Personal Identifiable Information Erasure in Large Language Models in a Globalized Context: a Review from the Theoretical Perspective of Ulrich Beck's Risk and Reflexive Modernity

Yuli Paola Vargas Rodríguez

Monograph Submitted as partial Fulfillment of the Requirement for the Degree in Sociology

Advisor

Jhon Freddy Duitama Muñoz, Ph.D. in Computer Science

Universidad de Antioquia

Facultad de Ciencias Sociales y Humanas

Sociología

Medellín, Antioquia, Colombia

2024

Citation	(Vargas Rodríguez, 2024)
Reference **APA 7 style (2020)**	Vargas Rodríguez, Y., (2024). *Personal Identifiable Information Erasure in Large Language Models in a Globalized Context: a Review from the Theoretical Perspective of Ulrich Beck's Risk and Reflexive Modernity*. [Professional degree work]. Universidad de Antioquia, Medellín, Colombia.

Acknowledgments

Without the network of exchange and cooperation, stablished through Jhon Fredy Duitama, Konstantin Beznosov, Jean Paul Sarrazín and Aleksandr Volosiuk, as mentors and advisors in different moments of this project, and the scholarship awarded by Mitacs, the National Research Organization of Canada, the present manuscript would not have been possible.

In addition, I want to thank Alejandro Rosso, for his adorable support, the highly stimulating conversations, and all the abundant leisure hours and not-so-abundant productive hours together.

I offer my sincere gratitude and I hope the plenitude surrounds them wherever they go.

Contents

List of tables

List of figures

Abbreviations, Acronyms and Initialisms

AI	Artificial Intelligence
LLM	Large Language Model
ML	Machine Learning
PII	personally identifiable information
STS	Science and Technology Studies
EU	European Union

Abstract

This review presents the trajectory and current scenario regarding the collection and processing of digital Personally Identifiable Information and its influence on the development of Artificial Intelligence applications such as Large Language Models. From this context, the state of existing legislative frameworks in the European Union, the State of California in the United States, and Colombian legislation in relation to the handling of personal data in technologies that use Artificial Intelligence is evaluated. At the same time, the frameworks for data processing by technology companies that have developed the most popular LLMs for the period between 2023 and 2024 are reviewed.

This review aims to analyze, from the perspective of Ulrich Beck's risk society and reflexive modernization, how the uncertainties arising from the side effects produced by technological developments with global reach, such as LLMs, are presented.

Keywords: Personal Identifiable Information, data treatment regulations, digital information, risk society, Large Language Model, Science and technology studies (STS).

Resumen

Esta revisión presenta la trayectoria y el escenario actual respecto a la recolección y el tratamiento de la Información de Identificación Personal digital y su influencia en el desarrollo de aplicaciones de Inteligencia Artificial como los modelos de lenguaje de gran tamaño. A partir de este contexto, se evalúa el estado de los marcos legislativos existentes en la Unión Europea, el Estado de California en Estados Unidos y las legislaciones colombianas en relación al manejo de datos personales en tecnologías que usan Inteligencia Artificial. Al mismo tiempo, se revisan los marcos para el tratamiento de datos de las compañías de tecnología que han desarrollado los LLMs más populares para el período comprendido entre 2023 y 2024.

Esta revisión tiene como objetivo analizar, desde la perspectiva de la sociedad de riesgo y la modernización reflexiva de Ulrich Beck, cómo se presentan las incertidumbres que surgen de los efectos colaterales producidos por los desarrollos tecnológicos con alcance global como lo son los LLM.

Palabras clave: Información de Identificación Personal, regulaciones de tratamiento de datos, información digital, sociedad de riesgo, Modelos de Lenguaje de Gran Escala, estudios de Ciencia y Tecnología (STS).

Chapter 1. Introduction

The popularity of Large Language Models (LLMs) grows in 2022 with the introduction of ChatGPT[1] by OpenAI, a model designed for conversational text generation (Gupta et al., 2023). The widespread adoption of LLMs is attributable to advancements in key elements as the development of more powerful Graphical Processing Units (GPUs) and expanded cloud computing resources, that address the substantial computational demands required for training these models. These advancements significantly enhance the efficiency and speed of model training. Additionally, the technology sector's ongoing interest in Artificial Intelligence and the exponential increase in available online data—such as web pages, digital books, news articles, and social media content—have provided extensive datasets crucial for training purposes. Typically, the training of multipurpose LLMs utilizes open-source repositories like OpenWebText2 and Commoncrawl[2], which compile textual data from billions of web pages. These datasets enable the models to learn statistical and structural linguistic patterns, including grammar, syntax, semantics, and style. However, such datasets may inadvertently include Personally Identifiable Information[3] (PII) from individuals who have disclosed identity-related information online.

This is why LLM's researchers have been testing different models using a technique known as prompt injection for identifing vulnerabilities. It has been proved that ChatGPT could be exploited by malicious users to extract PII, bypassing the model's ethical constraints (Gupta et al., 2023; Li et al., 2023; Lukas et al., 2023; Wu et al., 2024). Addressing these concerns requires navigating the legal frameworks governing data handling and the available legal actions for affected individuals, which vary depending on the jurisdiction of the individual's country of nationality and the geographical location of the developing company. This complexity is compounded by the misalignment between different regulatory environments.

The predominant data protection regulation all over the world is the General Data Protection Regulation (GDPR), established by the European Union in 2016. Additionally, California Consumer Privacy Act (CCPA) is notable due to the concentration of AI technology

[1] https://chatgpt.com/

[2] Refer to: https://openwebtext2.readthedocs.io/en/latest/ and https://commoncrawl.org/

[3] The distinction between Personal Identifiable Information and Privacy obey to the specificity in the actions referred to in the laws on data that identify an individual.

firms on this State. Colombia faces similar challenges and may need to update its legislation regarding the removal of PII from LLMs.

Beyond legal frameworks, technical considerations are also crucial when requesting the removal of PII, because of the complexity of data erasure process in Machine Learning (ML).

Models can be compared with the action of deleting experiences and memories which constitute personal traits in humans. For LLMs, modifying these traits is particularly challenging because their training and learning processes depend on the entirety of the dataset, meaning that altering specific data could disrupt the model's structure and functionality.

In this regard, national borders are less relevant in this context, as academic and independent research remains freely accessible. The feasibility of removing PII is complicated because of potential alterations to the model and implications for the technology developed by companies.

1.1 Problem statement

The German sociologist Beck (1992) introduces the concept of the "risk society" to describe a global context increasingly mediated by technology, characterized by the blurring of national borders and the intensification of transnational interdependencies. This phenomenon reflects not only the expansion of commercial and cultural networks but also signifies a shift towards a society where unintended consequences of technological progress, such as climate change, environmental pollution, pandemics, and technological hazards (e.g., nuclear disasters, cyberattacks) are prevalent. Beck (2009) terms these unintended consequences as "manufactured uncertainties," or *human-made uncertainties* which result from technological advancements and generate significant global risks.

These risks or *manufactured uncertainties* (Beck, 2009) introduces legal complexities, as regulatory frameworks in different countries strive to adapt to the rapid evolution of technology. Concurrently, corporations are continually updating their data handling policies to align with these evolving legal and technical standards. The challenge is particularly evident in the management of PII within LLMs. Regions pioneering regulatory frameworks for the inadvertent inclusion of PII in LLMs, emphasize on the implications that these technologies to pose beyond unforeseen side effects and present a fundamental challenge inherent in technological

development. This issue has mobilized modern institutions—such as the state, corporations, and academia—due to its profound impacts on individual privacy.

Beck's theoretical framework underscores the significance of these challenges, especially with the proliferation of freely accessible LLMs on the internet, which can be used globally. This situation raises concerns about the legitimacy of using data shared by users for ongoing model training, particularly when this data is deemed sensitive under the user's national legislation, which may not align with the regulations of the country where the platform-developing company is based. The treatment of PII and user requests for data modification or deletion may differ or be incompatible across jurisdictions, even though the technical methods for data handling in Machine Learning models remain consistent.

Beck's analysis calls for heightened awareness, as contemporary global risks such as climate change, financial crises, terrorism, and nuclear threats are increasingly recognizable and necessitate collective action beyond national borders. This scenario illustrates the limitations of nation-states in addressing global issues independently and highlights the need for integrated approaches to regulation and policies which consider the global 'other' among us, who cannot be excluded on those.

To address these concerns, this review proposes a comprehensive examination of the origins, evolution, and current issues related to digital data, particularly in the context of recent developments in LLMs. The study will assess the state of existing legislative frameworks and data controller policies with a specific focus on current AI advancements. This analysis will cover regulatory frameworks from the European Union, the State of California, and Colombia for the period between 2020 and 2024.

Such an examination aims to illuminate how the risk society and reflexive modernization—concepts defined by Ulrich Beck (1992)—expose the uncertainties arising from the interaction between social phenomena and globally significant events. The study will also explore how these factors, alongside the *non-knowing paradox* impact the erasure of PII in LLMs. This approach will provide insights into the challenges and regulatory considerations associated with managing PII in the era of *reflexive modernity* (Beck, 1992) as described by the author.

1.2 Objectives

General

The analysis aims to explore the interplay between Ulrich Beck's concepts of risk society, reflexive modernization, and the non-knowing paradox, with the issue of PII erasure in LLMs. This will be achieved through a comprehensive literature review from a legal perspective, examining privacy policy statements and data treatment practices of leading AI companies with popular LLMs from 2020 to 2024.

Specific objectives

1. To identify the digital data origin context, process, and its relationship with the recent developments of LLMs.

2. To describe current legislative frameworks focused on GDPR of the European Union, CCPA-California's data protection legislation in the United States and Colombian AI regulations, and data controller policy statements approached to AI for identifying the frameworks state, between 2020 and 2024.

3. To analyze the relation between risk society, reflexive modernization, and non-knowing paradox Ulrich Beck's theoretical concepts, with PII erasure in LLMs.

1.3 Methodology

1.3.1 Research internship and the laboratory's dynamic

In 2023, *Mitacs*, Canada's national research organization, provided me the opportunity to undertake a research internship at the *Laboratory for Education and Research in Secure Systems Engineering (LERSSE)* at The University of British Columbia located in Vancouver, Canada. In its initial phase, the process involved integrating me into the dynamics of the laboratory. These dynamics included group counseling sessions to exchange views on possible ways of delimiting one's chosen topic of study, bibliographic suggestions and the identification of conceptual relationships with topics that other members were simultaneously exploring. These exchanges occurred during weekly group meetings, individual consultations with the doctoral candidate, and weekly meetings with the laboratory director, who monitored progress. There, members presented their findings on the chosen topics to exchange experiences and give feedback. Thus, the dynamics served as a basis for me to select the research topic of interest, based on three focal areas addressed by the Lab: usability, privacy and security, in terms of their relevance to science and technology studies (STS), and impacts on populations in contemporary society. After evaluating the options, I chose "Privacy in LLMs", and started with the review of the state of the art and the systematization of the information found in this regard.

1.3.2 Systematization of key elements in documentary sources

The process of systematically organizing the documentary sources involved several steps, each of which was essential to gain a thorough understanding of the chosen topic and the need to delimit it. Initially, the selection of documentary sources was guided by criteria such as thematic specificity and year of publication, which made it possible to identify the state of current issues and the approaches through which IPI-related problems have been investigated in LLM. On the one hand, I identified legislative and regulatory concerns, and on the other hand, the tensions between the agents involved in the development of LLMs, their use and the place that governments are assuming. One of the recurring themes identified during the systematization was the technical process for the elimination of PII, its applicability and challenges to be solved.

In this sense, the review gradually led to the thematic delimitation that would later be analyzed from a sociological framework.

The time frame of the publications considered for systematic reviews ranged from February 2021 to July 2024. It was conducted by searching multiple databases and open access archives, including Scopus, Web of Science, ACM Digital Library, IEEE Xplore, Google Scholar, ArXiv.org, Proceedings, Elsevier (open access publisher), and PubMed, which redirect to conferences, journals, or seminary memories. Each paper was meticulously reviewed to code significant aspects relevant to the state of the art. This information was then used to create a diagram illustrating the relationships between the topics addressed by various authors. The bibliographic characteristics, objectives, and keywords for each study were organized into a matrix as shown in Table 1.

Table 1

Matrix for the systematization of key elements identified in consulted documents

ID	Age	Title	Conference / Journal	University / Institution	Authors	Objective	Key words
1	2023	"A Design Theory for Transparency of Information Privacy Practices".	Article accepted for publication at: Information Systems Research (Academic journal)	- Karlsruhe Institute of Technology - KASTEL Security Research Labs. (Germany)	-Tobias Dehling -Ali Sunyaev	The manuscript presents an Information Systems Design Theory (ISDT) that establishes a theoretical foundation for Information Systems (IS) that are less prone to undermine privacy as a social value by establishing transparency of information privacy practices (TIPP). In other words, TIPP theory explains and prescribes what transparency artifacts should be built to reveal the information consumers need to interact with IS in line with their privacy expectations.	-Transparency -Information privacy -Information privacy practices -Consumer information systems -Design -relevant explanatory/predictive theory -Information systems design theory -Theory development
2	2023	"When AI	IEEE Access	Institute of	-Abdul	AI also has strong synergy	-AI-powered

		Meets Information Privacy: The Adversarial Role of AI in Data Sharing Scenario"	(Volume: 11)	Information and Communications Technology Planning and Evaluation (IITP)	Majeed - Seong Oun Hwang	with the information privacy (IP) one of them is as a threat tool (i.e., AI used to compromise an individual's privacy), with a special focus on synthetic data (SD) that can serve as background knowledge leading to various kinds of privacy breaches. For instance, SD can encompass pertinent information (e.g., total # of attributes in data, distributions of sensitive information, category values of each attribute, minor and major values of some attributes, etc.) about real data that can offer a helpful hint to the adversary regarding the composition of anonymized data, that can subsequently lead to uncovering the identity or private information. We perform reasonable experiments on a real-life benchmark dataset to prove the pitfalls of AI in the data publishing scenario (when a database is either fully or partially released to public domains for conducting analytics).	attacks -Artificial intelligence -Background knowledge -Compromising privacy -Data publishing -Personal data -Privacy -Safeguarding privacy -Synthetic data -Utility
3	2023	"How Is Privacy Behavior Formulated? A Review of	Multimodal Technologies and Interaction. (Journal)	Multidisciplinary Digital Publishing Institute (MDPI)	-Ioannis Paspatis - Aggeliki Tsohou -	We seek to investigate the underlying factors that influence individuals' privacy-conscious behavior in the digital domain, as well as	-Privacy behavior -Determinant factors

		Current Research and Synthesis of Information Privacy Behavioral Factors"		(Switzerland)	Spyros Kokolakis	effective interventions to promote such behavior. Privacy decisions regarding the disclosure of personal information may have negative consequences on individuals' lives, such as becoming a victim of identity theft, impersonation, etc. Moreover, third parties may exploit this information for their own benefit, such as targeted advertising practices. By identifying the factors that may affect Social Networking Service (SNS) users' privacy awareness, we can assist in creating methods for effective privacy protection and/or user-centered design. Examining the results of several research studies, we found evidence that privacy behavior is affected by a variety of factors, including individual ones (e.g., demographics) and contextual ones (e.g., financial exchanges).	

The bibliographic review and systematization of the relevant content for the selected topic was carried out by each of the members of the laboratory according to the chosen theme. During the weekly meetings, the main presentation covered the objectives of the selected publications, the databases consulted, the main contributions of the document in relation to the topic of interest, the results, and the conclusions. These reviews allow me to identify thematic categories of analysis reflecting the current landscape, regulatory frameworks, company data processing

policies and the technical field, after analyzing what the authors of the papers wrote or presented in the literature.

1.3.3 Identifying the relationships between concepts

During the process of identifying common conceptual categories within the reviewed documents, I achieved significant conceptual clarifications that will enhance the clarity of subsequent documentary searches in the forthcoming months of the internship. Furthermore, the insights gained from the literature review on the handling of personal information leaks in the context of Large Language Models (LLMs) have facilitated an ongoing exploration of the interrelationships between digital data, privacy, secondary uses, and regulation. This exploration also highlighted prevalent concepts in the literature. Consequently, the diagrams utilized in group meetings proved to be a valuable tool for elucidating these relationships, and for identifying potential key concepts, organizational approaches, governmental perspectives, company policies, and advancements in technical procedures by researchers

Figure 1

Diagram of relations between digital data, privacy, secondary usages, and regulations.

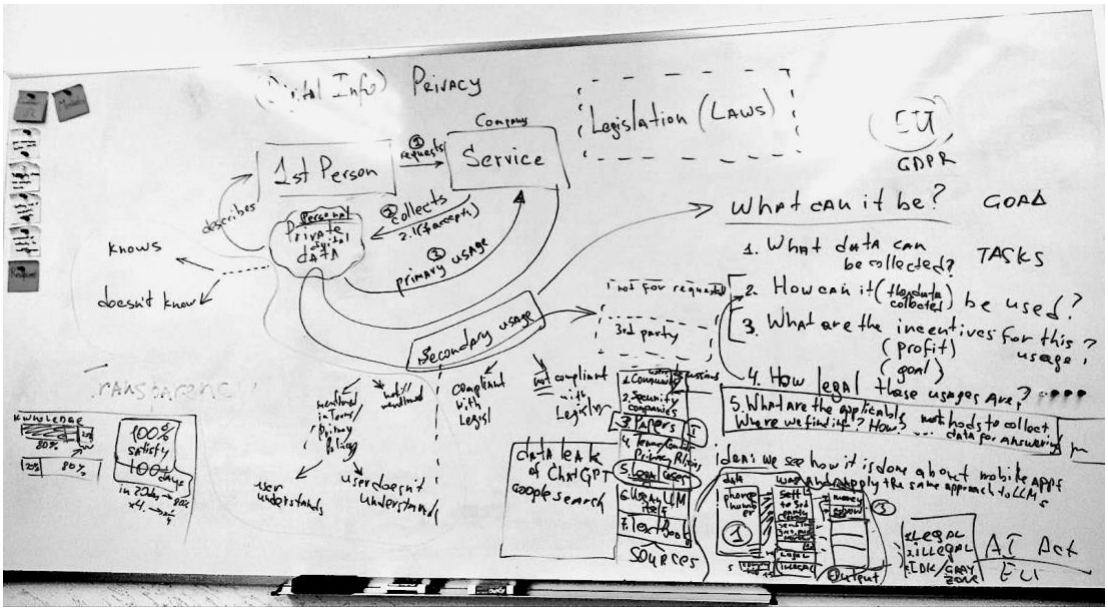
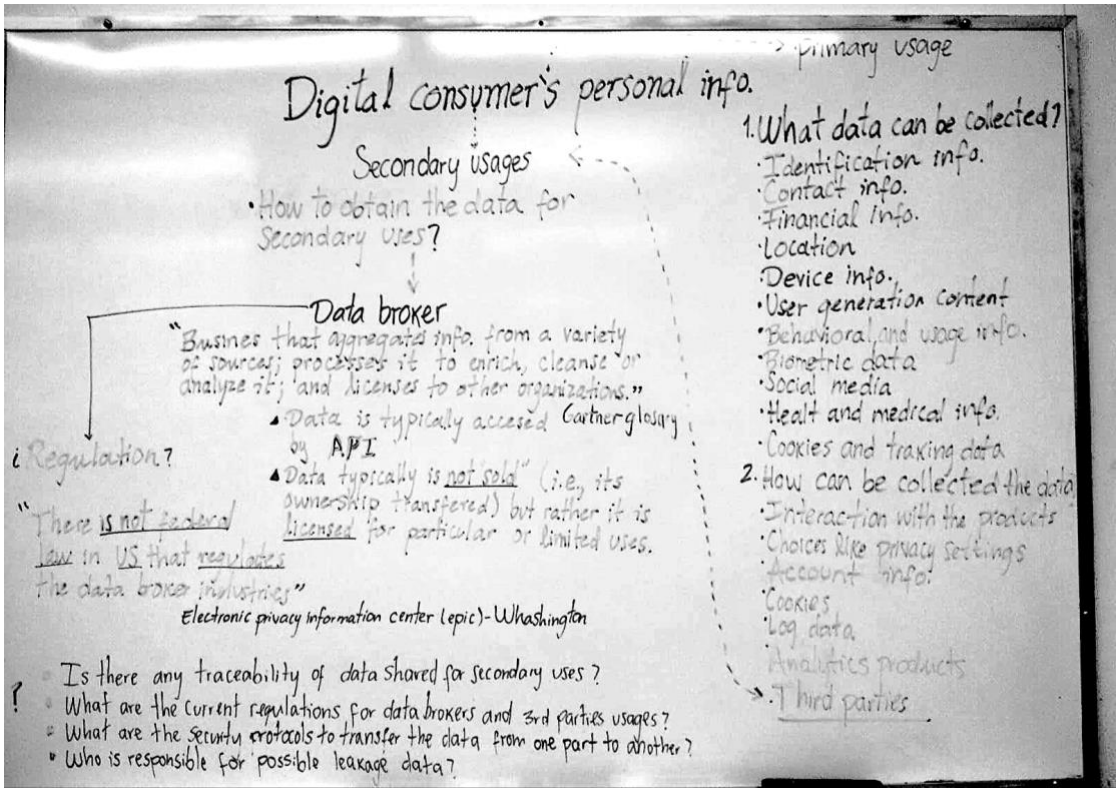


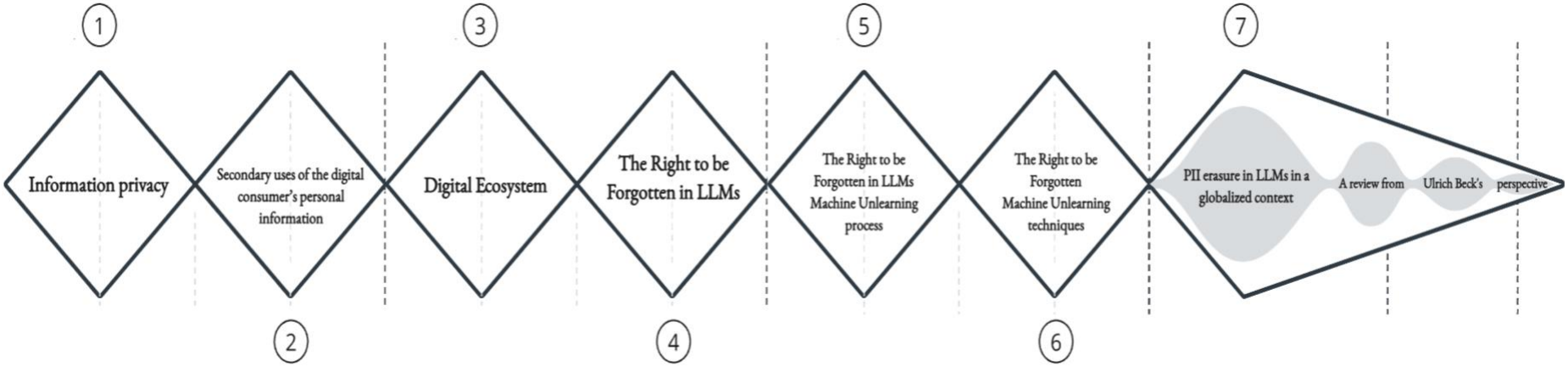
Figure 2

Relation between digital data secondary usages, data brokers and data's type collected.



The described process allowed me to delimitate the scope, the topic and to formulate a research problem based on the analysis provided not only by the literature reviewed but also by the laboratory member's significant feedback regarding the key point around PII in a digital context I would identify in relation with STS.

Figure 3
Topic delimitation process



The figures would also be a resource for further analysis from sociological theory, making evident elements related to risk, non-knowing paradox and reflexive modernization.

1.3.4 Purposed documentary analysis

The extraction of key citations and examples illustrating perceptions of risk and responses to data leakage facilitates the classification and coding of documents according to themes related to Beck's risk society theory. Data are categorized to reflect aspects of risk concerning digital data mining, non-knowing paradox in relation to regulations on PII, and globally applied technical methods. Initially, Chapter 2 provides a description of digital data mining, while Chapter 3 presents an overview of global regulations related to AI. This is followed by an analysis based on Beck's theoretical framework in Chapter 4.

The analysis elucidates how the documents reflect the global and interconnected nature of risks associated with information breaches in LLMs, examining the global repercussions of such breaches. It explores how these risks are presented in terms of unpredictability and uncertainty and evaluates individual responsibility in protecting personal information. This includes an assessment of how policies and regulations, both proposed and existing, address the issue of data breaches and whether they are adequate for the risk society context. Finally, the analysis identifies points of convergence and divergence regarding potential consensus on handling these issues and notes whether the documentation reflects varied responses.

Chapter 2. An overview on digital data mining

The incursion of the Internet most widespread global information system, known as the World Wide Web, the arrival of software and web services as commercial products, and the subsequent development of search engines made building a complex documentation system for information exchange possible (Berners-Lee, 1999).The Web, launched in 1989 at the European Organization for Nuclear Research (CERN), is the work of Tim Berners-Lee, who was interested in developing a resource that would solve the problem of linking unconnected information and make it accessible to people in any geographical location (Berners-Lee, 1999). Before WWW was launched there were not optimal and global resources for sharing massive amounts of data and files with different formats and extensions, even though the Internet, as a global system of interconnected computer networks, was available for sharing data from the 70's.

What provides WWW, is the possibility of access to Web site contents by using a web browser installed in a personal computer and connected to the internet. The important fact is not just the communication established between the server, that allows the website, and the personal computer but the decoded information in Hypertext Markup Language (HTML), a Markup language with embedded content, linked to another different content via hyperlinks (Berners-Lee, 1999), which in turn is linked to another content and so on, by using the Hypertext Transfer Protocol (HTTP), an stateless file transfer protocol which makes a request for a file named by the web address or Uniform Resource Locator (URL), and gets the file in response to finally disconnect the communication established by a client and the server which is allowed the web page visited. Basically, this is why HTTP is considered stateless, once the transference between the server and the browser which makes the request is done, the server will not attempt to keep the connection alive (Nikiforakis et al., 2013). Thus, with the model for accessing multiple websites and in turn the possibility of sharing data, no matter the physical location of each user, the digitalization of basically all life aspects becomes part of a society. Over time, it is possible to make quantitative analysis of digital data by training machine learning models which can result in social effects in different levels (Emmert-Streib, 2021).

2.1 Earliest digital information collection on internet context

With the increasing popularity of WWW, different browser tools were implemented for improving the user experience on the internet[4]; as a product of commercial necessities or as part of business strategy methods of technology companies. In its origin, the principal impeller of online tracking was behavioral advertising, the personalization of ads based on search history as a technical practice which may include customized price discrimination and steering (Hannak et al., 2014), (Mattioli, 2012) and personalized search results (Xing et al., 2014).

Web marketers started collecting information about their customers via the Internet in two main ways: browser and device fingerprinting, with the time were developed tools as web scraping for collecting information around internet.

2.1.1 Browser fingerprint

There are two main categories for browser fingerprinting, active and passive. In the first one, the user is explicitly consulted for preferences and options to define the level of information that is willing to allow for recording. For the second one, web marketers use data collection tools in a way that requires no explicit actions on the user's part, in which are included cookies and HTTP logging.

The *cookie*[5] is a tool developed to solve the problem posed by the stateless feature of the HTTP protocol, which involves not retaining the information of a user who has previously visited the requested website. This scenario appears in the business need of the Netscape company when developing an e-commerce application with a virtual shopping cart that must remember the customer's choices (Thomas, 1997). According to Kesan & Shah (2003), without a stateful mechanism, such as cookies, buying goods would be like using a vending machine and it would not be possible to buy more than one product in a transaction. At the time cookies successfully solved a technical challenge and even the first Netscape's web browser included them by default. However, according to (Kesan & Shah, 2003), there were no notification

[4] See "Kant, T. (2021). A History of the Data Tracked User" for a timeline with relevant events regarding data users tracking.

[5] A cookie is just a code such as a reference number or account number that the server assigns to the browser to recognize it when the same person returns. (Berners-Lee, 2000. p. 145)

mechanisms to inform users when cookies were being placed on their computers, in addition to other privacy concerns that lead to the conclusion that "cookie technology was not transparent." (p. 300) The *cookies* combined with referrer information[6] make it possible to track a person's movement by the visited websites, which means to monitor and to store user's movements around the web. (Kesan & Shah, 2003)

Another common tool included in web browsers is *HTTP logging*, a middleware[7] that store log files on a company's web server, with the record of every transaction between a web client and server, such as IP address, country code, domain name, date and hour of the request, HTTP method of the request, number of bytes transferred, type of browser used, among others[8]. (Matsuda et al., 1998)

In 2010, Peter Eckersley examine how much modern web browsers can be fingerprinted by using the version and configuration data they send to websites upon request. He concluded that browser fingerprinting is a potent tactic and that, when discussing web privacy and user trackability, fingerprints should be considered in addition to cookies, IP addresses, and supercookies[9]/evercookies, whereupon technical design as well as privacy policies are affected.

Gunes Acar et al., in their 2014 study mention different forms of mitigation to defend against tracking such as "blocking third party cookies" and making sure to "delete all possible storage locations." (p. 682) Also Eckersley (2010), highlight the Tor project[10] and its remarkable efforts for design and development against fingerprintability, similarly, Acar et al., (2014) refer the design and implementation of the Tor browser features regarding privacy. The list of items removed by that browser, includes HTTP auth, cookies, content and image cache, offline and memory cache, offline storage, SSL state, Cache storage, IndexedDB, DOM storage among others. (The Tor Project, 2019)

[6] The referenced [sic] field is part of the HTTP protocol advocated by Berners-Lee in 1992. It provides a website with the previous URL visited by the person. Its intended purpose was to allow websites to detect web sites that had linked to them with the hope that they would then add links back to the referring sites. (Kesan, J. P., & Shah, R. C., 2003. p. 299)

[7] According to Rick-Anderson (2023), HTTP logging is a middleware that records details about incoming requests and responses to the HTTP protocol.

[8] For details about Potential Marketing Application refer to "HTTP logging" section of *Data Collection: Defining the Customer* in Matsuda, Y., Rosenstein, P., Scovitch, C., & Takamura, K. (1998). Direct Marketing on the Internet.". *Massachusetts Institute of Technology*.

[9] The supercookie is a persistent cookie that is used by websites to replicate and store cookies that users have explicitly removed or changed.

[10] https://www.torproject.org/

2.1.2 Device fingerprint

The collection of information about the hardware and software of a remote computing device has been knowing as device fingerprinting. The data set, associated with a user's identity, can be stored even when cookies are restricted from being stored in the web browser or the browser is changed on the same device (Cao et al., 2017). To estimate the time deviation of a device to remotely obtain the device's hardware fingerprint, Kohno et al., (2005) identifies three main classes of remote physical device fingerprinting techniques: 1) passive fingerprinting, which consists of observing the device traffic, 2) active fingerprinting, in which the fingerprinter must have the ability to initiate connections to the fingerprintee, and 3) semi-passive fingerprinting, in which once a connection initiated by the fingerprinter is initiated, the fingerprinter can interact with the fingerprintee.

2.1.3 Web scraping

To understand when web scraping is useful, Richard Lawson (2015) presents a couple of illustrative examples:

What if you want to buy shoes on sale but you don’t know the specific date? or, imagine you are a store’s owner who needs to be aware of competitors product prices? How to obtain the information avoiding a repetitive manual task of daily check websites information update?

Richard Lawson (2015) propose is, instead of repetitive search, use an automated solution implementing web scraping techniques to replace the manual option.

To do it is relevant to have in account the web scraping process which is divided basically into three stages: 1) Fetching: the website is accessed via HTTP protocol by sending a GET request to the URL (web address) to get the HTML page as a response, 2) Extraction: there are different approaches as using Regular expressions to identify patterns in a string and scans for any matches on it; XPath queries for detecting information in documents and HTML parsing libraries to identify the relevant data according to the target and, 3) Transformation: due to the unstructured nature of the data collected, those must be converted to a structured format for storage or analysis for business intelligence purposes (Khder, 2021).

In the process, automated *bot* programs are developed to scrape specific information, which made the extraction faster. However, currently, some websites provide application programming interfaces (APIs), to share the hosted information, on the website, in a structured format. An API is "a set of protocols that enable different software components to communicate and transfer data" (Postman, Inc., 2024), in other words, it is an interface that allow the communication between client and server through a request and response cycle but the end user no necessarily must be a person, it could be a programmer or a set of instructions scheduled to *call* an specific part of API's content. "Indeed, some websites do provide APIs, but they are typically restricted by what data is available and how frequently it can be accessed". (Lawson, 2015, p.1) this is why developers are supposed to learn still web scraping techniques.

2.2 Large-scale collection of digital data

Since the early 1990s, the amount of data published on websites has grown exponentially, mostly due to user-generated content from "regular people who voluntarily contribute data, information, or media that then appears before others in a useful or entertaining way, usually on the Web—for example, restaurant ratings, wikis, and videos". (Krumm et al., 2008, p. 10) The potential of be informed based on real data from people[11] that has already visited a restaurant, touristic place, or store, make it attractive and consistent for contributors and followers to continue generating and consuming the information in a massive scale. Also, the public-mass network is constantly developing new technical approaches, strategies, and tools for gathering vast amounts of data in variated file format, the text is the most extended quantity of data on internet until now.

Based on licensing policies, required fees or special access to those, the datasets are divided into two categories: non-open data and open data. Rob Kitchin, (2014), argues that even if the datasets are available they have required specialist equipment and tools, such as computers and software, skills such as statistics and mapping know-how, and contextual knowledge concerning a field or topic, to make sense of them, much of which is beyond the capabilities of the general population.

[11] Krumm et al., 2008, identified four example categories of pervasive user-generated content and experiences to go into other parts of our lives. As an example, the OpenStreetMap project, for creating free road maps based on the work of volunteer's data contributions to trace roads on satellite images or out of-copyright maps. (p. 10)

This is why civil society organizations and other agents as researchers promote the *open data movement*, not just to open the access, but giving easy-to-use research tools to population without specific skills and statistical background. The value of data for society can be realized through collaboration, sharing, and transparency. In contrast to restricting the power of data to its producers and those able to pay for access, its goal is to democratize the ability to produce knowledge and information. (Kitchin, 2014) Using real data, such as web scraping COVID-19 news items, to generate datasets for sentiment and emotion analysis, is an excellent example of data science applications and an overview of large-scale datasets potential use (Khder, 2021).

2.3 Current large-scale data sets applications

The field of developing computers and robots that can simulate human thought and perform tasks in real-world settings is known as artificial intelligence (AI). AI has expanded its application beyond the development of high-performance systems for specific uses to become a general-purpose science. AI is the ability of computers to emulate human thought and perform tasks in real-world environments (Dhar, 2023). This new paradigm is a new way of thinking that better fits the current anomalies and challenges in the contemporary society. The impact on the change in methods for problem solving, then, requires its reading as a scientific paradigm, according to Dhar (2023) interpretation of Thomas Kuhn (1997), lecture of scientific revolutions in the human history. In 80´decade one of the first approaches of AI gained popularity due to its fiability: the Expert Systems. Accurately process task-specific instructions based on human interactions from which experience is derived. The main foci of this paradigm were diagnosis, design, and planning, and it was thought that they may be regarded as intelligent given their degree of competence, which was comparable to that of humans. It was difficult to express uncertainty in terms of the symbolic representations of knowledge that were then in use, such as "IF/THEN" rules, "semantic networks," or "structured object representations." This led to the use of probability theory models to represent uncertainty in knowledge (Dhar, 2023 citing Pearl, 1988). Probability and statistical data management were fundamental for AI’s approach known as Machine Learning, which aim to develop technologies and algorithms, as a key system for finding patterns, make decisions, and improve themselves through experience and data for becoming most reliable to application directed to identify patterns of natural language

comprehension, as an example. The diversity and quantity of the available datasets have facilitated the development of more precise and reliable AI systems that can manage difficult tasks and generate exact probability estimates based on trained models.

2.3.1 Large Language Models

Natural Language Processing (NLP) is an interdisciplinary field within computer science that focuses on enabling computers to support and manipulate human language. To achieve this, NLP processes natural or "ordinary" language datasets, such as text or speech corpora, using machine learning techniques based on neural networks and statistics. A computer that can comprehend the context of papers is the end aim. In the twenty-first century, the quantity of data available on the internet has increased, pushing NLP academics to study similarities in linguistic meaning or syntax using deep learning or artificial neural network methods. This approach involves systems learning a language through exposure to large quantities of text and using those text to generate its own prediction challenges. Examples of such challenges include identifying each word in the text given the previous words in phrase or a text sequentially. By repeating these prediction tasks billions of times and learning from its mistakes, the model becomes more proficient the next time it encounters a similar textual context (Manning, 2022). The technical advances in neural networks were implemented also in Large Language Models (LLMs), whose main feature, is the ability to generate language for general purpose, i.e., applicable in variated human contexts. Along with the syntax, semantics, and world knowledge that these sources inherently convey, they are thought to be able to detect biases and mistakes present in corpora of human language. The most widely used LLMs include Meta's LLaMA, xAI's Grok, Google's PaLM, BERT, Microsoft Copilot, Chat-GPT, and Gemini. Large volumes of data are needed for LLM training in order to fully capture and generalize the many nuances of human language. Because of this, researchers frequently mix data from many sources, including books, StackExchange, Github, Commoncrawl, C4 (Colossal Clean Crawled Corpus), Wikipedia, Red Pajama, and web text (Shen et al., 2023). Concatenation of multiple datasets by merging various corpora, covering diverse topics, styles, and sources, is one of the most often utilized methodologies for data combination for LLMs. Additionally, web literature that has been carefully selected by scraping content from the internet offers a varied range of training materials

because it is a rich combination of formal, informal, factual, and opinionated writing. (Gao et al., 2020)

According to Kumar et al., (2024) LLMs have significantly advanced tasks such as text categorization, named entity recognition, sentiment analysis, and language translation, enhancing the precision and efficiency of NLP applications. Are also crucial in computational linguistics due to their ability to comprehend and generate text akin to human language, serving as foundational components in conversational AI and chatbots by enabling more natural and contextually relevant interactions in domains like customer service and virtual assistants, thereby enhancing user experiences through tailored responses. Additionally, LLMs excel in information retrieval and question answering tasks, facilitating efficient access to relevant information across applications such as search engines and recommendation systems. They also play a pivotal role in software development by assisting in code generation, debugging, and programming assistance, ultimately streamlining development processes and enhancing productivity. Moreover, LLM-powered Personalization and Adaptive Systems leverage user input to deliver personalized user experiences and content recommendations, proving invaluable in content platforms and e-commerce applications where customized interactions foster user satisfaction.

Nevertheless, these models are susceptible to security and privacy threats, including data poisoning, jailbreaking, and PII leakage attacks, even if they provide a lot of benefits (Das et al., 2024).

Chapter 3. Government Regulations and Companies' PII Management Practices related to LLMs

"We've updated our Privacy Policy below. These updates do not apply to individuals located in the European Economic Area, UK, and Switzerland. If you reside in those areas, this version of our Privacy Policy applies to you."

OpenAI Privacy policy statement

Regulating Artificial Intelligence (AI) and particularly PII in LLMs is essential for several reasons that encompass both ethical and practical concerns. Across various national legislations, there is a shared interest in ensuring privacy and the secure and confidential handling of personal data to protect individuals from unauthorized access or data breaches. Given that LLMs may contain highly sensitive information, safeguarding this data is crucial to prevent leaks and other security risks.

AI regulations also promote transparency regarding how data is collected, processed, and stored. This enables individuals to exert greater control over their PII, allowing them to decide how and when it is used. Regulations establish accountability mechanisms for developers and users, ensuring that automated decisions are auditable, and processes are transparent.

The implementation of clear and robust standards also aids companies in competing in a global market by adhering to international standards, facilitating trade and international cooperation in the development and deployment of AI. Laws such as GDPR in Europe and the CCPA mandate that organizations obtain explicit consent from users for processing their personal data, specify how that data will be retained, and ensure its protection. These regulations aim to prevent the misuse of personal data and ensure that companies act responsibly.

Privacy regulations guarantee rights such as the "right to be forgotten," which allows individuals to request the deletion of their personal data. They also ensure data portability, enabling users to transfer their data between services. Given that LLMs handle large volumes of data, they must comply with these rights to respect individual autonomy.

3.1 Tensions between States and AI companies

The collection, storage, and use of large amounts of data in AI models by their processors and controllers[12] raise significant concerns for country governments and technology companies. Concerns include ethical and social implications related to privacy, fairness, accountability, and data transparency. By which governments aim to regulate how data is collected, stored, processed and shared to protect the rights of individuals and prevent its misuse, but one of the most complex concerns is in relation to regulations for the deployment and development of LLMs as there are regulatory gaps and inconsistencies in AI governance, as well as a lack of regulatory standards across countries through which to address the cross-border challenges of companies to ensure the safe and responsible development and deployment of technologies such as LLMs.

Regulations such as GDPR in Europe and similar laws elsewhere require explicit consent for data processing, specify data retention periods, and mandate security measures to safeguard sensitive personal data (General Data Protection Regulation, 2024). Another complex issue is defining data ownership and control, especially in contexts where multiple parties are involved in its collection and processing. Regulations can define rights related to data ownership, intellectual property and access to data to ensure fair practices and prevent exploitation. Also, data localization is important to mention, as some governments require certain types of data to be stored within their jurisdiction to ensure regulatory compliance, facilitate law enforcement access, or protect national security (Chen, 2021). This can pose challenges for global AI companies operating across borders because data transfers facilitated in these scenarios must maintain data protection and privacy.

3.1.1 Legislative projects approached to LLMs

The global consulting firm Berkeley Research Group, LLC (BRG) is a company that offer to his customers advisory in key areas such as economics, disputes, and investigations; corporate finance; and performance improvement and advisory. In the Global AI Regulation Report of 2024, they consider that the AI Act of the European Union, passed by the European

[12] General Data Protection Regulation. Chapter 4. Controller and processor definition. Available on https://gdpr-info.eu/chapter-4/

Parliament in March 2024, contains the most extensive regulations around the world. With an interventionist and risk-based approach, the AI Act forbids the use of AI in contexts such as sorting job applications or for detecting emotions in workplaces and schools. It also places limitations on the use of GenAI tools, and recent reports indicate that the United Kingdom is beginning to draft regulations centered on the potent language models that underpin ChatGPT. (Berkeley Research Group, 2024)

Fairly's Regulation and Policy Tracker[13] is also one of the initiatives which captures in a map the current state of AI and AI-adjacent regulation around the world. On it are showed the regulations into categories such as general regulations or focuses on generative AI, data, internet, privacy, and advertising or government and Military uses of AI. Also, regulation in the healthcare sector for both public and private parties; finance and Insurance industries.

The current quantity of General regulations on AI are 8 in effect, 6 passed, 41 proposed and 64 policies and for Data, Internet, Privacy, and Advertising there are 53 in effect, 13 passed, 20 proposed and 4 policies, those are the categories with the highest and more advanced regulations. (Fairly AI, 2024)

It is relevant to point out two data privacy frameworks and compliance by its relevance on defining other regulations around the world. The first one is the European Union's General Data Protection Regulation (EU GDPR) implemented in 2018, which stands as a landmark regulation governing data protection and privacy for individuals within the European Union and the European Economic Area. The second one is the United States comprehensive federal data privacy law, specifically the CCPA that focuses on specific industries or types of data, leading to a more fragmented regulatory landscape and the EU GDPR adopts a comprehensive and rights-based approach, emphasizing individual rights to privacy, data portability, and the "right to be forgotten”. (Bakare et al., 2024)

[13] See https://github.com/fairlyAI/fairly-regulation-policy-tracker

3.1.2 Colombian case regulations

According with the Forum on Administration, Management, and Public Policy Forum GPP[14], dedicated to public administration, public management, and public policy current research, in Colombia until now there are 4 legislative bills specifically concerning artificial intelligence that have been filed or withdrawn, and 4 active ones (Foro Administración, Gestión y Política Pública, 2022). Additionally, there are 2 active decrees and 1 active regulatory bill that address issues related to artificial intelligence systems, summarized as follows:

Table 2
Colombian Bills Related to the Use of AI-Based Tools and Personal Data Processing

Category	Bill Name	Development	Status	Notes
Archived or Withdrawn Bills	Bill No. 021 / 2020 Chamber	Withdrawn by the authors before being approved in the first debate.	Withdrawn	
	Bill No. 354 / 2021 Chamber	Not approved in the first debate during the legislature, thus archived.	Archived	
	Bill No. 253 / 2022 Senate	The report for the first debate was published but it was not approved during the legislature.	Archived	
	Bill No. 200 / 2023 Chamber	Filed on September 6, 2023. Positive report published on November 10, 2023. Comments from SIC, MinTIC, and other organizations. Associated with two public hearings organized by Representative Alirio Uribe. Not approved in the first debate in the 2023-2024 Legislature, hence archived.	Archived	Citizen interventions by the research group from the Universidad de Los Andes. (Muñoz et al., n.d.)
	Statutory Bill 111 / 2022 Senate – 418 of 2023 Chamber	Reconciled in Chamber and Senate, now under automatic constitutional review by the Constitutional Court since October 2023.	Not Specified	Relevant article on data analytics and AI in the National Registry of Civil

[14] Forum GPP) is an independent venue for engagement among scholars, professionals, public officials, and students from Schools of Government in Latin America, Europe, the United States, Oceania, and other research organizations such as observatories and government research institutes.

				Status.
	Bill No. 156 / 2023 Chamber	Published in Gazette No. 1188 of 2023. Awaiting scheduling for the first debate.	Pending	Relevant articles on personal data protection and use of AI.
Active Bills in Congress	Bill No. 059 / 2023 Senate	Filed on August 1, 2023. Report for the first debate published in Gazette No. 1354/2023. Approved in the first debate on November 14, 2023. Report for the second debate published in Gazette 233 / 2024.	Approved in First Debate	Identical to Bill No. 253 / 2022 Senate. Technical concept available.
	Bill No. 091 / 2023 Senate	Filed on August 9, 2023. Explanatory statement published in Gazette 1068/2023. Approved in the first debate.	Approved in First Debate	
	Bill No. 130 / 2023 Senate	Filed on September 6, 2023. Published in Gazette 1228/2023. SIC concept suggests changes. Approved in the first debate. Awaiting report for second debate.	Approved in First Debate	
	Bill No. 225 / 2024 Senate	Filed on February 21, 2024. Report for the first debate published in Gazette 463/2023. Approved in the first debate. Report for the second debate published in Gazette 111/2024.	Approved in First Debate	
	Bill No. 447 / 2024 Chamber	Approved in the first debate on June 3, 2024. Relevant articles on emerging technologies and data interoperability.	Approved in First Debate	Relevant articles on compliance with digital government policies and data utilization.
Public Hearings in Congress	Public Hearing on AI and Human Rights	Constitutional First Commission of the House of Representatives, December 1, 2022.	-	Organized by Representative Alirio Uribe.
	Public Hearing on AI regarding Bill No. 200 / 2023 Chamber	Constitutional First Commission of the House of Representatives, December 1, 2023.	-	Organized by Representative Alirio Uribe.
	Technical	First Commission of the Senate of the	-	Organized by Senator

	Meeting on AI Implications	Republic, April 5, 2024.		David Luna.
Current Regulations	Decree 1078 of 2015	Defines artificial intelligence software in Title 16, classification of software for digital content.	In Force	
	Decree 403 of 2020	Principles for the use of AI in surveillance and fiscal control.	In Force	

Several legislative projects concerning artificial intelligence (AI) in Colombia have either been archived or withdrawn. Notable among these are Bill No. 021/2020 from the Chamber of Representatives and Bill No. 354/2021, both of which aimed to establish regulatory guidelines for AI but were either withdrawn or archived without progressing further. Additionally, Bill No. 253/2022 from the Senate, which sought similar regulatory objectives, was archived after its report for the first debate was published but did not advance beyond this stage in the legislative process.

Currently, several AI-related bills are active. Bill No. 059/2023 from the Senate, which provides guidelines for AI development and use, was approved in the first debate and is progressing to the second debate, mirroring the previously archived Bill No. 253/2022. Bill No. 091/2023, also from the Senate, focuses on the responsible use of AI and has been approved in its initial debate. Bills No. 200/2023, No. 130/2023, and No. 225/2024 are in various stages of development, addressing issues from AI regulation to labor rights and amendments to the Penal Code. Moreover, Statutory Bill 111/2022 and Bill 156/2023 focus on data protection and the regulation of AI technologies, with ongoing constitutional review and pending debates. Existing regulations include Decree 1078 of 2015 and Decree 403 of 2020, which define AI software and integrate it into fiscal control measures, respectively.

3.2 Current Privacy policies and data treatment frameworks of AI companies with popular LLMs

AI companies with global extended LLMs must implement robust privacy policies and data treatment frameworks to ensure regulatory compliance and protect user trust. Adherence to data protection laws like the GDPR and CCPA is essential to avoid legal penalties and ensure lawful data handling practices. These policies also foster transparency by clearly outlining how user data is collected, used, and safeguarded, which is crucial for maintaining user trust and engagement with AI systems.

Additionally, privacy policies and data frameworks are vital for ensuring data security, ethical handling of information, and operational efficiency. They help mitigate risks associated with data breaches and misuse by establishing comprehensive security measures and protocols. Furthermore, these frameworks support consumer rights by providing mechanisms for users to manage their data and adapt to evolving regulatory standards. Strong privacy practices not only protect against legal and reputational risks but also offer a competitive advantage in a data-conscious market.

Table 3

Main features of privacy policies and data treatment frameworks of LLMs companies

Company	Privacy Policy Highlights	Data Treatment Framework	Key Features
OpenAI (GPT-4)	- **Data Collection:** OpenAI collects data from user interactions to improve service functionality and enhance model performance. - **Anonymization**: Personal data is anonymized and aggregated to protect individual identities. - **Data Usage:** Data is primarily used for model training and development, with strict limitations on its application beyond these purposes.	- **Data Storage:** Data is stored securely with limited retention periods based on its usage. - **Research Use:** Data may be used in internal and external research to advance AI technology. - **Third-Party Sharing:** Data may be shared with partners and vendors under stringent conditions to ensure it is used appropriately. - **Data Access Controls:** Only authorized personnel have access to user data, with strict protocols	- **Transparency Reports:** OpenAI publishes regular transparency reports detailing data usage, privacy practices, and any breaches. - **Access Controls:** Implement robust access control measures to ensure data security. - **Opt-Out Mechanisms:** Users have options to opt out of certain data collection practices, particularly for non-essential uses.

	- **User Consent:** Users are informed about data collection practices and must agree to them. (*Privacy Policy*, n.d.)	in place to prevent unauthorized access.	
Google DeepMind (Gemini)	- **Consent:** Data collection is based on explicit user consent, with clear explanations provided about data handling practices. - **Data Aggregation:** Data collected is anonymized and aggregated to analyze trends without compromising individual privacy. - **Data Limitation:** Data is used specifically for improving AI models and performance, with limitations on its application for other purposes. - **Compliance:** Adheres to stringent data protection regulations, including GDPR. (*AI Safety Summit*, 2024)	- **Training and Improvement:** Data is utilized to train and improve AI models, with detailed protocols for its handling and storage. - **Retention Policies:** Data retention periods are defined and enforced, with mechanisms for secure data deletion. - **Data Sharing:** Sharing with third parties is controlled and regulated to ensure compliance with privacy policies and legal requirements. - **Regulatory Compliance:** Ensures adherence to GDPR and other international data protection laws through regular audits and assessments.	‘ **Privacy Audits:** Regular independent privacy audits to ensure compliance with policies and regulations. - **User Consent Management:** Detailed processes for obtaining and managing user consent. - **Disclosure and Transparency:** Comprehensive disclosures regarding data collection, usage, and any third-party sharing practices.
Microsoft (Azure OpenAI Service - GPT-4)	- **Functionality:** Data is collected to support and enhance the functionality of products and services. - **Privacy by Design:** Privacy considerations are integrated into product design and development processes. - **Anonymization:** Data is anonymized to prevent identification of individuals. - **User Consent:** Users are required to consent to data collection practices, with clear	- **Model Training:** Data is used to train and refine AI models, with specific protocols for handling and usage. - **Data Sharing:** Data may be shared with partners under strict agreements to maintain privacy and compliance. - **Retention Policies:** Defined policies for data retention and deletion ensure that data is not kept longer than necessary. - **Compliance:** Adheres to GDPR and other relevant regulations,	- **Privacy Notices:** Clear and detailed privacy notices explain data collection and usage practices. - **Data Protection Measures:** Implemented robust data protection measures, including encryption and secure storage. - **Regular Updates:** Frequent updates to privacy policies to reflect changes in data handling practices and regulatory requirements.

	information provided. (gluckd, 2024)	with regular updates to privacy practices.	
Anthropic (Claude 3)	- **User Privacy:** Emphasizes strong protection of user data and privacy. - **Minimal Data Collection:** Collects only minimal data necessary for AI improvement. - **Anonymization:** Data is anonymized to ensure privacy. - **Transparency:** Provides clear information about data practices and uses. (*Privacy Policy Anthropics*, n.d.)	- **Model Improvement:** Data is used to enhance and develop AI models, with strict protocols in place for its use. - **Data Retention:** Data retention policies are implemented to limit the duration of data storage. - **Third-Party Sharing:** Limited sharing of data with third parties, with strong contractual and privacy protections. - **Data Security:** Employs advanced security measures to protect data from unauthorized access and breaches.	- **Privacy by Design:** Integrates privacy protections from the outset of model development. - **User Control:** Provides users with control over their data and how it is used. - **Transparency Reports:** Regularly publishes transparency reports detailing data handling and privacy practices.
Meta (LLaMA)	- **Personalization:** Uses data to personalize user experiences and improve service delivery. - **Consent:** Requires user consent for data collection and usage. - **Anonymization:** Implements anonymization techniques to protect user identities. - **Data Security:** Employs robust security measures to protect data from unauthorized access. (*Meta Privacy Policy - How Meta Collects and Uses User Data*, n.d.)	- **AI Enhancement:** Data is used to enhance AI algorithms and functionalities, with well-defined protocols for data usage. - **Global Compliance:** Ensures compliance with global data protection laws, including GDPR and CCPA. - **Data Sharing:** Data sharing with partners is regulated by strict privacy agreements and practices. - **Retention Policies:** Data retention policies are in place, with secure methods for data deletion.	- **Privacy Settings:** Extensive privacy settings available for users to manage their data and preferences. - **Policy Updates:** Regular updates to privacy policies to reflect changes in data handling and regulatory requirements. - **Data Protection Practices:** Implemented comprehensive data protection practices, including encryption and secure storage.
IBM (Watson)	- **Service Improvement:** Data collected to improve services and AI capabilities. - **Anonymization:** Personal data is	- **Training and Analytics:** Data is used for training AI models and conducting analytics, with strict protocols for its handling. -	- **Governance Policies:** Strong data governance policies are implemented to ensure data security and privacy.

	anonymized to protect user privacy. - **Compliance:** Adheres to industry standards and regulations for data protection. - **User Consent:** Users are informed and must consent to data collection practices. (*IBM Data Privacy Framework Policy for Certified IBM Cloud Services*, n.d.)	**Retention and Deletion:** Data retention and deletion policies ensure data is not kept longer than necessary. - **Standards Compliance:** Complies with industry standards for data protection and privacy. - **Data Sharing:** Controlled data sharing with partners, governed by privacy agreements.	- **Privacy Assessments:** Regular privacy assessments to evaluate and improve data protection practices. - **User Rights Management:** Management of user rights regarding data access, correction, and deletion.

Chapter 4. Analysis

The integration of artificial intelligence (AI)-based tools into products or services for communication, information retrieval, or access to everyday goods has highlighted legal gaps that challenge the frameworks established by international agreements. This suggests the need to assess potential risks arising from such implementations. Thus, the proliferation of AI technology in international trade presents a broad range of legal challenges, including issues related to liability, data protection, intellectual property rights, and regulatory frameworks. Frequently, AI is not specifically addressed in current international trade agreements, leading to uncertainty and inconsistencies in legal interpretation. The scenario becomes more complex as the pace of technological development outstrips the ability of legal systems to regulate it.

This situation results in tensions between data handling policies proposed by technology companies and the regulatory frameworks of nation-states, not to mention the technical challenges involved in applying the processes specified in these frameworks. As outlined in Chapter 3, the digitization of personally identifiable information (PII) has spurred regulatory efforts by governments and companies to address legal debates surrounding data processing. These actions underscore the need for frameworks that facilitate the management of PII in machine learning (ML) models by users, while also considering the ethical and practical concerns raised by all stakeholders involved in the issue.

4.1 Reflexive modernization and risk society

To interpret the challenges posed by AI-based technologies and the use of regulatory frameworks concerning public interest issues such as the extent of personally identifiable information (PII) processing by companies and their product users, Ulrich Beck’s concept of the "risk society" (1986) is utilized. In Beck’s framework, social, political, economic, and individual dangers tend to escape the control and protective institutions typical of industrial society. In this context, the systematic production of effects and self-threats is not subject to public discussion or political disputes, as the self-concept of industrial society remains dominant. This situation exacerbates and legitimizes the threats generated by decision-making, labeling them as residual risks (Beck & Ritter, 1992). Global and systemic risks generated by technological advancements

and the interdependencies that support, for example, websites on the World Wide Web, coupled with the ineffective attention to individuals' claims regarding the right to manage and monitor their PII due to legal gaps, underscore the inadequacy of traditional governance structures in managing complex risks. This situation necessitates new decentralized approaches that emphasize transparency and collective accountability. The return of uncertainty to society implies that an increasing number of social conflicts are regarded as risk problems rather than being resolved as issues of order. Risk problems lack clear solutions; instead, they are characterized by fundamental ambiguity, which may be expressed through probabilistic calculations without eliminating their existence (Beck et al., 1994).

In Beck's theory, reflexive modernization represents the transition from industrial society to the risk society, where the focus shifts to self-confrontation and adaptation to new forms of social risks (Beck et al., 1994), this is exemplified by the rapidly growing technologies based on AI and the need for not only each government and company to confront their own actions but also for self-confrontation as members of the collective institutions responsible for regulating all participants involved in the creation and consumption of these technologies.

4.2 Global Complexities of Regulations for AI Applications

Despite the broad scope of international trade agreements, there are notable gaps and ambiguities in addressing the specific legal challenges posed by products and services incorporating AI-based tools into their processes. A significant gap is the lack of explicit provisions that directly address these tools, as most existing trade agreements were negotiated prior to the widespread adoption of AI (Igbinenikaro & Adewusi, 2024). The uncertainty and interpretation issues regarding the application of current norms to AI-driven exchanges stem from this lack of specific provisions. This gap is indicative of the risk society described by Beck, where traditional legal and regulatory systems are ill-equipped to confront the emerging and global risks presented by technologies such as AI.

In 2021, Rong Chen presented “Mapping Data Governance Legal Frameworks Around the World,” a study that evaluates regulatory frameworks for data management and examines not only the presence and strength of laws related to data protection and e-commerce but also the management of data governance. Chen identified that, although many countries have

implemented data regulations, these are neither sufficiently extensive nor aligned with best regulatory practices to create a conducive environment for the data economy. This finding reflects the inequality in risk management described by (Beck & Ritter, 1992), where disparities between countries impact their ability to effectively manage global risks. Particularly, low-income countries exhibit deficiencies in technical and institutional infrastructure, whereas high-income countries score better in administrative aspects of data governance, though their laws may still fall short of all best regulatory practices. The study reveals a strong correlation between regulatory quality and public trust in the data economy, suggesting that a robust regulatory framework for data governance is often linked to countries where citizens have a higher perception of participation and trust in government. Although direct causal relationships cannot be established, the data suggest that data governance tends to be more effective in contexts with high levels of public trust in the overall regulatory environment (Chen, 2021). To strengthen public trust, transparency and open access to the processes involved in generating machine learning (ML) models, particularly large language models (LLMs), are required. This includes knowledge not only of the algorithms applied but also of the flows and agents participating in data transaction markets, data pricing, and privacy calculations.

Xu et al., (2023), outlines common techniques for data pricing and privacy calculation issues, based on three typical data transaction scenarios drawn from previous research: 1. A single data owner and multiple data buyers, 2. Multiple data owners and a single data buyer, and 3. Multiple data owners and multiple data buyers. The techniques employed to preserve data integrity vary depending on each scenario; for example, in scenario 3, data brokers are often involved to regulate compensation among all participants. Thus, data is a critical asset (Igbinenikaro & Adewusi, 2024) and its increasing use raises significant concerns regarding the privacy and protection of personally identifiable information (PII), particularly because the disclosure of the processes followed is not a priority for the actors involved in the data management ecosystem.

4.3 The Unknown and Ambiguity of Risks Associated with Data Processing in ML Models

According to Beck (2009), the formal definition of risk extends beyond mere uncertainty and encompasses two often-overlooked states of knowledge: ambiguity and the unknown. Ambiguity, as defined by the author in the context of risk assessment, manifests in the difficulty of reaching a consensus on how to frame and interpret results, rather than merely evaluating their probabilities. This difficulty impacts the formulation of questions, consideration of various perspectives, and the application of methodologies, as seen in the regulatory challenges faced by AI-based technologies and the diverse interpretations of international scenarios related to PII.

On one hand, there is a lack of understanding about the most effective methods to achieve consensus on regulation; on the other hand, there is the ignorance of users regarding the handling of their data. Based on the Australian Community Attitudes to Privacy Survey 2020 - Office of the Australian Information Commissioner, Quach reveals that 58% of the population admitted they do not understand what firms do with the data they collect, and 49% feel unable to protect their data due to a lack of knowledge or time (Quach et al., 2022). This ignorance also refers to the inability to question what is problematic to the public, not due to a lack of evidence of the effects of certain developments, but rather in the sociological sense of "knowledge construction as expectation" described by Beck, regarding the implications of the unknown in human living conditions vis-à-vis the knowledge of expert systems and control, as well as notions of sovereignty and state authority.

In this context, Beck (2009) revisits the cover article of Time magazine titled "Living with Risk" of 2003, which explores how individuals in advanced countries confront diffuse risks and scientific uncertainty. Although scientists can calculate probabilities regarding risks associated with genetically modified foods, mobile phones, and chemicals, these calculations do not guarantee whether the risks are real, nor do they provide clear guidance for consumers to make informed decisions. Tim Berners-Lee had already highlighted this risk in relation to the use of cookies on early websites on the Internet, stating:

> The problem is not with the cookie, over which the user has control. The problem is that it is not known what information the server will collect and how it will be used. Without

> this information, users may make choices based solely on fear and doubt; this is not a stable foundation for building a society on the web. (Berners-Lee, 1999, p. 145)

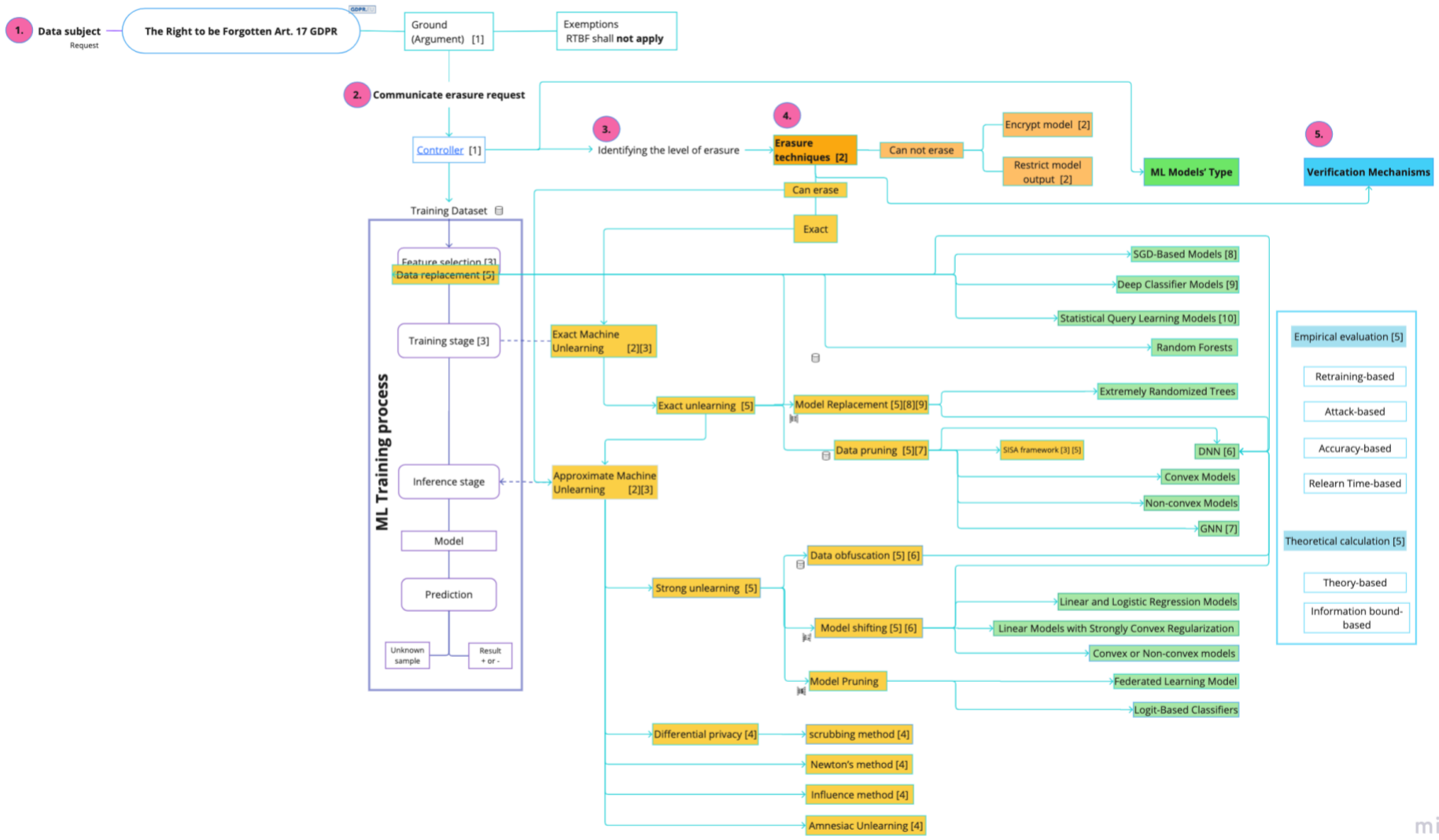
Although, for the case of PII, the GDPR describes that the data subject[15] GDPR stipulates that data subjects have the rights to access, rectification, restriction of processing, data portability, the right to object, the right not to be subject to solely automated decision-making, and the right to erasure commonly known as the "Right to Be Forgotten" *(General Data Protection Regulation, 2024)*, the processes and flows following data subjects' requests are not disclosed or made public regarding how data is processed once a request is approved.

The following diagram illustrates the phases of the request process for data subjects, citizens of the European Union, exercising the Right to Erasure.

[15] GDPR defines the Data Subject as the individual the personal data relates to.

Figure 4

Data treatment process in the application of The Right to be Forgotten, under GDPR requesting



The diagram begins with the data subject's request and proceeds with communicating the deletion request to the data controller or to the AI company that developed the model. Once the level of deletion is identified, the appropriate deletion technique is chosen based on the type of machine learning (ML) model. Finally, the success of the process is verified through mechanisms such as empirical evaluation or theoretical calculation. This schema arises from the need to understand how information is deleted in a large language model (LLM) once the request is accepted by the data controller. However, this understanding required extensive academic literature review and continuous evaluation of the correct interpretation of its content. I emphasize this aspect particularly because acquiring knowledge of the processes results from the lack of such information from companies that ultimately perform the technical process of total or partial data deletion in ML models like LLMs. This situation exemplifies the lack of transparency regarding data processing and highlights the role of researchers and open-access academic sources, as well as the importance of global knowledge networks that operate parallel to industries.

5 Conclusions

The social structuring of knowledge and ignorance, reflecting a hierarchy of power between experts and non-experts, affects how risks, such as the side effects of technologies or industrial products, are perceived and managed. The capacity to understand and manage these risks is often concentrated in groups with power and resources, while other actors are excluded from this specialized knowledge. In the context of LLMs, this manifests as the opacity of algorithms. Technical details and internal functioning of these models are often reserved for experts, leaving ordinary users without a clear understanding of how their data, including personally identifiable information (PII), is processed, and managed. The complexity and lack of transparency in these algorithms contribute to a limited understanding of data usage, generating a dynamic like the "not-knowing" described by Ulrich Beck.

The concept of side effects highlights that a lack of knowledge about potential negative impacts can intensify risk rather than mitigate it. In the case of LLMs, the absence of complete understanding regarding their operation and PII management can lead to exacerbated risks, such as exposure of sensitive data or information manipulation. Understanding and disclosing these risks is essential for properly managing emerging technologies and protecting users from potential dangers, underscoring the need for greater transparency and access to information within the field of artificial intelligence.

Although regulatory frameworks in Colombia are nascent, efforts to establish guidelines regarding the handling of PII in AI technologies are significant, especially as ministries, civil society representatives, and other governmental agents are participating in the global discussion about the impacts of this technological development.

References

Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014). The web never forgets: Persistent tracking mechanisms in the wild. *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 674–689.

Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. *Computer Science & IT Research Journal*, *5*(3), 528–543.

Beck, U. (2009). *World at risk*. Polity Press.

Beck, U., Giddens, A., & Lash, S. (1994). *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Stanford University Press.

Beck, U., & Ritter, M. (1992). *Risk society: Towards a new modernity*. Sage Publications. http://catdir.loc.gov/catdir/enhancements/fy0656/92050272-t.html

Berkeley Research Group. (2024). *What Would Good AI Governance Look Like? | BRG Global AI Regulation Report Insights BRG*. https://www.thinkbrg.com/insights/publications/airegulation/

Berners-Lee, T. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper San Francisco.

Cao, Y., Li, S., & Wijmans, E. (2017). (Cross-) browser fingerprinting via OS and hardware level features. *Proceedings 2017 Network and Distributed System Security Symposium*.

Chen, R. (2021). *Mapping Data Governance Legal Frameworks around the World: Findings from the Global Data Regulation Diagnostic*. The World Bank. https://doi.org/10.1596/1813-9450-9615

Das, B. C., Amini, M. H., & Wu, Y. (2024). Security and privacy challenges of large language models: A survey. *arXiv preprint arXiv:2402.00888*.

Dhar, V. (2023). The Paradigm Shifts in Artificial Intelligence. *arXiv preprint arXiv:2308.02558*.

Eckersley, P. (2010). How unique is your web browser? *Privacy Enhancing Technologies: 10th International Symposium, PETS 2010, Berlin, Germany, July 21-23, 2010. Proceedings 10*, 1–18.

Emmert-Streib, F. (2021). From the digital data revolution toward a digital society: Pervasiveness of artificial intelligence. *Machine Learning and Knowledge Extraction*, *3*(1), 284–298.

Fairly AI. (2024). *Global AI Regulation Tracker*. https://www.fairly.ai/blog/map-of-global-ai-regulations

Foro Administración, Gestión y Política Pública. (2022). *Regulación sobre IA*. https://forogpp.com/inteligencia-artificial/regulacion-sobre-ia/

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., & Nabeshima, N. (2020). The pile: An 800GB dataset of diverse text for language modeling. arXiv. *arXiv preprint arXiv:2101.00027*.

General Data Protection Regulation. (2024a). *Art. 17 GDPR – Right to erasure ('right to be forgotten')—General Data Protection Regulation (GDPR)*. https://gdpr-info.eu/art-17-gdpr/

General Data Protection Regulation. (2024b). Recital 51—Protecting Sensitive Personal Data. *General Data Protection Regulation (GDPR)*. https://gdpr-info.eu/recitals/no-51/

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*. https://ieeexplore.ieee.org/abstract/document/10198233/

Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014). Measuring price discrimination and steering on e-commerce web sites. *Proceedings of the 2014 conference on internet measurement conference*, 305–318.

Igbinenikaro, E., & Adewusi, A. O. (2024). Navigating the legal complexities of artificial intelligence in global trade agreements. *International Journal of Applied Research in Social Sciences*, *6*(4), 488–505.

Kesan, J. P., & Shah, R. C. (2003). Deconstructing code. *Yale JL & Tech.*, *6*, 277.

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, *13*(3).

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Kohno, T., Broido, A., & Claffy, K. C. (2005). Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, *2*(2), 93–108.

Krumm, J., Davies, N., & Narayanaswami, C. (2008). User-generated content. *IEEE Pervasive Computing*, *7*(4), 10–11.

Kuhn, T. S. (1997). *The structure of scientific revolutions* (Vol. 962). University of Chicago press Chicago.

Kumar, A., Murthy, S. V., Singh, S., & Ragupathy, S. (2024). *The Ethics of Interaction: Mitigating Security Threats in LLMs* (arXiv:2401.12273). arXiv. http://arxiv.org/abs/2401.12273

Lawson, R. (2015). *Web scraping with Python*. Packt Publishing Ltd.

Li, H., Guo, D., Fan, W., Xu, M., Huang, J., Meng, F., & Song, Y. (2023). *Multi-step Jailbreaking Privacy Attacks on ChatGPT* (arXiv:2304.05197). arXiv. http://arxiv.org/abs/2304.05197

Lukas, N., Salem, A., Sim, R., Tople, S., Wutschitz, L., & Zanella-Béguelin, S. (2023). Analyzing leakage of personally identifiable information in language models. *2023 IEEE Symposium on Security and Privacy (SP)*, 346–363.

Manning, C. D. (2022). Human language understanding & reasoning. *Daedalus*, *151*(2), 127–138.

Matsuda, Y., Rosenstein, P., Scovitch, C., & Takamura, K. (1998). Direct Marketing on the Internet.". *Massachusetts Institute of Technology*.

Mattioli, D. (2012). On Orbitz, Mac users steered to pricier hotels. *Wall Street Journal*, *23*, 2012.

Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., & Vigna, G. (2013). Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. *2013 IEEE Symposium on Security and Privacy*, 541–555.

OpenAI. (2024). *Privacy policy*. https://openai.com/policies/row-privacy-policy/

Postman, Inc. (2024). *What is an API? A Beginner's Guide to APIs*. Postman API Platform. https://www.postman.com/what-is-an-api/

Quach, S., Thaichon, P., Martin, K. D., Weaven, S., & Palmatier, R. W. (2022). Digital technologies: Tensions in privacy and data. *Journal of the Academy of Marketing Science*, *50*(6), 1299–1323. https://doi.org/10.1007/s11747-022-00845-y

Shen, Z., Tao, T., Ma, L., Neiswanger, W., Hestness, J., Vassilieva, N., Soboleva, D., & Xing, E. (2023). Slimpajama-dc: Understanding data combinations for llm training. *arXiv preprint arXiv:2309.10818*.

The Tor Project. (2019). *The Design and Implementation of the Tor Browser*. https://2019.www.torproject.org/projects/torbrowser/design/

Thomas, B. (1997). Recipe for E-commerce. *IEEE Internet Computing*, *1*(6), 72–74.

Wu, X., Duan, R., & Ni, J. (2024). Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence*, *2*(2), 102–115.

Xing, X., Meng, W., Doozan, D., Feamster, N., Lee, W., & Snoeren, A. C. (2014). Exposing inconsistent web search results with bobble. *Passive and Active Measurement: 15th International Conference, PAM 2014, Los Angeles, CA, USA, March 10-11, 2014, Proceedings 15*, 131–140.

Xu, J., Hong, N., Xu, Z., Zhao, Z., Wu, C., Kuang, K., Wang, J., Zhu, M., Zhou, J., & Ren, K. (2023). Data-driven learning for data rights, data pricing, and privacy computing. *Engineering*, *25*, 66–76.