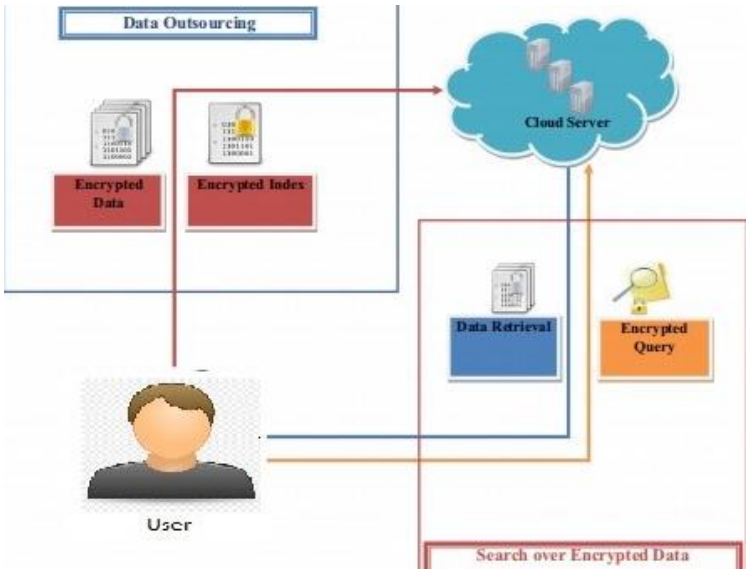

Figure 7.1: Searching through encrypted data using an adaptively secure encryption technique. The accessible symmetric encryption system's process is portrayed in the figure. To enable searching across encrypted data, our technique creates and keeps an encrypted index at the remote server.

The adaptively safe SSE technique put forth by Curtmola et al. is discussed in this section. The technique is based on the symmetric key cryptographic configuration, as the name would imply, and is therefore most appropriate for a single reader/single writer scenario. It makes use of the index-based technique (Goh,2003), in which the user pre-processes the contents to create a keyword index that facilitates search functionality. An encrypted index file "I" that comprises a collection of "m" encrypted keywords retrieved from the data set "D" is created by a user U after they encrypt a set of data D = "D1, D2,..., Dn". User U outsources index I and the encrypted data set D to the cloud server in order to run a search on the encrypted data. "U" produces an encrypted query while conducting a search and sends it to the server. Using

the encrypted index on the server, the cloud server utilizes the encrypted query as input to get links to the document(s) that contain the searched term. The client is given the encrypted document(s) containing the searched term after the search result has been achieved.
The main functional components of the technique are key synthesis, document preprocessing, encryption, search flag generation, retrieval, and decryption. The following list of each entity's and block's features:

- It uses a symmetric key primitive to produce the keys needed for encryption and decryption. For the encryption of index contents and documents, respectively, two keys, k1 and k2, are produced. At the user's device, both keys are created and safely kept. Because this primitive uses symmetric keys, the same key is utilized for both encryption and decryption.

- Reprocessing of documents: This component offers the essential functionality to prepare a group of records and launch the encryption process. The user must pre-process the document set in order to extract the keyword and create an index before encrypting it. Let "D" provide a collection of documents that will be secured before being uploaded to a remote server. The user must create an index table in D that lists all the keywords from each document along with the corresponding documents. Each document in the bundle will also receive a content ID attributable to the approach.

- The user will pre-process this collection of records during the pre-processing stage to create the index table depicted in Table 1. The term is represented by wi(i = 1, 2, 3, , n) and n is the total number of keywords. The content IDs are given out in order, starting with number 1. The procedure will then provide an inverted index table listing the corresponding content IDs for each keyword wi. Table 2 displays the intermediate representation for the aforementioned illustration. The mechanism to encrypt the index and document set is provided by this component. Using the encryption keys created during the key generation process, the encryption is carried out. It basically consists of record and index encryptions and functions as stated below: Index protection This generates an encrypted index and encrypts the phrase set generated during the pre-processing stage. The computation for the keyword encryption is ENCPK1 (wi||ni), where ENCKP1 denotes encryption with key K1, wi denotes the keyword I, and ni denote the associated document ID containing the keyword wi. The encoded index table lists the ID of the corresponding record ID for each of the encrypted keywords. The classified index table for our example document set is displayed in Table 3. Document security: Every content from record set D is then encrypted with key K2 and saved in the database. ENCPK2 (Di), which represents the encryption of document Di with key K2, is used to calculate the document's encryption. The encrypted document lists for the aforementioned scenario are displayed in Table 4.

Content/ document	Content ID	Keyword
D_1	1	W1 and w5
D_2	2	W1,w2,w8
D_3	3	W2,w3,w8
D_4	4	W6, w7,w9
D_5	5	W1,w4,w10

Table I: Record index created

- This module is utilized when a user wishes to search for a record that provides a specific keyword. It gives the user or client the ability to create a search token or sinkhole that can be used on the server to search through encrypted texts. The user will enter the search keyword and then compute the search token to complete the search operation. A keyword wq's search token is calculated as follows: The number of documents in the document set D is represented by the search token "t": ENCPK1 wq||1, ENCKP1 wq||2,, and ENCKP1 wq||n. After being calculated, the search token is transmitted to the server to identify the page that corresponds to the requested term.

Keyword	Content ID
	1, 2, 5
2	3
3	2
4	5
5	1, 3
6	4
7	4
8	2
9	3, 4
10	5

Table II: Index is inverted

- The search feature returns a list of records containing the requested phrase after receiving the search token and the encrypted index table as inputs. If any of the values in t match the encrypted keyword, it will compare the search token with the encrypted index. It sends back the relevant records to the client after outputting the matching document IDs that match the search token.

Encrypted keyword	Content ID
$ENCp_{K1}(w_1\|\|1)$	1
$ENCp_{K1}(w_1\|\|2)$	2
$ENCp_{K1}(w_1\|\|5)$	5
$ENCp_{K1}(w_2\|\|3)$	3
$ENCp_{K1}(w_3\|\|2)$	2
$ENCp_{K1}(w_4\|\|5)$	5
$ENCp_{K1}(w_5\|\|1)$	1
$ENCp_{K1}(w_5\|\|3)$	3
$ENCp_{K1}(w_6\|\|4)$	4
$ENCp_{K1}(w_7\|\|4)$	4
$ENCp_{K1}(w_8\|\|2)$	2
$ENCp_{K1}(w_9\|\|3)$	3
$ENCp_{K1}(w_9\|\|4)$	4
$ENCp_{K1}(w_{10}\|\|5)$	5

Table III: Here we can observe that the indices are encrypted

- After obtaining the encrypted document set containing the sought-after terms, the client decrypts the original document. The decryption of the document is computed as DECPK2 which decrypts the record using key K2.

Content ID	Encrypted Document
1	$ENCp_{K_2}(D_1)$
2	$ENCp_{K_2}(D_2)$
3	$ENCp_{K_2}(D_3)$
4	$ENCp_{K_2}(D_4)$
5	$ENCp_{K_2}(D_5)$

Table IV: Records list is encrypted

Let's look at the model in simpleton's words. I have implemented fundamental encryption and decryption methods using Python to store the encrypted message on the server using hashing. This facilitates the encryption of the user's personal data and the server-side storage of the encrypted data. The Administrator will be held accountable for the security and protection of the data with tight access controls and user credentials. With decrypted data, organizations can still offer the service with a minimum of information (as shown in figure 7.3). The information is encrypted when the user enters it (as shown in figure 7.2)

We can utilize the key to decode the data when necessary, using different encryption techniques. In order to reduce the danger of a data breach, this encrypted message is saved on the server.

The business has two options for user verification: OTP for online support and biometrics for banking and healthcare.

The server has the option of returning messages that are encrypted since it has decrypted the IDs and knows which have been requested. how our system is being used right now. As an alternative, the client might make a second request for the messages themselves once the server returns a list of IDs.

To update, search, or store the data, we employ tuple mapping. Let me offer an example to demonstrate this mapping:

Alice will utilize the keywords (which represent the personal information or data to be kept), and this software will generate hashes of the keywords using symmetric encryption and a secret key. Bob (a company employee) keeps this data from Alice's file on the server. If Alice wishes to get her information back, she transmits the keyword's hash. Bob maps to the server's files using the hash. Character sensitivity and unique hashes are generated for each keyword using the hashing technique. Bob will provide the corresponding files to Alice using these hash cipher keywords.

SAMPLE MODEL WORK OUTPUTS:

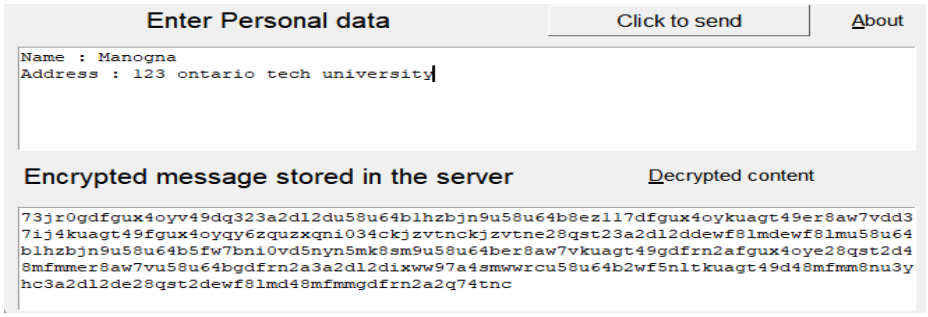

Figure 7.2: Encrypting the records of personal data using a hash function.

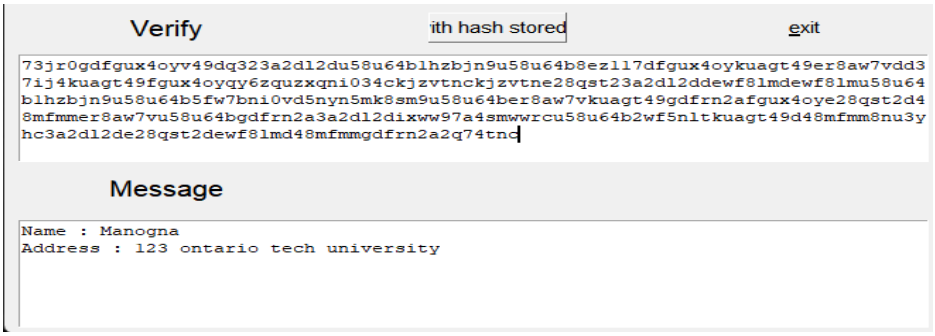

Figure 7.3: Decrypting the retrieved records from the model server

8. CONCLUSION

The increasing use of extensive, cutting-edge big data-driven research approaches challenges existing ethical frameworks and assumptions that researchers and ethics committees rely on to ensure adequate protection of human subjects. doing. As a result, there are many conceptual gaps in how big data research leverages established business model implementations and established research ethical principles. Most importantly, using decision heuristics from Nissenbaum's contextual consistency theory will help us as researchers to better understand and manage organizational ethical aspects and business model development, and bridge conceptual gaps. We have shown through model examples how it can help to fill in and ultimately secure research. Data security and accountability to win user trust and serve.

We propose a logic model for contemplating and articulating transmission norms for personal data. Contextual integrity is a theoretical structure for comprehending privacy standards that have been created in the research on law and public policy as well as concepts of ethics. This framework formalizes some of its key concepts. "Respect for privacy" in any service delivery is discussed. Ethics and statutory requirements for applications for encrypting and decrypting sensitive data using a symmetric encryption model are depicted. Achieving accountability and confidentiality lowers the danger of data breaches and improper use of data. Creating the server-side program to implement the suggested approach is the future scope. Transferring and storing data using cutting-edge encryption and hashing methods safeguards data accountability prior to access. Encrypting the data provided to the organization is still another option for the application of this concept for data collecting in relation to real-time.

"Where can we find the facts and verification?" is the question while encrypting the files. Considering locally decrypting the data, there are performance problems with this concept. So, we suggest that data be encrypted so that the server may return the encrypted data with very little data present.

With a slightly larger index size trade-off, this method indexes the entire document rather than a subset of the keywords in each document schema, allowing users to search for any keyword within the document. Also, clients are not required to maintain a keyword index on their site. However, the implementation's static index update mechanism prevents new files from being added or existing files to be modified. However, the encryption module takes a little longer, as it also involves converting the documents into text files for the keyword extraction process. A faster encryption method can be achieved if the keyword extraction module works without the document conversion process.

REFERENCES

1) Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. IEEE Security Privacy, 3(1), 24– 33.
2) 6. Adya A, Bolosky WJ, Castro M, Cermak G, Chaiken R, Doucer JR et al (2002) Farsite: federated, available, and reliable storage for an incompletely trusted environment. In: Proceedings of the 5th Symposium on Operating systems design and implementation, vol. 36, pp 1–14
3) Aggeliki Tsohou, M. K. (2015). Analyzing the role of cognitive and cultural biases in the internalization of information security policies: Recommendations for information security awareness programs. DOI: https://doi.org/10.1016/j.cose.2015.04.006.
4) Backhaushuus, L. (2012). The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI. the 2012 ACM annual conference, (pp. 367–376). New York: ACM. https://doi.org/10.1145/2207676.2207727.
5) Benaloh J, Chase M, Horvitz E, Lauter K (2009) Patient controlled encryption: ensuring privacy of electronic medical records. In: Proceedings of the 2009 ACM workshop on Cloud computing security, ACM, pp 103–114
6) Boneh D, Waters B (2007) Conjunctive, subset, and range queries on encrypted data. In: Salil Vadhan P (ed) Theory of cryptography, LNCS 4392, Springer, Berlin Heidelberg, pp 535–554
7) Boneh D, Waters B (2007) Conjunctive, subset, and range queries on encrypted data. In: Salil Vadhan P (ed) Theory of cryptography, LNCS 4392, Springer, Berlin Heidelberg, pp 535–554
8) Chang, Alvin. “The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram.” Vox, 23 Mar. 2018, www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-Cambridge-Analytica-trump-diagram.
9) Chan, Rosalie. "The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections". Business Insider. Archived from the original on January 29, 2021. Retrieved May 7, 2020.
10) Chang YC, Mitzenmacher M (2005) Privacy preserving keyword searches on remote encrypted data. In: Ioannidis J, Keromytis A, Yung M (eds) Applied Cryptography and Network Security, LNCS 3531. Springer, Berlin Heidelberg, pp 442–455
11) Curtmola R, Garay J, Kamara S, Ostrovsky R (2006) Searchable symmetric encryption: improved definitions and effi- cient constructions. In: Proceedings of the 13th ACM conference on Computer and communications security, ACM, pp 79–88
12) Fowler M. UML distilled: a brief guide to the standard object modeling language. The Addison-Wesley object technology series. Boston, MA: Addison-Wesley; 2004.
13) Fung BCM, Ke W, Rui C, Yu PS. Privacy preserving data publishing: a survey of recent developments. ACM Computer Survey 2010;42(4) 14:1–14:53.
14) Goh EJ (2003) Secure indexes. In: Cryptology ePrint Archive: Report 2003/216

15) Gillespie, Tarleton. “Governance of and by platforms”, in Burgess, Jean, Poell, Thomas and Marwick, Alice, eds., SAGE Handbook of Social Media (New York: SAGE Publishing, 2017) 1.
16) Grundstein-Amado R. (1991). An integrative model of clinical-ethical decision making. Theoretical medicine, 12(2), 157–170. https://doi.org/10.1007/BF00489796
17) Henriksen-Bulmer, J., Faily, S., & Jeary, S. (2019). Privacy risk assessment in context: A meta-model based on contextual integrity. In Computers & Security (Vol. 82)
18) Hacigümüs. H, Iyer B, Li C, Mehrotra S (2002) Executing sql over encrypted data in the database-service-provider model. In: Proceedings of SIGMOD, ACM, pp 216–227
19)
20) Hoel, T., Chen, W. & Pawlowski, J.M. Making context the central concept in privacy engineering. RPTEL 15, 21 (2020). https://doi.org/10.1186/s41039-020-00141-9
21) H. Nissenbaum. (2004). Privacy as contextual integrity. Number 1 Symposium: Technology, Values, and the Justice System. Volume 79
22) H. Nissenbaum, A. B. (2006). Privacy and contextual integrity: framework and applications. *2006 IEEE Symposium on Security and Privacy (S&P'06)*, 15 pp.-198. doi:10.1109/SP.2006.32
23) In re GOOGLE INC. COOKIE PLACEMENT CONSUMER PRIVACY LITIGATION., 13-4300. (United States Court of Appeals, Third Circuit. November 12, 2015).
24) J. Morrison, "Context integrity measurement architecture: A privacy-preserving strategy for the era of ubiquitous computing," *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2016, pp. 1-10, DOI: 10.1109/UEMCON.2016.7777816.
25) Joshi. (2020, September). Privacy Theory 101: Privacy as Contextual Integrity - Centre for Law & Policy Research. Centre for Law & Policy Research. Retrieved November 29, 2022, from https://clpr.org.in/blog/privacy-theory-101-privacy-as-contextual-integrity/
26) J.Prins. (2004). The propertization of personal data and identities. Electronic Journal of Comparative Law.
27) Kamara S, Lauter K (2010) Cryptographic cloud storage. In: Sion R, Curtmola R, Dietrich S, Kiayias A, Miret JM, Sako K, Sebé F (eds) Financial Cryptography and Data Security, LNCS 6054. Springer, Berlin, Heidelberg, pp 136–149
28) Kenny, S. & Borking J. (2002). The value of privacy engineering. The Journal of Information,, Law and Technology (JILT). Online: http://elj.warwick.ac.uk/jilt/02-1/kenny.htm
29) Krupa Y, Vercouter L. Handling privacy as contextual integrity in decentralized virtual communities: the privacy as framework. Web Intel Agent Syst 2012;10(1):105–16.
30) Kubiatowicz J, Bindel D, Chen Y, Czerwinski S, Eaton P, Geels D et al (2000) Oceanstore: an architecture for global- scale persistent storage. In: Architectural support for programming languages and operating systems, ACM, pp 190–201

31) Lahlou, S., Langheinrich, M., & Rucker, C. (2005). Privacy and trust issues with invisible computers. Communications of the ACM, 48(3).
32) Li M, Lou W, Ren K (2010) Data security and privacy in wireless body area networks. IEEE Wireless Communications Magazine, vol. 17, IEEE, pp 51–58
33) Li M, Yu S, Ren K, Lou W (2010) Securing personal health records in cloud computing: patient-centric and fine- grained data access control in multi-owner settings. In: Jajodia S, Zhou J (eds) Security and Privacy in Communica- tion Networks, LNCS 50. Springer, Berlin Heidelberg, pp 89–106
34) Liu Q, Wang G, Wu J (2009) An efficient privacy preserving keyword search scheme in cloud computing. In: Interna- tional Conference on Computational Science and Engineering (CSE), Vol. 2, pp 715–720
35) Liu Q, Wang G, Wu J (2012) Secure and privacy preserving keyword searching for cloud storage services. J Netw Comput Appl (JNCA) 35(3):927–933
36) Li M, Lou W, Ren K (2010) Data security and privacy in wireless body area networks. IEEE Wireless Communications Magazine, vol. 17, IEEE, pp 51–58
37) Lomas, N. (2020, December 10). *France fines Google $120M and Amazon $42M for dropping tracking cookies without consent*. TechCrunch. Retrieved November 29, 2022, from https://techcrunch.com/2020/12/10/france-fines-google-120m-and-amazon-42m-for-dropping-tracking-cookies-without-consent/
38) Lyons, T., Courcelas, L., & Timsit, K. (2018). Blockchain and the GDPR. Report produced by ConsensSys AG on behalf of the European Union Blockchain Observatory and Forum.
39) Mansour, R. F. (2016). Understanding how big data leads to social networking vulnerability. Computers in Human Behavior, 57(C), 348–351. https://doi.org/10.1016/j.chb.2015.12.055.
40) Moor, James H. "What Is Computer Ethics?", 16:4 (1985), Metaphilosophy 266-275.
41) O'Boyle, Edward. (2002). An Ethical Decision-making Process for Computing Professionals. Ethics and Information Technology. 4. 267-277. 10.1023/A:1021320617495.
42) Rumbaugh J, Jacobson I, Booch G. Unified modeling language reference manual. (2nd ed). Pearson Higher Education; 2004.
43) Shaffer, G. (2021, December 1). Applying a Contextual Integrity Framework to Privacy Policies for Smart Technologies. *Journal of Information Policy*, *11*, 222–265. https://doi.org/10.5325/jinfopoli.11.2021.0222
44) Shvartzshnaider, Y., Apthorpe, N., Feamster, N., & Nissenbaum, H. (2018). Analyzing privacy policies using contextual integrity annotations. Online: https://arxiv.org/pdf/1809.02236. Accessed: 2019-12-11.

45) Squicciarini, A.C., Xu, H., & Zhang, X. (2011). CoPE: Enabling collaborative privacy management in online social networks. Journal of the American Society for Information Science and Technology, 62(3), 521– 534.
46) Schwartz, Mark S. Business Ethics: An Ethical Decision-Making Approach. John Wiley & Sons Inc. (e-book). (2017). Chapter 2 & 3
    a. Online: https://ebookcentral-proquest-com.uproxy.library.dc-uoit.ca/lib/oculuoit-books/reader.action?docID=4812597&ppg=101.
47) Thomson D, Bzdel L, Golden-Biddle K, Reay T, Estabrooks CA.Central questions of anonymization: a case study of secondary use of qualitative data. Forum: Qual Soc Res 2005;6(1):1–16.
48) Tseng FK, Chen RJ, Lin BS (2013) iPEKS: Fast and secure cloud data retrieval from the public-key encryption with keyword search. International Conference on Trust. Security and Privacy in Computing and Communications, IEEE, pp 452–458
49) Wu, Philip & Vitak, Jessica & Zimmer, Michael. (2019). A Contextual Approach to Information Privacy Research. Journal of the Association for Information Science and Technology. 71. 10.1002/asi.24232.
50) Wu, P.F. (2019). The privacy paradox in the context of online social networking: A self-identity perspective. Journal of the Association for Information Science and Technology, 70(3), 207– 217.