THE UNIVERSITY OF CHICAGO

The Right to Be Forgotten in the Age of Artificial Intelligence:

Meta's Implementation Challenges

Submitted to the Pozen Center for Human Rights

The University of Chicago

In Partial Fulfillment of the Requirements

For the Degree of Bachelor of Arts

By: Harper Schwab

Faculty Advisor: Lake Polan

Chicago, Illinois

April 2025

CONTENTS

I. Introduction

The digital age has fundamentally transformed our relationship with personal information, creating unprecedented challenges for privacy protection frameworks worldwide. The right to be forgotten, also known as data erasure rights, is a principle which states that an individual has the right to request their personal data to be erased from internet services and search engines. The right to be forgotten often exists within larger legal frameworks of user data protections protecting the amount and content of personal data that can be utilized or stored by an organization. Frameworks protecting this right have been established within the European Union through the General Data Protection Regulation (GDPR); however, not federally in the United States. Given the interconnected and complicated nature of the internet, a request for data erasure cannot always be achieved, typically it is only legally protected to a technically feasible level. As artificial intelligence systems become increasingly sophisticated, traditional conceptions of privacy rights face new tests that require innovative legal and technical solutions.

This thesis examines the right to be forgotten and its complex implementation in the context of artificial intelligence advancement, with particular focus on Meta's approach to data privacy in the United States and European Union jurisdictions. Meta serves as an ideal case study due to its massive global user base, extensive data collection practices, and significant investment in AI technology. Meta is at the intersection of competing privacy frameworks and commercial imperatives and must navigate implementation challenges in both regions as it advances. The tension between data collection practices that fuel AI advancement and individuals' right to control their own digital footprint

represent a critical inflection point in privacy governance. While the EU has established comprehensive data protection frameworks through the GDPR, the United States has adopted a more sectoral approach, creating fragmented regulatory landscapes that corporations such as Meta must navigate. This disparity between different regions' privacy frameworks underpins a fundamental difference in privacy conceptualization and highlights the challenges in enforcing data erasure rights in a digital ecosystem where information permanence is the default.

Through a systematic analysis, this thesis will first establish the historical evolution of privacy rights and their theoretical foundations, tracing key developments from Warren and Brandeis to contemporary digital challenges. It then compares the legal frameworks governing privacy in the EU and US, with particular attention to how the right to be forgotten has been codified in the GDPR and interpreted through landmark cases. An examination of Meta's specific data collection and privacy policies follows, highlighting the company's responses to different regulatory environments and the strategic compromises made in each jurisdiction.

As Large Language Models (LLMs) and other AI systems continue to evolve, particularly with agentic capabilities, this thesis investigates the increasingly pressing questions surrounding data control, algorithmic transparency and the practical implementation of the right to be forgotten in systems where data permanence is the default. The contrast between different regions' privacy frameworks not only reveals fundamental differences in privacy conceptualization but also highlights the practical difficulties in enforcing data erasure rights. Ultimately, this thesis proposes a reconceptualization of privacy rights that moves beyond binary public/private distinctions toward a

progressive realization approach that accounts for the technical realities of AI systems while preserving meaningful individual control over personal information.

II. Historical Context and Theoretical Framework

Evolution of Privacy Rights

The need to protect personal information and data on the internet stems from a history of legal protections for personal and private property as well as personal information. Within the United States, privacy rights were brought to the forefront in 1890 by Samuel Warren and Louis Brandeis through their Harvard Review article "The Right to Privacy".[1] Characterizing the changing nature of the press at the time, Warren and Brandeis connect the right to life with the "right to enjoy life– be let alone." The context of the creation of this article was the increase of photography and newspapers producing intimate details on people's personal lives. Their article traced how political, social, economic, and particularly technological changes inherently require new protections of rights. The new technologies of photography and newspapers enabled the press to invade personal privacy to such an extent that Warren and Brandeis describe them as producing "mental pain and distress, far greater than could be inflicted by mere bodily injury." Publication, they believe, is the primary way to understand why a right to privacy is important in the sense that publication is the act of making the private, public. The press was making the choice of publication without respect for ownership of information. Through this, the authors connected the common law around physical property with

[1] Warren and Brandeis, 'The Right to Privacy.'

personal information, photographs, and the only caveat made was for when information was within the public interest or of privileged communication. In response, they stated that there must be legal remedies such as civil damages for violations of the right to privacy and legal frameworks need to be established recognizing privacy as a distinct legal right protecting one from the publication of private matters. The Warren and Brandeis article established intellectual property and the basis for personal privacy laws within the United States.

Further conceptions of privacy in the United States have evolved beyond Warren and Brandeis' right to be "let alone." Scholars such as Daniel Solove have outlined a progression of privacy theories.[2] Primarily, Solove determined that the many definitions of privacy that have been posited, such as a protection of personhood or concealment of personal information, are either too broad to hone in on what is commonly considered as private, as is the concept of personhood, or too narrow, as is the concept of the concealment of personal information leaving important personal concepts outside of the private realm.[3] However, the theory that most closely aligns with the EU's conception of privacy in the GDPR is the conception of privacy as a protection of personhood. Solove describes the protection of personhood as an extension of Warren and Brandeis' idea of "inviolate personality."[4] This conception utilizes normative assumptions about what personal attributes are necessary for an individual to protect to have security of the self. Solove draws on the philosopher Stanley Benn to describe how privacy is the protective force behind their "[engagement with] a kind of self-creative

[2] Solove, *Understanding Privacy*.

[3] Ibid.

[4] Solove, *Understanding Privacy*, 29.

enterprise".[5] Later, he notes privacy as a protection of personhood exemplified in US Supreme Court decisions such as Roe v. Wade, ensuring the ability for persons to make independent decisions.[6]

The shifting understandings of privacy further relate to a change from physical privacy to information security. In Warren and Brandeis' conception of the right to be let alone, they believed their privacy to be violated through an overstep from the press into their personal lives almost physically. The press was able to get closer to homes and family affairs than ever before. With increased technological advancement, particularly with the advent of the internet and increased government reliance on computer databases and data processing, privacy has changed to primarily protecting personal information. A member of the press no longer needs to physically go to one's home to learn private information. Instead, they may be able to simply perform a search on Google. Digital privacy transcends the original physical concerns of privacy and creates concerns about knowing what is public on the internet.

Digital Age Privacy Challenges

Following the advent of the internet, it is important to note that the data that is published online cannot feasibly be deleted. When content is uploaded to the web, it exists on every machine that has access to it. For instance, if user X uploads a photo to Facebook, and user Y downloads it, the

[5] Solove, 30

[6] Ibid.

original photo being deleted by X does not affect Y's copy. Additionally, different backups of the internet will not reflect the changes that one user makes. Non-profit and research organizations such as the Internet Archive regularly scrape the internet to preserve a historical record, fundamentally making the content on certain pages permanent. The permanence of data on the internet is a contemporary issue and conflicts with the traditional understanding of privacy, like what Warren and Brandeis wrote about in 1890.[7] The gap between society's recognition of the internet's permanence (social advancement) and the vast expansion of information available online (material advancement) exemplifies what sociologists call "cultural lag".[8] In addressing this phenomenon, Kwak suggests that understanding people's motivations for wanting their digital data erased creates a pathway to reduce this lag through collective agreement on appropriate practices. Overall, there are six reasons for users to have their data erased: information disclosure, content sensitivity, social reputation, control over further processing, system/process, and sociality.[9] The validity of a given user's request for erasure stems from their role as the data subject– whether they or another organization have the ability to edit , remove, or access their data. When the user is in control of their data, they are more concerned with information disclosure, meaning the understanding of the lifecycle of their data–where will it end up even if it was placed on the internet with private intentions, and who will use it? On the other hand, when the user is not in control of their online content, they may be concerned with content

[7] Waddell, "Your Data Is Forever."

[8] Kwak, Lee, and Lee, "Could You Ever Forget Me?"

[9] Ibid.

sensitivity–what are the repercussions for uploading specific content? Fundamentally, both concerns surround the uses of one's "digital footprint".

When maintaining one's digital footprint, data permanence becomes a primary concern. Any amount of data placed on the internet will most likely stay there. Content believed to be private has appeared in court hearings and caused individuals to be terminated from their place of employment.[10] Companies create acknowledgments of digital permanence when uploading content to their platform. Facebook explicitly states all of the data and information collected from users and the many ways in which they plan to use it after it is collected.[11] However, it is not immediately clear to the average user the vast amount of data that is collected, and the use that is gained from it. The vast amounts of data that are collected are used, tracked, and modeled to assist Facebook and other platforms in selling advertisements and keep one on the platform. Warren and Brandeis' article did not conceptualize the scale in which private information could be utilized by an organization for its own benefit.

Corporations such as Meta are able to collect basic personal information based on what a user provides them, and additionally are able to extrapolate details based on interactions with their platform. Post interactions such as "likes," the length of time spent reading a post, comments, and simply who a user has followed on a platform such as Facebook allow the creation of predictive profiles for a user.[12] These profiles are then used as analytic markers that are sold to advertisers.

---

[10] Keenan, "On the Internet, Things Never Go Away Completely."

[11] Facebook, "Privacy Policy," Facebook, last modified April 4, 2023, https://www.facebook.com/privacy/policy/.

[12] Facebook Business Help, "About Data Sources," Facebook, accessed April 22, 2025, https://www.facebook.com/business/help/318580098318734?id=369013183583436.

Research has revealed that certain characteristics of data collected, such as the time elapsed since its collection, its duration, and its frequency, pose increasing privacy risks.[13] Additionally the level of dimensionality in a dataset, the mode of analysis, sample size, and population demographics increase privacy concerns for data collection.[14] No matter the specific context of the data collection, these factors provide increased ability for an organization to pose privacy risks including personal identifiability and vulnerability.[15] Primarily, an organization mitigates privacy concerns through control of computation, inference, and use of collected data. These controls are typically placed through a mix of policy, disclosure, and data infrastructure protections.[16]

Some of the difficulty in modern data privacy laws is the fact that they focus on "personal" instead of strictly "private" information. Between Europe and the United States, a difference in protections for personal information is apparent. European conceptions of privacy extend the category of a "fundamentally vulnerable individual not just in the private setting of their home, but in the public domain of employment, community, and polity and extends an entitlement to basic human dignity to those public realms".[17] The United States, in contrast, has a more limited extension of privacy into the public domain. Consequently, it defines privacy, to a certain extent, as personal information protected by the Fourth Amendment, which safeguards individuals against unreasonable

[13] Altman et al., "Practical Approaches to Big Data Privacy over Time."
[14] Ibid.
[15] Ibid.
[16] Altman et al., "Towards a Modern Approach to Privacy-Aware Government Data Releases."
[17] Kohl, "The Right to be Forgotten in Data Protection Law and Two Western Cultures of Privacy."

searches.[18] These differences create a stark contrast in the legal foundation for data privacy in both regions, and allow European laws to more explicitly protect users' personal information. It is important to note however that within the United States, specific categories of data such as medical data are explicitly protected and extend into the public sphere.

Using disparate data sources together through data aggregation poses further risks to personal privacy. A user may willingly provide specific information to one source but not another on the internet; however, a third source may collect both data sources and combine them, providing multiple links to information that was assumed by the user to be separate. Additionally, separate data sources are anonymous; combining them may lead to de-anonymization. A high-profile example of this is the combination of anonymized Netflix data with public IMDB account data to match personally identifiable information with a previously anonymized dataset.[19] Cases such as this create a direct understanding that disparate data sources on the internet can have direct consequences for an individual's anonymity on the internet.

With increased data collection and aggregation, behavioral tracking has increased on social media. Studies that added one explicit form of tracking for users' interactions on Facebook were able to create direct connections between post image content and user emotional state.[20] In the digital age,

[18] Ibid.

[19] Narayanan and Shmatikov, "Robust De-Anonymization of Large Sparse Datasets."

[20] Šola, Mikac, and Rončević, "Tracking Unconscious Response to Visual Stimuli to Better Understand a Pattern of Human Behavior on a Facebook Page."

one's backyard is no longer at risk. Instead it is a protection of one's self in the purest sense: personality, agency, and longevity.

Surveillance Capitalism Framework

Shoshana Zuboff, in her book The Age of Surveillance Capitalism, introduces the concept of "behavioral surplus" to describe the excess personal data collected through online interactions, such as clicks and time spent on webpages.[21] Surveillance capitalism transforms human experiences into behavioral data that exceeds what's needed for product improvement. This surplus becomes a proprietary resource processed through machine intelligence to create predictive products about future human behavior, which are then traded in what Zuboff calls "behavioral futures markets".[22] Crucially, this system goes beyond merely predicting behavior. Zuboff emphasizes that surveillance capitalism actively seeks to modify and intervene in individual behavior through targeted content and design choices. This manipulative capacity gives organizations unprecedented power not just to anticipate a user's actions based on personal information they didn't explicitly provide, but to influence those actions in real time. Large corporations build comprehensive user profiles that can predict personally identifiable information before users voluntarily disclose it. The economic incentive structure promotes ever-increasing data collection, as more detailed behavioral data enables more

[21] Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (New York: PublicAffairs, 2019)
[22] Zuboff, 2019, p. 14

accurate predictive models and greater opportunities for behavioral modification, creating a self-reinforcing cycle that serves corporate interests above all.

With the data collected in 2017, Facebook was able to produce "six million predictions of human behavior every second" using artificial intelligence.[23] Surveillance capitalism works to make predicted human behavior a product, and data collection allows for it to happen. However, Zuboff describes that surveillance capitalists such as those at Meta fear government laws, regulations, and citizen disapproval.[24] This is seen for models such as Meta's flagship AI chat service Meta AI, where integration into the European market, which has explicit data protections through the GDPR, was delayed by two years.[25]

Large Language Models pose a new avenue for data collection and encourage sensitive disclosures of personal information when interacting with a chat-based artificial intelligence model.[26] LLMs are increasingly utilized in private and public sector organizations as well as peoples' personal lives creating another scenario in which a new technological advancement is released and utilized without in-depth testing and guidelines. LLMs pose further privacy risks due to the model's ability to memorize information present in training sets. Many models explicitly do not choose to save user chat information for future training or customization in some ways to mitigate recitation of user-provided chat information. However, the only way to ensure that personal information provided to a chat

[23] Kohl, "The Right to be Forgotten in Data Protection Law and Two Western Cultures of Privacy."
[24] Ibid.
[25] Sawers, "Meta AI Is Finally Coming to the EU, but with Limitations."
[26] Zhang et al., "'It's a Fair Game', or Is It?"

interface is secure and preserve privacy is to only utilize datasets that were designed to be public.[27] The data collected for the training of a large language model, however, has no guarantee that it was intended for public use in this way.

III. Legal Frameworks for Privacy Protection

United States Sectoral Privacy Frameworks

Within the United States, privacy laws have primarily formed around specific industry sectors and needs, such as the Health Insurance Portability and Accountability Act (HIPAA) in 1996.[28] This act created national standards for sensitive patient health data. Similarly the Children's Online Privacy Protection Act (COPPA) established requirements for website and online services operators with services for children under the age of 13.[29] The State of California, with the California Online Privacy Protection Act (CalOPPA), was the first state requiring commercial websites and online services to have a privacy policy, shifting the digital privacy legal frameworks towards a state-centric approach.[30] Before the direct sectoral approaches, the United States passed the Privacy Act of 1974, regulating how federal agencies collected, maintained, used, and disseminated personally identifiable information.

[27] Brown et al., "What Does It Mean for a Language Model to Preserve Privacy?"

[28] CDC, "Health Insurance Portability and Accountability Act of 1996 (HIPAA)."

[29] "Children's Online Privacy Protection Rule ('COPPA')."

[30] "The California Online Privacy Protection Act (CalOPPA)."

Within the United States, this established the precedent of an individual's ability to access records and set limits on use and disclosure without personal consent.[31]

By splitting up privacy protections across different sectors such as health or minor safety on the internet, the United States' privacy frameworks exist in contrast to much of the European Union. The EU established the EU Data Protection Directive, which created congruent legal guidelines for data protection across member states and key government agencies and business sectors.[32] The shared guidelines developed a rights-based approach to data protection and ensured that every individual within the EU had the same protections no matter which country they were in. The contrast of the EU blanket protections and the US sectoral protections created the need for frameworks such as the EU-US Safe Harbor Framework in 2000.[33] This framework created mechanisms for American companies to certify compliance with the European data protections and ideally acted as the bridge between the two conceptions of data protections in the two regions. However, in 2015, the Safe Harbor Framework had to be replaced due to US companies failing to comply with the requirements needed to operate within the EU. The ability for companies to operate at different levels of protection within the United States and in the European Union provided the ability to do just that, causing the need for more stringent legislation and intergovernmental agreements for data transfers between the two regions.

[31] "Office of Privacy and Civil Liberties | Privacy Act of 1974."

[32] "The History of the General Data Protection Regulation | European Data Protection Supervisor."

[33] "U.S.-EU Safe Harbor Framework."

From Data Protection Directive to GDPR

The European Union has had specific data protection laws since 1995 with the Data Protection Directive, which protected individuals in regard to the processing of their personal data and the free movement of their data.[34] In 2012, there was a push to expand the data protections outlined in the Data Protection Directive for the digital economy that eventually resulted in the 2014 adoption of the GDPR. The GDPR introduced several requirements, such as "freely given, specific, informed, and unambiguous" consent.[35] Additionally, individuals in the EU have the right to access a confirmation about whether their personal data is being processed in Article 15, as well as a requirement for privacy by design and default in Article 25, stipulating that only necessary data can be provided to the "controller" by default. Specific data types, such as health, biometric, and genetic data, are explicitly protected under the GDPR. A data controller "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".[36] In simple terms, a data controller decides what data is processed, why and how it is processed.

The GDPR stands on principles such as data minimization–the act of only using adequate, relevant or limited data for processing– and assurance of security and confidentiality of personal data.[37] Furthermore, the data controller must be able to demonstrate compliance with the principles of

[34] "The History of the General Data Protection Regulation | European Data Protection Supervisor."
[35] Ibid.
[36] "Art. 4 GDPR – Definitions."
[37] "Art. 5 GDPR – Principles Relating to Processing of Personal Data."

the processing of personal data. Under Article 6 of the GDPR, processing of data is only lawful if the data subject provides consent for one or more specific purposes.[38] An important stipulation of the GDPR is the ability for the data subject to withdraw consent at any given time for the processing of their personal data. If the subject is to withdraw their consent, further data processing will cease to be legal.[39] Article eight of the GDPR refers to the consent of children for data processing. The legality of data processing only exists if the child is older than 16 years of age or has the explicit consent of the legal guardian of the child.[40] The GDPR also contains protections for data containing personal identities such as religious and political beliefs, sexual orientation, or genetic data.[41] Additionally, information relating to criminal convictions or past offenses is not able to be processed except by the control of the official authority–typically a government agency.[42]

For erasure rights in Article 17 of the GDPR, the data subject specifically is able to request the removal of data when it is no longer lawful under Article 6. The requirements for removal are proof of a lack of given consent, or if the user believes that the data being processed contains special categories of personal data under Article 9 including data revealing racial or ethnic origin, political opinion, religious or philosophical beliefs, union membership, genetic and biometric data for the purpose of identification, health data, and data concerning a person's sexual orientation.[43] Data controllers and

[38] "Art. 6 GDPR – Lawfulness of Processing."

[39] "Art. 7 GDPR – Conditions for Consent."

[40] "Art. 8 GDPR – Conditions Applicable to Child's Consent in Relation to Information Society Services."

[41] "Art. 9 GDPR – Processing of Special Categories of Personal Data

[42] "Art. 10 GDPR – Processing of Personal Data Relating to Criminal Convictions and Offences."

[43] "Art. 9 GDPR"

processors are required to have data protection by design and by default, and strict rules are in place for the sharing of information and data between different controllers, controllers to processors, and processors to processors. For instance, specific or general written authorization for the data controller must exist for processors to engage with other processors and guarantees of processing protections and standards must be met for a processor to utilize data to begin with.[44]

The Right to Be Forgotten: Legal Precedent and Implementation

The legal precedent for direct complaint-based data erasure rights stems primarily from Google Spain SL v. Agencia Española de Protección de Datos. Decided in 2014 by the Court of Justice of the European Union (CJEU), this case had a complaint form from Spanish national Costeja González against a newspaper, Google Spain, and Google Inc. The complaint was a request to remove or change the record of attachment and garnishment proceedings in 1998 that appeared when the complainant's name was entered on Google. González's argument was that the legal proceedings were previously adjudicated and therefore should not appear online. Spain's Data Protection Agency decided that while the newspaper could maintain its digital record of the court case, Google, as a search engine, was a "data controller", meaning that it processed personal data through its search engine algorithm.[45]

[44] "Art. 28 GDPR – Processor."

[45] "Google Spain SL v. Agencia Española de Protección de Datos."

Connected with Warren and Brandeis, this decision held up the idea that the subject's rights to their personal data must be in proportion to the general public's interest in the data.

The interpretation of the erasure request is what decides if data requested by a user will be erased. For instance, specifically in the UK personal information that may be important for reasons of public health, scientific research, or freedom of expression will not be erased. Article 17 of the General Data Protection Regulation outlines when the right to be forgotten applies and was derived from the Google Spain SL v. Agencia Española de Protección de Datos.[46] However, the method of erasure, what constitutes erasure, and direct implementation of the right allow for discretion on the part of the data controller.

> "Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data".[47]

This discretionary element of Article 17 reveals the gap between theoretical rights and practical implementation in data protection law. The regulation's language— "reasonable steps"—and considerations of "available technology and cost of implementation", create a framework where

[46] "Right to Erasure."

[47] "Art. 17 GDPR – Right to Erasure ('Right to Be Forgotten')."

interpretation varies based on organizational capacity.[48] Large technology companies, like Meta, with sophisticated data tracking systems and resources, face different expectations for data erasure implementation than others. One company's promise of erasure is different from another, and their ability to inform other controllers of a data erasure depends on their organizational means.

Organizations may make claims that one's personal data falls under a category that protects it from erasure. The organization can make a determination that the data itself is of public interest and should not be deleted, is for archival purposes, or even make a claim that the content the user is requesting to be deleted should stay intact due to protections for one's freedom of expression and information. However, organizations might claim that the data was properly anonymized before use, allowing for an organization to forgo the process of data deletion without violating the GDPR.

Additionally, countries within the EU have different requirements for the ability of an organization to not erase data upon request. For instance, Spain has removed consent as an exemption to following all other requirements in GDPR Article 9(2)(a) which state reasons one's personal data may be deemed necessary.[49] In the country, a data subject must provide consent for a specific reason such as the defense of legal claims. Outside of the country, if the data was collected with the data subject's consent, then no other qualifications need to be met. Punishment for GDPR violations also is not uniform across the European Union. Countries have discretion for determining punishment for

[48] Ibid.

[49] "Article 9 GDPR."

GDPR violations within the fines and penalties section of the GDPR.[50] Germany's data protection laws extend criminal offenses including imprisonment for breaches of the GDPR, whereas other countries merely levy fines.

IV. Meta's Approach to Data Privacy and AI

Current Policies and Practices

When a user signs up for an account with Meta or uses a product created by the company, they become subject to the organization's End User License Agreement (EULA).[51] Usually, this is shown to the user in an overwhelming block of text with an option to accept, or it is shown in the bottom of the sign-up page with a statement such as "By clicking Sign Up, you agree to our Terms, Privacy Policy, and Cookies Policy." This ensures all users of Meta products are subject to privacy policies that allow for user data to both be collected and used by the organization. To use their products, you must agree to their policies.

With the advent of Large Language Models (LLMs), new policies regarding Artificial Intelligence (AI) have been created. Along these lines, Meta has developed specific policies for AI.[52] These policies include requirements for eligibility, like the requirement to be 13 years of age or older in the United States, and standards of acceptable uses forbidding violations of local laws, disseminating

[50] "Fines / Penalties."

[51] Facebook, "Terms of Service," Facebook, accessed April 22, 2025, https://www.facebook.com/terms/.

[52] Facebook, "Meta AI Terms and Conditions," Facebook, accessed April 22, 2025, https://www.facebook.com/legal/eu-ai-terms.

adult content, or content that infringes upon the rights of others– specifically stipulating intellectual property rights and privacy rights.

In terms of personal information, Meta states that they "may process personal information, including any sensitive information, as described in the Privacy Policy, to improve AIs and related technology."[53] The Meta Privacy Center has a specific page about Generative AI that is geared towards instructing users about their AI products and the data that is used for their creation.

In both the US and the EU, Meta's privacy policy governs the processing of personal information for their AI products. Additionally, the language around the processing of personal information differs between the EU and the US; however, in both statements, the exchange of processed personal information for platform tools and services is apparent, for instance, in the EU AI policy: "The Meta Privacy Policy governs our processing of personal information processed through the AIs and explains how we collect, use, and share personal information - including personal information you submit as Prompts, Feedback, or other Content, your activity across our products, and personal information we collect from and about your device. We process personal information, as described in the Privacy Policy, to provide the service to you".[54]

In both regions, Meta declares their use of public information for the training of their models. Specifically, Meta does not use private information from messages with other users or family members

[53] Ibid.

[54] Facebook, "Meta AI Terms and Conditions," Facebook, accessed April 22, 2025, https://www.facebook.com/legal/eu-ai-terms.

for their model training. However, they do use any communication with their models for personalizing the results of the model.[55] Meta's use of personal information is documented and clear for the European Union. Its process for release in the region has taken longer than in the US due to the regulations set forth by the GDPR. In the United States, the Meta AI tools have been present since 2023; however, it was only March 19, 2025, that products were released in European countries.[56] Its delayed release stems from regulatory bodies requesting further actions such as an opt-out for public data scraping for EU users.

While the extent of the different categories of data collected by Meta are transparent through their privacy center, the extent of use is more vague. They utilize the collected data for personalization, product improvement, prevention of harmful behavior, advertisement, and research.[57] Targeted advertisements on Meta's platforms don't work by sharing your information to advertisers. Instead, Meta hides specific information from an advertiser and allows them to market to specific demographics, perhaps providing categories or recommendations for targeted advertising.[58] Additionally, external companies provide Meta with a given user's interaction with their own platforms, allowing Meta to provide increasingly targeted advertisements.[59] Between the EU and the

[55] Facebook, "Your Interactions with AI Features," Facebook, accessed April 22, 2025, https://www.facebook.com/privacy/dialog/your-interactions-with-ai-features.
[56] Meta, "Europe, Meet Your Newest Assistant: Meta AI," Facebook, March 19, 2025, https://about.fb.com/news/2025/03/europe-meet-your-newest-assistant-meta-ai/.
[57] Facebook, "How Do We Use Your Information? (Section 2)," Meta Privacy Policy, accessed April 22, 2025, https://mbasic.facebook.com/privacy/policy/printable/#2-HowDoWeUse.
[58] Facebook, "Does Meta Sell My Information?," Meta Privacy Center, accessed April 22, 2025, https://www.facebook.com/privacy/dialog/does-meta-sell-my-info.
[59] Ibid.

US, constraints exist for the transfer of data regarding European citizens to the United States. The EU-US Data Privacy Framework is built upon seven principles. The first principle is notice, ensuring that companies provide transparent information about their data processing, which Meta does through their privacy center. The second is choice, allowing individuals to opt out of personal information disclosure to third parties, which pairs with the third, placing responsibility and accountability for data use on the data controller. The security principle mandates that organizations take intentional steps to ensure protection against data loss, misuse, or alteration. The framework then states that individuals need to be able to view the data that is collected about them and correct, amend, or delete it when inaccuracies are discovered. Finally, individuals in the EU must have resource mechanisms if their data rights are infringed upon by a US company.[60]

Facebook somewhat follows this framework reciprocally and allows users in the United States to remove specific information related to their account activity or their entire account. However, there is no process federally in the United States that allows for data erasure requests similar to the GDPR. Therefore, unless a user resides in a state such as California with state laws mimicking the GDPR, they are only able to delete specific posts or revoke access to different public information on their accounts.[61] Meta's data collection practices align with Zuboff's understanding of data maximization through the incentive of behavioral surplus. In regions where Meta can collect more data, they do.

[60] "Key Principles and Considerations for Participation in the EU-US Data Privacy Framework."

[61] Meta, "Privacy Policy," accessed April 22, 2025, https://www.facebook.com/privacy/policy/.

Meta's maximization of data is exemplified by their attempts to implement a "Pay or Okay" model for data protection. In 2023, Meta announced a roughly € 156 per year subscription to access Facebook and Instagram without typical data collection.[62] This created a system of data privacy that is potentially inaccessible to the vast majority of EU users of Meta's platforms and allowed Meta to legally extract data that would face regulatory pressure otherwise. In the following year, Meta reduced the cost of the subscription and allowed for the ability for users to opt in to "less personalized ads" for free.[63] Meta views the paid reduction in data collection as a form of consent from the user, determining an ability to collect personal data. In a press release in December of 2024, Meta stated: "The option for people to purchase a subscription for no ads balances the requirements of European regulators while giving users choice and allowing Meta to continue serving all people in the EU, EEA, and Switzerland. In its ruling, the CJEU expressly recognized that a subscription model, like the one we are announcing, is a valid form of consent for an ads-funded service".[64] The "Pay or Okay" model allows Meta to move around stricter regulatory policies for data collection and maximize their personal data collection practices. As data collection for Large Language Models and other AI systems increases, the ability for personal data to be collected and used for these purposes may only be protected through a paywall in regions that it is legally protected.

[62] "Meta's 'Pay or Okay.'"

[63] Meta, "Facebook and Instagram to Offer Subscription for No Ads in Europe," About Meta (blog), November 2024, accessed April 22, 2025, https://about.fb.com/news/2024/11/facebook-and-instagram-to-offer-subscription-for-no-ads-in-europe/.

[64] Ibid.

AI Training Data Sources and Scraping Practices

To train the Llama family of models, Meta used a combination of different data sources including the English Common Crawl, C4, Github, Wikipedia, Gutenberg and Books3, ArXiv, and Stack Exchange.[65] CommonCrawl and C4 are extremely large datasets of scraped web data that were collected from any and all publicly available web sources. The Common Crawl is an open repository of raw web data that has been scraped from the internet starting in 2008.[66] Containing petabytes of data the common crawl exists as the largest free source of pre-training data for LLMs. Containing petabytes of data, the Common Crawl exists as the largest free source of pre-training data for LLMs. However, the Common Crawl itself is full of undesirable content, such as error codes, website menus, and even hate speech. [67] Meta utilizes parts of the Common Crawl after filtering content to target parts of the internet that are desirable for pre-training and additionally utilizes Alphabet's Colossal Clean Crawled Corpus (C4), which is a filtered and cleaned subset of the Common Crawl. The Llama models utilize content from GitHub, the web-based platform for Git which is the standard version control system used by developers; content from Wikipedia, Stack Exchange— a question and answer site typically used with STEM-related queries— and Project Gutenberg, an open-source repository of books in the public domain.[68] The last main source of pre-training data for the Llama models is Arxiv, "a curated research-sharing platform... [hosting] more than two million scholarly articles in eight

[65] Touvron et al., "LLaMA."

[66] "Common Crawl - Overview."

[67] Baack, "A Critical Analysis of the Largest Source for Generative AI Training Data."

[68] Touvron et al.

subject areas... "[69] Notably, the papers on Arxiv are not reviewed by the website itself and commonly contain copyrighted material that is simply allowed to be distributed using the paper or Arxiv's license.[70] Arxiv does maintain instructions for the harvesting of scholarly material from their website without enforcement for misuse or reproduction of content.[71]

While portions of the pre-trained content have no discernible concerns with the right to be forgotten, such as content that is within the public domain, some may raise concerns, particularly in relation to the ability of a subject possessing data erasure rights to withdraw consent for the use of their personal data, as seen in the GDPR Article 17 1(b).[72] Scraping on the internet solely exists on an opt-out system: To mitigate the scraping of a website, developers may, for instance, input a section within a robots.txt file disallowing an automated system from scraping the web page. Many websites created by large corporations disallow scraping from large language models or other systems; however, personal sites or smaller websites which may contain personal or private information most likely do not have the infrastructure, meaning that all the data is available for scraping and then training a Large Language Model.[73]

The Hamburg Commissioner for Data Protection and Freedom of Information produced a discussion on the applicability of the GDPR and erasure on LLMs. Primarily, a distinction was

[69] "About arXiv - arXiv Info."

[70] "Permissions and Reuse - arXiv Info."

[71] "arXiv Bulk Data Access - arXiv Info."

[72] "Art. 17 GDPR – Right to Erasure ('Right to Be Forgotten')."

[73] Google Developers, "Introduction to Robots.txt," Search Central Documentation, accessed April 22, 2025, https://developers.google.com/search/docs/crawling-indexing/robots/intro

established between the input and output of a LLM and the training data. The Hamburg Commission argued that the input and output of a LLM remain subject to the GDPR, whereas the data the model is trained on and the model itself are not.[74] This argument is underpinned by the tokenization of the training data as a process that removes the direct relationship between the data and the subject. Specifically, this determination lends credibility to any GDPR request that desires an inability for a model to produce information on a specific subject but not for said subjects' data to be used in the first place.

The interpretation of the GDPR protects results, however, not the collection, storage, and usage of a subject's personal data. Focusing on the resulting output of the model exemplifies the GDPR's retroactive approach to privacy protection. While it establishes guidelines for an organization for the storage and collection of data, negative effects are only seen through a retroactive, personal, grievance-based perspective. If a user does not have a complaint or desire for their personal data to be erased, then their personal data will remain to be utilized and reproduced.

V. Regional Approaches and Implementation Challenges

EU Implementation Framework

Between the United States and the European Union, Meta has two different privacy policies for AI. In the EU, Meta has an agreement with Ireland due to their more restrictive data privacy laws

[74] "Hamburger Thesen zum Personenbezug in Large Language Models."

with the Data Protection Commission.[75] This manifests through clear differences in language, with Meta's policy associated with the United States having intentionally broad statements such as "to provide more relevant or useful responses" to explain when Meta might send user data to third-party organizations. The EU, on the other hand, has specific statements on the sharing of personal data with third parties "if the AI cannot answer your query" or if a more relevant response can be provided not with the assistance of an AI tool. The primary difference, however, between the two is the inclusion of the EU online dispute resolution platform, which the US does not have. Along these lines, the EU privacy policy states Meta's obligation to "exercise professional diligence" for their services, including addressing complaints and requests for data erasure. The United States, with no such grievance system to account for, does not include a similar statement.

US Policy Evolution

With the ability of a large language model to produce information in a digestible format faster than a search engine and the near inability of private information that is present to be removed from a pre-trained model, a European-style privacy protection seems to be increasingly unrealistic. Privacy as a core concept has always been understood in both the United States and the European Union and protected in different manners. In the United States a user's right to privacy on the internet is not federally protected and limited protections are only available on the state level, particularly such as California's Consumer Privacy Act (CCPA).[76] Despite a seemingly binary definition of privacy–

[75] "Homepage | Data Protection Commission."

[76] "California Consumer Privacy Act (CCPA)."

something either being private or not– Europe's actual implementation of laws and regulations leaves citizens in a middle ground of protection. More accurately, a definition of privacy through a lens of progressive realization is needed. Progressive realization is understood as a gradual implementation of a law, focusing on progress instead of a binary true or false solution. Additionally, progressively realized human rights consider the abilities of organizations, and their available resources and ensure that goals are measurable with benchmarks throughout legal implementation. A framework based on progressive implementation would account for the discretion that organizations such as Meta have for upholding users' rights to be forgotten, and the vast amounts of data being collected. Furthermore, creating space for laws that call for progressive realization of privacy would allow for monitoring and regulation of data that is not explicitly given to Meta or that is used as training data for large language models. In both the United States and the European Union, this approach would allow for progress towards more secure personal data, both in a regulated EU and a divided United States.

In the United States, rotating priorities for AI regulation have taken place with the change of administration from President Biden to President Trump. Both administrations took different stances on regulation, with Biden adding both business and governmental regulations aimed at mitigating the risks and promoting the ethical use of AI. This included Executive Orders such as "Safe, Secure, and Trustworthy Development and Use of AI", and the AI Bill of Rights. Both documents have explicit statements supporting the protection of privacy, even stating "Agencies shall use available policy and technical tools, including privacy-enhancing technologies (PETs) where appropriate, to protect privacy and to combat the broader legal and societal risks—including the chilling of First Amendment

rights—that result from the improper collection and use of people's data".[77] President Biden created a clear intention for the protection of American users of AI technology and further posited this with rights-based language through the Blueprint for an AI Bill of Rights. The document released by the Biden administration provided the recommendations for future legislation surrounding Artificial Intelligence and included any protections that exist for Europeans under the GDPR, including data access and correction, consent withdrawal, and data deletion, and privacy by design and by default.[78] While providing extremely similar guidelines to European protections, most of the Biden administration's executive orders were rescinded by the Trump Administration.

Instead of focusing on user protections, the Trump administration has focused on research and development and reducing regulatory barriers for the production and commercialization of AI systems. With this, the United States demonstrated a lack of stability in position for data privacy projections for the quickly developing technology. The only progressive legislation for data privacy in the United States is on the state level, particularly the similar policies to the GDPR that exist in California with the CCPA. The CCPA provides California residents with privacy rights, including the right to erasure.[79] Meta benefits from the lack of personal data privacy policies in the United States and only chooses to enforce data protections when they are required to.

[77] "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

[78] "Data Privacy | OSTP."

[79] "California Consumer Privacy Act (CCPA)."

Corporate Challenges and Responses

One specific challenge with Meta's platforms is their real name policy. Worldwide, Facebook and Instagram require the legal names associated with the user to be attached to each account. Having direct personally identifiable information associated with each account creates easier access for data aggregation to connect outside information with a given Facebook account. While there exists the ability for users to make specific information public or private on their accounts, some argue that requirements for actual names necessarily restrict the choice that users have on their data privacy on Meta's platforms.[80] The adoption of the GDPR and a number of lawsuits have not caused any changes to the platform's policy.[81]

In compliance with the GDPR, Meta has received record-setting fines reaching $1.3 billion dollars by the Irish Data Protection Authority due to data transfer of European users to servers within the United States.[82] Previously, Meta was charged $263 million for data security failures, $425 million for the handling of minors' data, and $410 million for improper processing of user data for ad targeting services.[83] While the fines are steep, the staggering revenue of over $160 billion dollars turns what are supposed to be punishments given by governments into a price to pay for operation.[84] For

[80] "Kansas Journal of Law, Volume 28."
[81] Ibid.
[82] "1.2 Billion Euro Fine for Facebook as a Result of EDPB Binding Decision | European Data Protection Board."
[83] "GDPR Enforcement Tracker - List of GDPR Fines."
[84] "Annual revenue and net income generated by Meta, Statista"

corporations of Meta's size, there is little incentive to follow regulation when the primary form of punishment is monetary fines.

The large amounts of revenue also give Meta the ability to lobby governments it is subject to. In the United States, Meta spent over $24 million dollars on lobbying in 2024. One major point of lobbying and federal attention given to Meta was for the Protect Kids on Social Media Act, which sought to create stronger regulation on content and services that can affect minors on the internet. Meta, along with other tech and social media companies such as Alphabet, testified and lobbied against much of the legislation that worked to increase regulation.[85] The year-over-year increased political lobby spending demonstrates a further ability for Meta to shift legislation in their favor.

In addition to the simple ability to pay fines for GDPR violations and to lobby governments for reduced regulation, the GDPR allows for large amounts of discretion for carrying out their responsibilities, particularly for data erasure requests. Meta has the ability to determine if an exemption can apply for the maintenance of personal information such as a legal obligation or public interest. Additionally, they are the body who determines what data is verifiable as the data subject's, potentially allowing data that the subject wished to be deleted to remain on the platform. Meta has discretion on the exact implementation of erasure, which may not mean erasure. In the context of the GDPR, requests can be followed up with complete deletion or simply further anonymization. If Meta determines, under their guidelines, that the subject's data is properly anonymized, then they are able to

[85] OpenSecrets.org 2024

leave it on the platform. Finally, Meta has discretion to prove that they followed any data erasure request "without undue delay".[86] However, they notably do not have discretion over the requirement to respond to all valid erasure requests, document their decision-making process, and provide clear reasoning behind the refusal of data erasure. All entities subject to the GDPR, including subjects, processors and controllers have the ability to dispute determinations from one another or a regulatory body they are subject to, providing a right to an effective judicial remedy.[87] While the ability to dispute actions provide possibility for further accountability, it is hard to determine measurable results as the data on disputes are not readily available.

Meta's approach to GDPR compliance illustrates the limitations of current data protection frameworks when they are applied to corporations of the magnitude of Meta. Despite record-setting fines, Meta is able to fight against regulation through lobbying. Additionally, the discretion given to Meta allows them significant leeway in determining compliance with the framework. While Meta cannot avoid responding to requests, they have control over how the request is ultimately resolved, representing a fundamental power imbalance between the data subjects and the data controllers.

[86] "What Is GDPR Article 17 (Right to Erasure) and 4 Ways to Achieve Compliance."
[87] "Enforcement and Remedies under the GDPR."

VI. Future Implications: Agentic AI Systems and Data Control

The future of large language models is creating "agentic" systems which are able to operate on their own after an initial user prompt.[88] Meta is focusing on use cases with AI agents for future model releases such as Llama 4.0.[89] This ability to make decisions after human or institutional guidance begets the question of understanding Large Language Models as a potential data controller or processor, separate from the company it spawned from.

Meta has not released easy-to-use agentic abilities for the consumer with their large language models, but it is inevitable that they will, due to their track record with keeping up with their major competitors such as OpenAI or Anthropic with their release of agentic capabilities in February 2025.[90] If this model is able to make decisions and run processes unaided by a user, then it follows that it can conduct actions similar to a data processor. If this model is able to make decisions and run processes unaided by a user, then it follows that it can conduct actions similar to a data processor. The AI agent would be able to carry out an "operation or set of operations... performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation, or alteration..."[91] The AI would not be considered a processor, however, because it is not a legal entity. Authors such as Brandeis Marshall have outlined the important differences in logic between legal entities and an AI model, namely the pattern recognition

[88] "Practices for Governing Agentic AI Systems."
[89] Vanian, "Meta Product Chief Says Llama 4 Will Power AI Agents."
[90] "Claude 3.7 Sonnet and Claude Code."
[91] "Art. 4 GDPR – Definitions."

of AI compared to logical reasoning.[92] However, Marshall sheds light on a gap in the legislature about the legal repercussions for Artificial Intelligence, including potential frameworks for banning specific algorithms, systems, and tools when negative effects to society have been discovered.[93] Marshall circles questions about responsibility: if a Large Language Model provided the information of a past foreclosure for a general query about a citizen, is it the model, organization, or underlying data that must face repercussions, namely the removal of content upon the complainant's request?

A 2018 ruling by the CJEU determined that users must inform Google about the personal data they do not wish to be listed on the search results due to the impracticality of removing all "sensitive" data such as political or religious affiliation. Additionally, the CJEU determined that it is the company that should adjudicate the level of public interest compared to the individual complaint, leaving the precedent and decision for removal and the extent of such data removal to be left to Google instead of a legal adjudicating body.[94] Finally, the CJEU determined that it is only EU domains that need to be compliant with the GDPR, following a French regulator arguing that data protections should apply worldwide instead of just for those subject to European Union laws.[95]

If a similar case to Google Spain SL v. Agencia Española de Protección de Datos were to happen under different circumstances, namely a user whose garnishment proceedings appeared when asking a Large Language Model for information about said person, similar decisions would most likely

[92] Marshall, "No Legal Personhood for AI."
[93] Ibid.
[94] Massé, "EU Court Decides on Two Major 'Right to Be Forgotten' Cases."
[95] Ibid.

be made. The organization who created the model may be considered the data controller because they indexed, stored, and made the data available to users of the model according to the algorithm that underpins the model, much like Google's processing of the data for the complainant. Secondly, the CJEU clearly demonstrated the need for following the data protection laws even if the data controller did not reside in the EU. Meta, for instance, still has to protect the data of European citizens as a United States Corporation. Thirdly, even if a publication of the proceedings was lawful, just as it was found in the 2014 case, the company who created the LLM would be obligated to remove the information from the search. Importantly, this will be done after a determination of the general public's stake in the information.[96] However, this determination would, in fact, be easier to determine as the Google Spain decision was made before the GDPR. However, the difficult factor is the actual removal of content.

If a user desires to have all of or a portion of their data removed from Meta platforms, the company suggests either using the Meta Data Request Tool, which allows you to submit a direct petition to a data protection officer.[97] You then are able to select an option to delete a copy of your information and complete the erasure request form. Additionally, they provide direct access to a GDPR erasure form through their privacy center. Despite the multiple pages that a user must navigate through to access the GDPR form, the page is compliant with all legal mandates for clarity.

[96] Court of Justice of the European Union, "Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González," Judgment of the Court (Grand Chamber), Case C-131/12, May 13, 2014.
[97] "Kontaktiere Den Datenschutzbeauftragten | Facebook."

Despite the access to a data erasure request form, outcomes for data erasure are almost entirely up to Meta, as shown through the jurisprudence from the CJEU.[98] Meta does state that a data erasure request stemming from content produced by their large language models would be difficult. Namely, they cite data fragmentation and model complexity. Data fragmentation is the disparate locations that personal data can come from; there is no section of the pre-trained data for "Jane Doe." The statement on model complexity points to the current complexity in determining what data is selected at different points of a model's decision-making process.[99] Further approaches for erasure in LLM contexts are done before training, such as data anonymization, aggregation, and adjustments to the model's structured representation of knowledge, such as pointers to related or similar data points, for example: people, places, things, or concepts that are related. All of the listed attempts to erase are not actually erased. Instead, they work to prevent scenarios in which a user requests data erasure related to a LLM.

If one were to attempt to erase data that the model was trained on, they wouldn't be able to because a trained model cannot be changed. Attempts to adjust results through fine-tuning can happen and are legally adequate. But fundamentally, the data is still present "under the hood" of the large language model. Meta benefits from the discretion they are allotted because they can adjudicate compliance with the user's data erasure request despite the fact that the underlying data is still present.

[98] Court of Justice of the European Union, "Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González," Judgment of the Court (Grand Chamber), Case C-131/12, May 13, 2014.
[99] "On the Biology of a Large Language Model."

VII. Impact Assessment and Recommendations

Current Implementation Challenges

Data privacy regulations, while explicit and broad in the European Union and on the state level of the United States, are either way unable to concretely be fully protected in either region, particularly in the context of Large Language Models. The discretion given to organizations like Meta in the EU for the right to be forgotten undermines the benefits that a rights-based framework provides. Traditionally, rights are often conceived as positive or negative, with positive rights mandating a government provide rights-holders with goods, services, or protections. The right to be forgotten itself is a positive right because the data controller is responsible for providing their subjects with the ability to have their data erased. Negative rights, on the other hand, are withholding of specific action. The right to be free from unwanted tracking across websites is an example of a negative right in the context of data privacy and protections.

For a company like Meta, the common positive rights in the EU with the GDPR represent a need for infrastructure and explicit processes needed to fulfill the requests for services stated in rights frameworks. The negative data-related rights that rely on financial penalties for violation work to prevent harm to individuals, however, lose touch when brought to scale. Their business model incentivizes activity that maximizes data collection, and therefore, certain practices that exist in the US may violate a user's rights in the EU. This may have the potential for Meta to work through a regulation-first development cycle that emphasizes compliance with EU frameworks before further

deployment across the EU. This is already demonstrated through Meta's withholding of MetaAI in EU markets for over a year after it was available in the US.[100] However, if similar development and deployment cycles follow for other features, it is only a matter of time before specific features are adjusted until they are compliant with EU requirements. However, if personal data protection is the goal, the current frameworks are hindered by their own construction; a new formulation of privacy protection for digital technology is needed.

Proposed Solutions

In order to continually protect data privacy and rights such as the right to be forgotten as technology progresses, privacy needs to be reconceptualized for digital spaces. The binary classification of one's personal information as private or public provides organizations like Meta with the discretion to fulfill their own desires and not the desires of the data subject. Instead, a progressive approach to privacy may provide support for the users of Meta's platforms to maintain their listed rights. The gathered consent on the part of the user from terms and conditions, privacy policies, or EULAs makes that consent the default when users use their services. The administrative burden of withdrawing consent falls on the user instead of the organization itself. Administrative burden, as coined by Herd and Moynahan (2015), describes the labor required of a subject to access or properly utilize services that are provided by a governing body.[101] If the conception of privacy moved away from allowance and

[100] bettya, "Making AI Work Harder for Europeans."

[101] Herd and Moynihan

consent to accountability this would potentially create legislation that focuses on data stewardship regardless of the method in which a company such as Meta gained consent for data usage.

Algorithmic impact is also a large concern for Large Language Models and still is not fully understood as the emerging technology is continually evolving and being tested live. However, requirements for organizations to assess and disclose the data processing and algorithmic systems' impact to privacy and users' rights allow for users to remain informed about their personal data after it is collected, or after they provide consent.

One benefit of the EU system of privacy regulation is the harmonization of its regulation throughout the member states. Maintaining similar rights-based legal frameworks internationally ensures that rights are similarly protected, no matter where one is. Digitally in the EU, somewhat similar statements can potentially be made assuring that one's right to be forgotten will be recognized whether they are in Spain or Belgium. The US, however, lacks such federal and overarching legal frameworks and instead has left data privacy regulation to the states. The sectoral approach incentivizes Meta and other companies to only act on what is required, weakening the rights-based frameworks and language that is deployed in the laws mimicking EU legislation such as the GDPR. A unified federal legal framework is needed if data privacy rights are protected across the United States, as one State's strict guidelines will not force an organization to comply on the national level.

Fundamentally, between the EU and the US, a cultural shift is required if a progressive understanding of privacy is adopted. The hands-on approach of the EU and the hands-off approach of the US create a contradiction that allows for Meta and other companies to exploit users' personal data whenever available. A progressive framework acknowledges that privacy itself is not merely the absence

of interference or collection of one's personal data, but instead requires an affirmative protection and enablement of privacy-seeking practices.

VIII. Conclusion

The right to be forgotten stands at the forefront of privacy protection in the age of artificial intelligence. Current frameworks fall short when confronted with the realities of Large Language Models and the business imperatives of data-driven companies like Meta. This thesis has traced the evolution of privacy conceptualization from Warren and Brandeis's "right to be let alone" to the digital privacy challenges of the present day, where personal information exists in a state of near-permanent availability. The contrast between the European Union's GDPR framework and the United States' sectoral approach highlights how regional differences in privacy governance create exploitable gaps that undermine effective data protection. Meta's implementation challenges demonstrate the fundamental tension between maximizing data collection for AI advancement and respecting individual privacy rights.

The emergence of agentic AI systems further complicates this landscape, raising questions about data controller status and erasure implementation that current legal frameworks are ill-equipped to address. This analysis suggests that a binary conception of privacy as either public or private is increasingly inadequate in a digital system where personal information exists across diverse platforms. Instead, a progressive approach to privacy, emphasizing accountability rather than consent, offers a more promising path forward. Conceptualizing privacy in legal frameworks for the digital world

would shift the administrative burden of privacy protection from individuals to data controllers, establish more meaningful algorithmic impact assessment requirements, and create harmonized international standards for data erasure implementation. Without these reforms, the right to be forgotten will remain more theoretical than practical, particularly as AI systems continue to evolve. The future of privacy protection depends on a willingness to move beyond discretionary implementation and toward enforceable standards that account for both technological innovation and fundamental human rights in the digital sphere.

IX. Acknowledgments

I would like to thank my thesis advisor, Lake Polan, for the many emails, meetings, and support throughout the writing process. My preceptor, Kathleen Cavanaugh, for her belief in my topic and ability to accomplish this paper throughout the year. Additionally, I would like to thank my parents, Angie and Brandon, for their support throughout my entire academic and particularly lifelong journey. My roommates, London and Victor, for providing the table space, late-night conversations, and the best friends anyone can ask for. And Cara for giving me grace and support throughout the entire writing process.

X. Bibliography

"1.2 Billion Euro Fine for Facebook as a Result of EDPB Binding Decision | European Data Protection Board." Accessed April 23, 2025. https://www.edpb.europa.eu/news/news/2023/12-billion-euro-fine-facebook-result-edpb-binding-decision_en.

"About arXiv - arXiv Info." Accessed April 22, 2025. https://info.arxiv.org/about/index.html.

Altman, Micah, Alexandra Wood, David R O'Brien, and Urs Gasser. "Practical Approaches to Big Data Privacy over Time." International Data Privacy Law 8, no. 1 (February 1, 2018): 29–51. https://doi.org/10.1093/idpl/ipx027.

Altman, Micah, Alexandra Wood, David O'Brien, Salil Vadhan, and Urs Gasser. "Towards a Modern Approach to Privacy-Aware Government Data Releases." SSRN Scholarly Paper. Rochester, NY, May 1, 2016. https://doi.org/10.2139/ssrn.2779266.

"arXiv Bulk Data Access - arXiv Info." Accessed April 22, 2025. https://info.arxiv.org/help/bulk_data.html.

Baack, Stefan. "A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl." In The 2024 ACM Conference on Fairness, Accountability, and Transparency, 2199–2208. Rio de Janeiro Brazil: ACM, 2024. https://doi.org/10.1145/3630106.3659033.

bettya. "Making AI Work Harder for Europeans." *Meta* (blog), April 14, 2025. https://about.fb.com/news/2025/04/making-ai-work-harder-for-europeans/.

Brown, Hannah, Katherine Lee, Fatemehsadat Mireshghallah, Reza Shokri, and Florian Tramèr. "What Does It Mean for a Language Model to Preserve Privacy?" In 2022 ACM Conference on Fairness, Accountability, and Transparency, 2280–92. Seoul Republic of Korea: ACM, 2022. https://doi.org/10.1145/3531146.3534642.

CDC. “Health Insurance Portability and Accountability Act of 1996 (HIPAA).” Public Health Law, September 10, 2024. https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html.

"Claude 3.7 Sonnet and Claude Code." Accessed April 22, 2025. https://www.anthropic.com/news/claude-3-7-sonnet.

"Common Crawl - Overview." Accessed April 22, 2025. https://commoncrawl.org/overview.

Consumer Federation of California. “The California Online Privacy Protection Act (CalOPPA),” November 3, 2015. https://consumercal.org/about-cfc/cfc-education-foundation/california-online-privacy-protection-act-caloppa-3/.

Court of Justice of the European Union. "Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González." Judgment of the Court (Grand Chamber), Case C-131/12. May 13, 2014.

Exabeam. “What Is GDPR Article 17 (Right to Erasure) and 4 Ways to Achieve Compliance.” Accessed April 23, 2025. https://www.exabeam.com/explainers/gdpr-compliance/what-is-gdpr-article-17-right-to-erasure-and-4-ways-to-achieve-compliance/.

Facebook. "About Data Sources." Facebook Business Help. Accessed April 22, 2025. https://www.facebook.com/business/help/318580098318734?id=369013183583436.

Facebook. "Does Meta Sell My Information?" Meta Privacy Center. Accessed April 22, 2025. https://www.facebook.com/privacy/dialog/does-meta-sell-my-info.

Facebook. "How Do We Use Your Information? (Section 2)." Meta Privacy Policy. Accessed April 22, 2025. https://mbasic.facebook.com/privacy/policy/printable/#2-HowDoWeUse.

Facebook. "Meta AI Terms and Conditions." Facebook. Accessed April 22, 2025. https://www.facebook.com/legal/eu-ai-terms.

Facebook. "Privacy Policy." Facebook. Last modified April 4, 2023. https://www.facebook.com/privacy/policy/.

Facebook. "Terms of Service." Facebook. Accessed April 22, 2025. https://www.facebook.com/terms/.

Facebook. "Your Interactions with AI Features." Facebook. Accessed April 22, 2025. https://www.facebook.com/privacy/dialog/your-interactions-with-ai-features.

Federal Register. "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," November 1, 2023. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Federal Trade Commission. "Children's Online Privacy Protection Rule ('COPPA')," July 25, 2013. https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa.

Federal Trade Commission. "U.S.-EU Safe Harbor Framework," February 6, 2025. https://www.ftc.gov/business-guidance/privacy-security/us-eu-safe-harbor-framework.

"GDPR Enforcement Tracker - List of GDPR Fines." Accessed April 23, 2025. https://www.enforcementtracker.com.

GDPRhub. "Article 9 GDPR." Accessed April 22, 2025. https://gdprhub.eu/index.php?title=Article_9_GDPR.

General Data Protection Regulation (GDPR). "Art. 4 GDPR – Definitions." Accessed April 22, 2025. https://gdpr-info.eu/art-4-gdpr/.

General Data Protection Regulation (GDPR). "Art. 5 GDPR – Principles Relating to Processing of Personal Data." Accessed April 22, 2025. https://gdpr-info.eu/art-5-gdpr/.

General Data Protection Regulation (GDPR). "Art. 6 GDPR – Lawfulness of Processing." Accessed April 22, 2025. https://gdpr-info.eu/art-6-gdpr/.

General Data Protection Regulation (GDPR). "Art. 7 GDPR – Conditions for Consent." Accessed April 22, 2025. https://gdpr-info.eu/art-7-gdpr/.

General Data Protection Regulation (GDPR). "Art. 8 GDPR – Conditions Applicable to Child's Consent in Relation to Information Society Services." Accessed April 22, 2025. https://gdpr-info.eu/art-8-gdpr/.

General Data Protection Regulation (GDPR). "Art. 10 GDPR – Processing of Personal Data Relating to Criminal Convictions and Offences." Accessed April 22, 2025. https://gdpr-info.eu/art-10-gdpr/.

General Data Protection Regulation (GDPR). "Art. 17 GDPR – Right to Erasure ('Right to Be Forgotten')." Accessed April 22, 2025. https://gdpr-info.eu/art-17-gdpr/.

General Data Protection Regulation (GDPR). "Art. 28 GDPR – Processor." Accessed April 23, 2025. https://gdpr-info.eu/art-28-gdpr/.

General Data Protection Regulation (GDPR). "Fines / Penalties." Accessed April 22, 2025. https://gdpr-info.eu/issues/fines-penalties/.

Global Freedom of Expression. "Google Spain SL v. Agencia Española de Protección de Datos." Accessed April 22, 2025. https://globalfreedomofexpression.columbia.edu/cases/google-spain-sl-v-agencia-espanola-de-proteccion-de-datos-aepd/.

Google Developers. "Introduction to Robots.txt." Search Central Documentation. Accessed April 22, 2025. https://developers.google.com/search/docs/crawling-indexing/robots/intro.

Donald Moynihan, Pamela Herd, Hope Harvey, Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions, Journal of Public Administration Research and Theory, Volume 25, Issue 1, January 2015, Pages 43–69, https://doi.org/10.1093/jopart/muu009

HmbBfDI. "Hamburger Thesen zum Personenbezug in Large Language Models," July 15, 2024. https://datenschutz-hamburg.de/news/hamburger-thesen-zum-personenbezug-in-large-language-models.

Homepage | Data Protection Commission. "Homepage | Data Protection Commission." Accessed April 22, 2025. https://www.dataprotection.ie/.

Kansas Journal of Law & Public Policy. “Volume 28.” Accessed April 23, 2025. https://lawjournal.ku.edu/archives/volume-xxviii/.

Keenan, Thomas P. "On the Internet, Things Never Go Away Completely." In The Future of Identity in the Information Society, edited by Simone Fischer-Hübner, Penny Duquenoy, Albin Zuccato, and Leonardo Martucci, 37–50. Boston, MA: Springer US, 2008. https://doi.org/10.1007/978-0-387-79026-8_3.

"Key Principles and Considerations for Participation in the EU-US Data Privacy Framework." Accessed April 22, 2025. https://katten.com/key-principles-and-considerations-for-participation-in-the-eu-us-data-privacy-framework.

Kohl, Uta. "THE RIGHT TO BE FORGOTTEN IN DATA PROTECTION LAW AND TWO WESTERN CULTURES OF PRIVACY." International & Comparative Law Quarterly 72, no. 3 (July 2023): 737–69. https://doi.org/10.1017/S0020589323000258.

"Kontaktiere Den Datenschutzbeauftragten | Facebook," June 18, 2019. https://web.archive.org/web/20190618173807/https://www.facebook.com/help/contact/540977946302970.

Kordi, Mehrdad. "Surveillance Capitalism: The Transformation of Raw Online Data into Valuable Assets by High-Tech Companies—Is AI Governance a Threat or a Solution to Our Privacy Concerns?" In The Palgrave Handbook of Sustainable Digitalization for Business, Industry, and Society, edited by Myriam Ertz, Urvashi Tandon, Shouheng Sun, Joan Torrent-Sellens, and Emine Sarigöllü, 401–16. Cham: Springer International Publishing, 2024. https://doi.org/10.1007/978-3-031-58795-5_18.

Kwak, Chanhee, Junyeong Lee, and Heeseok Lee. "Could You Ever Forget Me? Why People Want to Be Forgotten Online." Journal of Business Ethics 179, no. 1 (August 2022): 25–42. https://doi.org/10.1007/s10551-021-04747-x.

Lexology. “Enforcement and Remedies under the GDPR,” September 18, 2017. https://www.lexology.com/library/detail.aspx?g=35f640a4-0a8a-4a81-becb-392fcb201042.

Marshall, Brandeis. "No Legal Personhood for AI." Patterns 4, no. 11 (November 10, 2023): 100861. https://doi.org/10.1016/j.patter.2023.100861.

Massé, Eliška Pírková, Estelle. "EU Court Decides on Two Major 'Right to Be Forgotten' Cases: There Are No Winners Here." Access Now, October 23, 2019. https://www.accessnow.org/eu-court-decides-on-two-major-right-to-be-forgotten-cases-there-are-no-winners-here/.

“Annual revenue and net income generated by Meta, Statista. https://www.statista.com/statistics/277229/facebooks-annual-revenue-and-net-income/”

Meta. "Europe, Meet Your Newest Assistant: Meta AI." Facebook, March 19, 2025. https://about.fb.com/news/2025/03/europe-meet-your-newest-assistant-meta-ai/.

Meta. "Privacy Policy." Last modified February 16, 2025. https://www.facebook.com/privacy/policy/.

"Meta's 'Pay or Okay': Is This the Final Challenge for EU GDPR?" Accessed April 22, 2025. https://epc.eu/en/Publications/Metas-Pay-or-Okay-Is-this-the-final-challenge-for-EU-GDPR~5672dc.

Narayanan, Arvind, and Vitaly Shmatikov. "Robust De-Anonymization of Large Sparse Datasets." In 2008 IEEE Symposium on Security and Privacy (Sp 2008), 111–25, 2008. https://doi.org/10.1109/SP.2008.33.

“Office of Privacy and Civil Liberties | Privacy Act of 1974,” June 16, 2014. https://www.justice.gov/opcl/privacy-act-1974.

OpenSecrets.org. 2024. "Federal lobbying spending tops $1.2 billion in first quarter of 2024." May 30, 2024. https://www.opensecrets.org/news/2024/05/federal-lobbying-spending-tops-1-billion-in-q1-first-quarter-of-2024/.

"Permissions and Reuse - arXiv Info." Accessed April 22, 2025. https://info.arxiv.org/help/license/reuse.html.

Pop, Cristina. “EU vs US: What Are the Differences Between Their Data Privacy Laws?” *Endpoint Protector Blog* (blog), November 15, 2023. https://www.endpointprotector.com/blog/eu-vs-us-what-are-the-differences-between-their-data-privacy-laws.

"Practices for Governing Agentic AI Systems," February 14, 2024. https://openai.com/index/practices-for-governing-agentic-ai-systems/.

"Right to Erasure," January 10, 2025. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/right-to-erasure/.

Sawers, Paul. "Meta AI Is Finally Coming to the EU, but with Limitations." TechCrunch (blog), March 20, 2025. https://techcrunch.com/2025/03/20/meta-ai-is-finally-coming-to-the-eu-but-with-limitations/.

Šola, Hedda Martina, Mirta Mikac, and Ivana Rončević. "Tracking Unconscious Response to Visual Stimuli to Better Understand a Pattern of Human Behavior on a Facebook Page." Journal of Innovation & Knowledge 7, no. 1 (January 1, 2022): 100166. https://doi.org/10.1016/j.jik.2022.100166.

Solove, Daniel J. Understanding Privacy. First Harvard University Press paperback edition. Cambridge, Massachusetts London, England: Harvard University Press, 2009.

State of California - Department of Justice - Office of the Attorney General. "California Consumer Privacy Act (CCPA)," October 15, 2018. https://oag.ca.gov/privacy/ccpa.

"The History of the General Data Protection Regulation | European Data Protection Supervisor," May 25, 2018. https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en.

The White House. "Data Privacy | OSTP." Accessed April 22, 2025. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/data-privacy/.

Touvron, Hugo, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, et al. "LLaMA: Open and Efficient Foundation Language Models." arXiv, February 27, 2023. https://doi.org/10.48550/arXiv.2302.13971.

Transformer Circuits. "On the Biology of a Large Language Model." Accessed April 22, 2025. https://transformer-circuits.pub/2025/attribution-graphs/biology.html.

Vanian, Jonathan. "Meta Product Chief Says Llama 4 Will Power AI Agents." CNBC, March 5, 2025. https://www.cnbc.com/2025/03/05/meta-product-chief-says-llama-4-will-power-ai-agents.html.

Waddell, Kaveh. "Your Data Is Forever." The Atlantic (blog), June 2, 2016. https://www.theatlantic.com/technology/archive/2016/06/your-data-is-forever/485219/.

"Warren and Brandeis, 'The Right to Privacy.'" Accessed April 22, 2025. https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html.

Zhang, Zhiping, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. "'It's a Fair Game', or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents." In Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–26. Honolulu HI USA: ACM, 2024. https://doi.org/10.1145/3613904.3642385.