Jean Monnet Saar

EUROPARECHT ONLINE

# Saar Blueprints

**Asena Serdaroglu**

From Search Engines to Large Language Models: Re-evaluating the Right to Be Forgotten in the Age of Artificial Intelligence

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

**10/2025 EN**

**About the author**

Asena Serdaroglu, is a graduate of the Master of Laws program at the Europa-Institut of Saarland University, where she obtained an LL.M. in "European Economic Law" and "Foreign Trade and Investment". She completed her LL.B at Istanbul University.

**Preface**

This publication is part of an e-paper series (Saar Blueprints), which was created as part of the Jean-Monnet-Saar activity of the Jean-Monnet Chair of Prof. Dr. Thomas Giegerich, LL.M. at the Europa-Institut of Saarland University, Germany.

The opinions and analysis within these papers reflects the author's views and is not to be associated with Jean-Monnet-Saar or the respective employers or institutions that the author works for.

**Editor**
Lehrstuhl Prof. Dr. Thomas Giegerich
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

**ISSN**
2199-0050 (Saar Blueprints)
DOI: 10.17176/20251013-090218-0

**Citation**
*Serdaroglu*, From Search Engines to Large Language Models: Re-evaluating the Right to Be Forgotten in the Age of Artificial Intelligence, Saar Blueprints 10/25, accessible via: https://jean-monnet-saar.eu/wp-content/uploads/2025/10/Saar-Blueprint_Asena-Serdaroglu.pdf.

Funded by the **Deutsche Forschungsgemeinschaft** (DFG) – Project No.: 52557664

**Table of Contents**

## List of Abbreviations

| | | |
|---|---|---|
| 1. | AEPD | Agencia Española de Protección de Datos |
| 2. | AG | Advocate General |
| 3. | AI | Artificial Intelligence |
| 4. | CEDPO | Confederation of European Data Protection Organisations |
| 5. | CJEU | Court of Justice of the European Union |
| 6. | CNIL | Commission Nationale de l'Informatique et des Libertés |
| 7. | DPA | Data Protection Authority |
| 8. | ECHR | European Convention on Human Rights |
| 9. | EDPB | European Data Protection Board |
| 10. | EDPS | European Data Protection Supervisor |
| 11. | ENISA | European Network and Information Security Agency |
| 12. | EPRS | European Parliamentary Research Service |
| 13. | EU | European Union |
| 14. | GDPR | General Data Protection Regulation |
| 15. | ICO | Information Commissioner's Office |
| 16. | LLM | Large Language Model |
| 17. | PIA | Privacy Impact Assessment |
| 18. | RTBF | Right to Be Forgotten |
| 19. | WP29 | Article 29 Data Protection Working Party |

## A. Introduction

In the digital age, personal data has become one of the most valuable and contested resources. As digital services expand into nearly every facet of daily life, individuals face growing concerns over how their personal information is collected, processed and retained by powerful technological actors. This concern is further grown by the rise of artificial intelligence (AI) systems, particularly large language models (LLMs), which rely on vast datasets, including potentially personal data, to generate human-like outputs. As a result of the inclusion of personal data, the question of how fundamental data protection rights apply to these emerging technologies has become increasingly urgent.

One of the most prominent legal mechanisms for safeguarding personal data in the European Union (EU) is the Right to Be Forgotten (RTBF), codified in Article 17 of the General Data Protection Regulation (GDPR or the Regulation)[1]. The RTBF emerged from the landmark *Google Spain SL v Agencia Española de Protección de Datos* (*Google Spain*) case in 2014, where the Court of Justice of the European Union (CJEU or Court) held that individuals have the right to request the removal of links to personal information from search engine results.[2] This judgment was groundbreaking in that it held search engine operators accountable as data controllers and obligated them to respond to individuals´ erasure requests. Since then, the RTBF has primarily been applied in the context of search engines due to their role in processing and amplifying access to personal data.

However, since the adoption of the GDPR in 2018, the technological landscape has shifted dramatically and so has the way personal data is being processed or collected. Recently, individuals have been introduced to the LLMs and their use in chatbots, which have shortly become very popular. These models are trained on extensive datasets scraped from the internet, including text that may contain personal information.[3] LLMs store and process data differently from search engines´ indexing.[4] Despite this difference, both technologies raise parallel concerns about the accessibility and processing of personal data.

This evolving context of personal data usage raises a critical legal question: Should the RTBF extend to LLMs from search engines? To address this question, the present thesis adopts a

[1] Regulation (EU) 2016/679 Of The European Parliament and of The Council of 27 April 2016 On The Protection of Natural Persons with Regard to The Processing of Personal Data and On the Free Movement of Such Data, And Repealing Directive 95/46/EC (GDPR), OJ L 119, 4.5.2016, p. 1; Article 17 GDPR.
[2] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12.
[3] *Zhang et al.*, AI and Ethics, 2024, p. 2445, 2445.
[4] Ibid.

comparative approach by analysing both technologies through the lens of their functions, legal classification and implications for the RTBF. Search engines have long been scrutinised for their personal data processing practices, but LLMs´ opaque and technically complex nature creates much more challenges on protecting personal data. Their probabilistic and non-indexed outputs complicate the implementation of the RTBF.

The analysis draws on the legal treatment of search engines under the GDPR and CJEU jurisprudence as a reference point to assess how existing frameworks might be extended to LLMs. It critically evaluates the legal grounds on which data subjects might assert the RTBF in the context of LLMs, assesses the technical feasibility of enforcing such requests and proposes potential solutions to uphold the RTBF in the age of generative AI.

The structure of this thesis is as follows: First, it sets out how the EU data protection law has evolved, with particular focus on the emergence of the GDPR and the codification of the RTBF. Second, it analyses the technical and legal nature of the search engines, highlighting their role in the application of the RTBF. Third, it introduces LLMs, provides a comparative analysis with the search engines to identify relevant similarities and differences in the RTBF context and outlines on what grounds the RTBF can be activated for LLMs. Fourth, it addresses the challenges of applying the RTBF to LLMs, including technical limitations and regulatory gaps. Finally, the thesis offers proposals for adapting legal and technical measures to ensure the effective implementation of the RTBF in the age of AI.

## B. The Right to be forgotten: legal foundations and developments

This chapter outlines the legal framework of the RTBF under the European data protection law. It traces the development of the RTBF from its origins in early privacy instruments and the Data Protection Directive 95/46/EC (Directive) through its landmark judicial recognition in *Google Spain* case to its formal codification in Article 17 of the GDPR.[5] Understanding this evolution provides the necessary basis for later analysing the applicability and enforceability of the RTBF in emerging technological contexts, particularly LLMs.

### I. Evolution of Data Protection in the EU

The right to privacy is recognized as a fundamental human right in Europe, most notably enshrined in Article 8 of the European Convention on Human Rights (ECHR), which guarantees

[5] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12; Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31.

respect for private and family life. However, the rise of digital technologies and the increasing flow of information across borders, proven that general privacy protections are insufficient to address the challenges of widespread data processing.

The first legally binding international instrument on data protection was the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, which was adopted by the Council of Europe.[6] It aimed to safeguard individuals against data misuse and established fundamental principles.

Building on these foundations, the European Commission adopted the Directive in 1995 to harmonise data protection laws across Member States while ensuring the free flow of personal data within the internal market. Although the Directive represented significant progress, when Member States implemented it into their national laws in a divergent way, it resulted in a fragmented and inconsistent data protection framework across the EU.

The European Commission recognised these challenges and proposed the GDPR in 2012 to replace the Directive with a regulation that would apply directly in Member States. While the GDPR retained many core principles from the Directive, it also introduced significant innovations, including stronger rights for data subjects, enhanced accountability obligations for data controllers, and more effective enforcement tools. One of its notable innovations was the formal codification of the RTBF in Article 17, which grants individuals the right to request the erasure of their personal data under certain conditions.

## II. Foundations of the Right to Be Forgotten

The codification of the RTBF in Article 17 of GDPR represents a major development in the field of data protection. However, it did not originate entirely with the GDPR.[7] It actually has deep historical and legal roots, shaped by growing societal concerns about individuals´ ability to regain control over their personal information in a digital age where their past becomes increasingly difficult to leave behind.

The groundwork of the RTBF was laid by Article 12 of the Directive, which provided a limited right to rectification and erasure where personal data was inaccurate, incomplete, or unlawfully processed (Directive, Article 12(b)). However, at the time of the Directive's adoption, the

[6] Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108, Strasbourg, 28 January 1981, as amended by the Protocol of 10 October 2018 (not yet in force), CETS No. 223.

[7] *Reding*, The EU Data Protection Reform 2012: Making Europe the Standard Setter for Modern Data Protection Rules in the Digital Age, SPEECH/12/26, 22 January 2012, https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_12_26, (last accessed 10 May 2025), p.5

internet was in its infancy.[8] Yahoo had only just been created, and Google had not yet been founded.[9] The Directive was therefore not designed for the complexities of large-scale online indexing and retrieval of personal information by search engines.

By the late 1990s and early 2000s, internet usage expanded rapidly, which led to an unprecedented increase in the amount of online accessible personal information.[10] The growing dominance of search engines raised concerns about whether the existing rules of the Directive could sufficiently protect individuals from this new form of exposure. Nearly two decades later, these concerns culminated in the landmark ruling by the CJEU in the *Google Spain* case, which addressed the issue directly in the context of search engine operations.

**III. Famous Ruling: The Google Spain Case**

As personal data became increasingly available online, the EU's legal framework began to shift toward formally recognising the RTBF. A key milestone in this development occurred in 2010, when Mario Costeja González, a Spanish citizen, filed a complaint with the Spanish Data Protection Authority (AEPD). He complained that when someone searched his name on Google, the results showed two newspaper articles about a real estate auction of his home, which had been conducted to recover social security debts he had owed at the time.[11] He argued that since the proceedings had been resolved a long time ago, its continued visibility online was no longer relevant to his present life.[12] Therefore, González requested that the newspaper to remove the articles and that Google to delist the links to them.[13]

The AEPD rejected the request against the newspaper, reasoning that the publications were legally valid at the time and formed part of the public record.[14] However, it upheld the complaint against Google. It found that search engines, by organising and disseminating information, have independent obligations under the Directive to protect personal data.[15]

Google Spain SL and Google Inc. appealed the decision to the Spanish Supreme Court, which referred preliminary questions to the CJEU. In its judgment, the CJEU held that search engine

[8] *EDPS*, The History of the General Data Protection Regulation, https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en, (last accessed on10 May 2025).
[9] *Alessi*, Emory Int'l L. Rev., 2017, p. 145, 154.
[10] *Alessi*, Emory Int'l L. Rev., 2017, p. 145, 155.
[11] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para 15.
[12] Ibid.
[13] Ibid.
[14] Ibid.
[15] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para 17.

operators qualify as "data controllers" under Article 2(d) of the Directive, as they determine the purposes and means of processing personal data.[16] Furthermore, the Court found that the indexing, organising, and making information accessible through search results constitute "processing" under Article 2(b), therefore, Google must comply with the obligations applicable to data controllers under the Directive.[17]

The CJEU emphasised that search engines significantly impact individuals' privacy by making personal data widely accessible through name-based searches.[18] Although Google argued that indexing and displaying results formed part of its legitimate business model, the CJEU rejected that argument by holding that economic interests alone cannot outweigh fundamental rights to privacy and data protection.[19]

At the same time, the Court acknowledged that the delisting of information could affect the legitimate interests of internet users to access information. It therefore, established an important balancing test: when individuals request the removal of links, the operator must weigh the data subject's rights against the public's interest in accessing the information.[20] While the rights of the data subject will generally prevail, this balance may vary depending on the nature of the information such as its sensitivity for the individual's private life and the interest of public to have access to it, particularly where the data subject plays a role in public life. Thus, the CJEU clarified that the RTBF is not an absolute right.[21] In González's case, the Court prioritized his privacy and control over his personal data over freedom of information and the search engine's economic interests because the information was outdated and did not hold any significant public interest.[22]

Finally, the CJEU decided that search engine operators must remove links (delinking) from search results that lead to web pages containing personal data about individuals, even if the data

[16] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 32.
[17] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 33-41.
[18] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 80.
[19] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 81.
[20] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 81–85.
[21] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 81.
[22] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 97, 99.

is lawfully on those web pages, and that this obligation exists regardless of whether the data is removed from the web pages themselves.[23]

The *Google Spain* judgment established the judicial foundation for the RTBF, interpreting the Directive in light of fundamental rights and paving the way for its formal codification under Article 17 of the GDPR.

## IV. Scope of the Right to Be Forgotten

Article 17 of the GDPR introduces the "Right to Erasure", commonly referred to as the "Right to Be Forgotten". This dual terminology is a very important legal development, as the GDPR not only codifies the right to erase but also incorporates the broader concept of being forgotten. While the right to erasure obliges data controllers to delete personal data, the RTBF extends this right's effect retrospectively, allowing individuals to request the removal of data that is no longer relevant or necessary from public access.[24] In this thesis, the terms the right to be forgotten and right to erasure are used interchangeably, and it should be understood to encompass both aspects of Article 17 of the GDPR.

To analyse the applicability of the RTBF, it is first necessary to clarify the material and personal scope of the GDPR, followed by the determination under what conditions this right may be exercised. This foundational analysis is essential for the later assessment of whether and how the RTBF can be meaningfully enforced against LLMs.

### 1. Material Scope: Defining Personal Data

Under Article 4(1) of the GDPR, personal data is defined as *"any information relating to an identified or identifiable natural person"*. This broad definition has three key elements. First, "any information" includes any kind of data, regardless of its format or nature, that concerns an individual. This includes not only direct identifiers such as names, online identifier and location data but also indirect identifiers such as professional activities, public engagements, personal beliefs and online behaviours tracked by cookies.[25] Second, "relating to" requires a necessary link between the data and the individual.[26] According to the Article 29 Data Protection Working Party (WP29), data relates to a person when it concerns their identity, characteristics or behaviour.[27] Third, the term "identified or identifiable natural person" refers to individuals who

[23] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 100.
[24] *Politou, Alepis, Patsakis*, Journal of Cybersecurity, 2018, p. 1, 9.
[25] European Union Agency for Fundamental Rights (FRA), Council of Europe, pp. 88- 89.
[26] WP29, Opinion 4/2007 on the Concept of Personal Data, 2007, 01248/07/EN WP 136, p. 9.
[27] WP29, Opinion 4/2007 on the Concept of Personal Data, 2007, 01248/07/EN WP 136, p.10.

can be recognised either directly or indirectly by combining different pieces of information.[28] It is important to note that anonymised data falls outside the scope of the GDPR, provided that it can no longer be linked to an individual.[29] Accordingly, the RTBF application depends fundamentally on whether the data in question qualifies as personal data under this definition.

**2. Personal Scope: Data Subjects and Controllers**

The personal scope of the RTBF concerns both the beneficiaries of the right and the entities responsible for its implementation. Article 17 of the GDPR grants the right to be forgotten to data subjects, the natural person to whom the personal data relates (Article 4(1) GDPR). It ensures individuals' ability to control the dissemination of their information.

The responsibility to comply with the RTBF requests falls on the data controllers. Because the data controllers are the entities that determine the purposes and means of personal data processing (Article 4(7) GDPR), data controllers are the primary addressees of RTBF obligations. While this chapter provides a preliminary overview regarding the data controllers, a more detailed analysis of the legal classification of search engine operators and LLM providers as data controllers will be undertaken in the following chapters. And these analyses will provide more information on data controllers.

**3. Conditions for Exercising the RTBF**

As outlined in this thesis so far, one of the core aims of the RTBF is to enable data subjects to regain control over their personal data. However, this right is not absolute. Therefore, it cannot be invoked only because an individual finds their information undesirable or inconvenient. To ensure legal certainty, Article 17(1) of the GDPR sets out six specific grounds under which data subjects may request erasure.

The first condition applies where the personal data is no longer necessary for the purposes for which it was originally collected or processed (Article 17(1)(a) GDPR). This condition reflects the principle of purpose limitation, which mandates that data processing be specific, explicit, and legitimate (Article 5(1)(b) GDPR). And it reinforces data minimisation by preventing the retention of irrelevant or outdated information (Article 5(1)(c) GDPR).

The second ground of the RTBF is withdrawal of consent (Article 17(1)(b) GDPR). According to GDPR, personal data may only be lawfully processed where a valid legal basis exists, one of

[28] FRA, Council of Europe, p. 86; WP29, Opinion 4/2007 on the Concept of Personal Data, 2007, 01248/07/EN WP 136, pp.12-13.
[29] Recital 26 GDPR.

which is the data subject's consent. According to Article 4 of the GDPR consent must be freely given, specific, informed, and unambiguous. Upon withdrawal, the individual may request erasure, unless another valid legal basis exists for continuing the processing, such as the controller's legitimate interests (Article 6(1)(f) GDPR).

Third, the RTBF is warranted where the data subject has successfully exercised the right to object under Article 21(1) GDPR. The distinction between Article 17 and 21 lies in their focus. Under Article 17 (c) of the GDPR, when there is no lawful basis, the continued processing or storage of personal data is prohibited. Although Article 21(1) of the GDPR allows individuals to object to further processing under certain conditions the controller may retain the personal data. This provision effectively provides data controllers with discretion to assess whether processing should cease based on legitimate interests.

Fourth, erasure is required if the personal data has been unlawfully processed. Under Article 6(1) GDPR, processing is lawful only if it meets at least one of six legal bases, such as consent, contractual necessity, or compliance with a legal obligation. If no valid justification exists for the processing, it is considered unlawful and the data must be erased in accordance with 17(1)(d) of the GDPR.

The fifth ground for erasure concerns situations where deletion is necessary to comply with a legal obligation under Union or Member State law to which the controller is subject (Article 17(1)(e) GDPR).

Finally, erasure may be requested where the personal data was collected from a child in connection with the provision of information society services (Article 17(1)(f) GDPR).

As the RTBF is not absolute, Article 17(3) of the GDPR offers several exceptions which allow data controllers to retain and process data where necessary. These exceptions might be exercising the right to freedom of expression, fulfilling a legal obligation, or carrying out tasks in the public interest. This illustrates that the RTBF is subject to a balancing test, as also established in the *Google Spain* judgment.

### C. The role of search engines in the RTBF framework

As the foundation of the RTBF has been set out, its crucial to outline the role of search engines in this framework. Search engines play a central role in the information society by determining

how personal data is accessed and displayed online.[30] This has positioned them as key actors in the application of the RTBF. Yet, neither the GDPR nor earlier EU data protection directives provide a clear legal definition of search engines.[31] Instead, their classification has evolved through case law, most notably in *Google Spain*, where the CJEU recognised search engine operators as data controllers.[32] This chapter examines the legal and technical characteristics of search engines relevant to the RTBF and provides a foundation for comparison with LLMs in the following chapter.

**I. Legal Classification of Search Engines under GDPR**

The *Google Spain* case not only introduced the RTBF but also provided a legal qualification of search engine operators as data controllers. Moreover, the CJEU's interpretation of search engines under EU law has significant implications for determining their responsibilities under data protection regulations.

As the CJEU based its judgment in its famous ruling on the Directive, it is important to clarify how the Directive applies to search engines. The Directive may apply to websites that create and publish original content. Google argued that, since search engines do not produce content themselves but merely facilitate access to it, they should not be considered data controllers under the Directive.[33] However, the CJEU found that, by providing links to personal data in response to name-based search queries, search engines could negatively impact an individual's privacy.[34] The Court relied on the Directive's definitions of controller, personal data, and processing of personal data, to justify its position.[35]

Under the Directive Article 2(d), a "data controller" is defined as any natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data. "Personal data" refers to any information relating to an identified or identifiable natural person (data subject) (Article 2(b) Directive). Notably, this definition is similar to the GDPR's definition of data controller under Article 4(7). "Processing of personal data" includes a broad range of operations, such as

[30] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras.36-38.
[31] *Kerr*, Chicago Journal of International Law, 2016, p. 217, 221.
[32] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 41.
[33] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 22.
[34] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 37-38.
[35] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras.25, 26, 32.

collection, recording, organization, storage, retrieval, consultation, use, dissemination or otherwise making available, erasure of personal data, whether automatic or not (Article 2(b) Directive).

The Court held that, the operator, through its indexing programmes, collects, retrieves, records and organises the personal data, which it then stores on its servers, discloses and makes available to its users in the form of search results.[36] These operations clearly fall within the scope of the Directive's definition of processing.[37] Because the search engine operator determines how and why this data is processed, specifically how it is ranked and presented, the CJEU concluded that it acts as a data controller. Importantly, the Court emphasised that these obligations apply even if the personal data originates from content lawfully published on third-party websites.

Furthermore, the Court observed that the activity of a search engine goes beyond that of content publishers, as it has the capacity to significantly affect the fundamental rights to privacy and data protection. For this reason, search engine operators must ensure compliance with the Directive within their technical and legal capabilities. This ensures that the protections afforded to data subjects are meaningful and enforceable.[38]

The CJEU further noted that due to the activities of the search engines, their operators must meet the Directive´s requirements so that the protection of the data subjects can be ensured. This is because search engines are widely used which allow them to infringe on privacy rights more easily than website publishers.[39] They do more than passively present links, they store websites containing personal data and disseminate information to users who may not have otherwise discovered or accessed that information.[40] Because of the central role search engines play in shaping access to personal information online, the interference with privacy rights becomes more significant. This broad accessibility and ubiquity are precisely what the Directive and the CJEU's interpretation aim to address.

Before the *Google Spain* judgment, the legal and conceptual understanding of search engines was generally narrower, often limited to their technical function as tools for locating content on

[36] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 28.
[37] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 26.
[38] CJEU, Press and Information, Press Release No 70/14, Luxembourg, 13 May 2014, p. 2.
[39] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 36-38, 85.
[40] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 37-38.

the web.[41] The CJEU in *Google Spain* adopted a broader functional view of search engines than traditionally assumed, recognised their role not merely as content locators, but as active processors of personal data, which might indicate that the CJEU intended the term "search engine" to be interpreted broadly, without limiting its scope of application.[42]

This analysis under the Directive and *Google Spain* was crucial because in its subsequent case law, such as *Google LLC v CNIL or GC and Others v. CNIL*,[43] the CJEU has maintained the rationale established in *Google Spain* regarding the concept of the data controller, which continues to apply under the current GDPR framework.[44] As data controllers, search engine operators remain directly responsible for evaluating and responding to valid RTBF requests.

## II. Technical Functionality of Search Engines

From the Court´s overview and analysis, this thesis will further define search engines from a technical point of view. This chapter explores the technical foundations of search engines as in later chapters it will compare these specifications to LLMs.

A web search engine can be defined as a service that allows users to retrieve content hosted on external websites based on keywords.[45] It functions by analysing previously indexed data and delivering a list of results ranked by relevance using statistical algorithms.[46] While often associated with retrieving web pages, modern search engines can access diverse content formats, such as videos, images, and audio files.[47] This shows the extent of search engines' capability to access information. Search results typically include a title identifying the content, a link to the original content and a short quotation of the content related to the research query.[48] Their architecture can be broken down into four core processes: Crawling, Indexing, Ranking and Retrieval & Display.

---

[41] *Kerr*, Chicago Journal of International Law, 2016, pp. 219-221.; Opinion of AG Jääskinen, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos and Mario Costeja González, Case C-131/12, ECLI:EU:C:2013:424, paras. 32-35; CJEU, Judgment of 23 March 2010, Case C-236/08, Google France SARL and Google Inc. v Louis Vuitton Malletier SA, ECLI:EU:C:2010:159, para. 22.
[42] *Kerr*, Chicago Journal of International Law, 2016, p. 227.
[43] CJEU, Judgment of 24 September 2019, Case C-136/17, GC and Others v. CNIL, ECLI:EU:C:2019:773; CJEU, Judgment of 24 September 2019, Case C-507/17, Google LLC v. Commission nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:772.
[44] EDPB, Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR, Version 2.0, 2021, p. 9, points 13-14.
[45] *Grimmelmann*, Iowa Law Review, 2007, p. 1,7; *Kerr*, Chicago Journal of International Law, 2016, p. 220.
[46] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4558, 4560.
[47] *Grimmelmann*, Iowa Law Review, 2007, pp.7-8; *Xiong et al*., IEEE Trans. Serv. Comput. 2024, p. 4560.
[48] Grimmelmann, Iowa Law Review, 2007, pp.7-10.

The process begins with crawling,[49] where automated programs, often called "web crawlers" or "spiders", systematically browse the web to gather a wide variety of web resources including web pages, images, videos, and other multimedia content.[50] Crawlers copy the website content, which is then temporarily cached for indexing.[51]

“Indexing” forms the backbone of a search engine’s ability to deliver results quickly and accurately. It enables search engines to retrieve results quickly by organising crawled content into a searchable structure.[52] The index stores metadata such as page titles and short snippets generated by algorithms to preview content in response to queries. Although site owners can restrict indexing through content removal or technical exclusions, given the scale of web data, updates to the index may be delayed, and cause removed content to remain temporarily visible in search results.[53]

Once an internet user submits a query, the search engine ranks indexed results by assigning relevance scores based on query term frequency, document structure, and semantic content.[54] Algorithms like PageRank and machine learning models analyse these factors to prioritise the most relevant results.[55]

Finally, “retrieval and display” refers to how search engines present the most relevant results to users based on query interpretation and ranking. Results appear in a structured format, typically showing the title, URL, and snippet generated from indexed content.[56] Search engines also personalise results based on user history, location, and device type.[57] This layer of personalisation illustrates the active role search engines play in shaping how personal data is accessed.

The technical functionalities outlined above show how search engines organise, index, and disseminate personal information (such as names, affiliations, and other identifiers). These activities bringing them within the scope of the GDPR as they are processing personal data and supporting their legal classification as data controllers. These capabilities, combined with their

[49] *Moffat*, Harv. J.L. & Tech, 2009, p. 475, 481.
[50] Ibid.; *Xiong et al*., IEEE Trans. Serv. Comput. 2024, p.3.
[51] *Moffat*, Harv. J.L. & Tech, 2009, p. 481.
[52] Ibid.
[53] *Gürkaynak, Yılmaz, Durlu,* Computer Law & Security Review, 2013, p.40, 41.
[54] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4560.
[55] Ibid.
[56] *Gürkaynak, Yılmaz, Durlu*, Computer Law & Security Review, 2013, p.41.
[57] Ibid.

broad dissemination power, form the basis for their legal accountability under Article 17, as confirmed by the CJEU.

### III. Implementation of the RTBF by Search Engines

The *Google Spain* judgment left the application of the RTBF to search engines. They are required to assess and respond to erasure requests submitted by data subjects.[58] Since Google began accepting requests on 29 May 2014, it has received 1,783,078 delisting requests for the removal of 7,049,744 URLs, as of 8 June 2025.[59] Operators must determine whether the conditions for erasure under Article 17 GDPR are satisfied. If a delisting request is denied, the data subject may appeal to the national data protection authority or initiate legal proceedings before the national courts.[60] If either body supports the data subject's claim, the search engine is then obligated to remove the link in question. If the original decision is upheld, the content remains accessible.[61]

The judgment also triggered debate regarding the territorial scope of delisting obligations. In *Google v. CNIL,* the CJEU held that neither the Directive nor the GDPR requires global delisting.[62] However, the Court clarified that EU law does not prohibit such a practice, and the national authorities may still request global delisting, provided that such action appropriately balances the right to privacy with freedom of expression and access to information.[63] Therefore, although the Court ensured that the identification of an individual is reduced, it didn't make it impossible.

The GDPR is ambiguous regarding the extent of erasure required under the regulation. In *Google Spain* and *Google v CNIL*, the CJEU interpreted the RTBF primarily as a right to de-referencing, which is the removal of links from search engine results. Following the *Google Spain* ruling, the WP29 provided guidelines to clarify that the right to erasure only applies to delinking search results based on a person's name and that the underlying information itself

[58] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 77; Personal data removal request form: https://reportcontent.google.com/forms/rtbf, (last accessed on 29 May 2025).
[59] Requests to delist content under European privacy law, https://transparencyreport.google.com/eu-privacy/overview, (last accessed on 8 June 2025).
[60] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras.78-79.
[61] *Wechsler*, Colum JL & Soc Probs, 2015, p.135, 142.
[62] CJEU, Judgment of 24 September 2019, Case C-507/17, Google LLC v. Commission nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:772, paras. 62-72.
[63] CJEU, Judgment of 24 September 2019, Case C-507/17, Google LLC v. Commission nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:772, para 72.

won't be deleted from the search engine's index.[64] Interestingly, some scholars find this approach an elegant solution as it keeps historical records intact while making them less readily accessible.[65] However, the EDPB has emphasised that search engine providers are not exempt from the obligation to fully erase personal data in exceptional cases.[66] Furthermore, the European Network and Information Security Agency (ENISA) outlines a spectrum of interpretations, from strict erasure such as removal from all sources and storage layers, to weaker forms such as excluding data from query results or public indices.[67] Therefore, it's clear that there is a lack of clarity and guidance in terms of the application of the RTBF.

Another issue is that despite the transition from the Directive to the GDPR, the CJEU has consistently relied on the reasoning established in *Google Spain* and focused primarily on search engine operators, which limits the exploration of how this right should apply to newer technologies. This lack of guidance together with the GDPR's technology-neutral design and the ambiguity surrounding the extent of erasure, places the burden of interpretation and application of the RTBF on courts, regulators, and other competent authorities and creates uncertainty for both individuals and data controllers navigating this area of data protection law.

## D. The evolution of AI and the rise of LLMs

The evolution of AI is widely regarded as a major technological milestone. It is commonly defined as the ability of the machine to perform tasks that typically require human intelligence.[68] Although AI currently receives significant attention, the origin of AI can be traced back to Alan Turing's 1950 paper, "Computing Machinery and Intelligence" in which he shifted the question from whether machines can "think" to whether they can show human-like intelligence through their behaviour.[69] This conceptual shift laid the foundation for the field of artificial intelligence, a term coined by John McCarthy in 1956.[70]

In recent progress of AI research has accelerated due to technological advancements that enabled access to vast volumes of data. At first, research in AI was mostly focused on

[64] WP29, Guidelines on the Implementation of the Court of Justice of the European Union Judgment On "Google Spain and Inc v. AEPD and Mario Costeja González" C-131/12, 14/EN WP 225, 2014, p.2, para. 4.
[65] Gorzeman, Korenhof, Philosophy & Technology, 2017, p. 73, 89.
[66] *EDPB*, Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR, Version 2.0, 2020, p.5, para. 10.
[67] *ENISA*, The Right to Be Forgotten - Between Expectations and Practice, 2012, p. 7.
[68] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p. 5; EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p.2; *Turing*, Philosophia Mathematica, 1996, p. 256, 257.
[69] *Turing*, Mind, 1950.
[70] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p.7.

understanding machine logic, but later it has gradually evolved to prioritise the use of data. And the emergence of large datasets has become essential for the development of machine learning. In Web 2.0, technology companies gained the ability to collect, store, process, and combine vast amounts of data to train machine learning algorithms.

In the late 1980s, AI models shifted to so-called "knowledge-based systems" and the goal was to transform expert human knowledge into computer form.[71] This led to the development of machine learning, a field focused on designing algorithms that allow computers to learn from data and improve through experience.[72] This data is known as "training data", which teaches the system to make predictions or complete specific tasks.[73] These machine learning algorithms use "neural networks", computational systems inspired by the structure and functioning of the human brain.[74] These networks have layers of interconnected nodes (or neurons), and each of them receives inputs, processes them, and passes the result to the next layer.[75] Every node is assigned a weight, which determines the value of the incoming information and applies an activation function to decide whether to forward the output.[76] The node calculates a weighted sum of its inputs and processes it through the activation function. If the result meets a certain threshold, the node activates and forwards the information to subsequent layers; if not, the information flow stops at that point.[77]

The early 2000s marked the beginning of the Big Data era, enabling the large-scale collection and transmission of information. This widespread data availability prepared the foundation for deep learning which is a subfield of machine learning that uses multi-layered neural networks.[78] These networks are structured hierarchically, where each layer processes specific aspects of the input data.[79] Through this structure, deep learning analyses the input to extract necessary patterns, classify information, and generate outputs derived from its classification.[80] Unlike

[71] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p. 9.
[72] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020. p.6.
[73] Ibid.; *Chahal, Gulia*, Int. J. Innov. Technol. Explor. Eng., 2019, p. 4910, 4911.
[74] *Chahal, Gulia*, Int. J. Innov. Technol. Explor. Eng., 2019, p. 4913; EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p.13.
[75] Ibid.
[76] Ibid.; EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p.13.
[77] Ibid.
[78] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p.11.
[79] *Chahal, Gulia*, Int. J. Innov. Technol. Explor. Eng., 2019, p.4913; European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p.11.
[80] *Chahal, Gulia*, Int. J. Innov. Technol. Explor. Eng., 2019, p. 4913.

earlier systems that relied on human intervention for extracting patterns, deep learning acquires this ability autonomously from large datasets.[81] And it refines its performance over time through feedback loops.[82] It has been realised that increasing the amount of training data significantly improves the performance of AI systems.[83] As a result, the advancement of AI has been closely tied to the rise of Big Data.[84] This close connection continues to accelerate advancements in AI.

As deep learning gained prominence, the scope of AI systems extended significantly, in particular with the rise of generative AI. Generative AI refers to algorithms capable of creating new content, including text, images, or code, based on the data they were trained on.[85] Within this field, LLMs have emerged as one of the most impactful developments. As a powerful subset of generative AI, LLMs are designed to generate human-like text. They are based on deep neural network architectures and are trained on vast amounts of textual data, allowing them to learn complex linguistic patterns and contextual relationships.[86] Their ability to produce coherent, human-like responses has made them central to numerous applications, including chatbots. A well-known example is ChatGPT, developed by OpenAI.

Chatbots powered by LLMs are capable of generating synthetic content in various formats, such as text, audio, images, and video.[87] Unlike traditional rule-based systems, these models respond dynamically to user prompts by drawing on patterns internalised during training. This method of interaction, popularised by GPT-3, enables users to submit detailed prompts which the model responds by relevant, often creative, outputs.[88] To improve alignment with human expectations, newer models like InstructGPT and ChatGPT are trained using Reinforcement Learning from Human Feedback, a method that optimises responses through a reward function derived from manually generated prompt-response pairs.[89] As a result of these advancements, various LLM-driven chatbots have been released by major tech companies and research communities,

[81] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p.11,12; EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p. 8, 9.
[82] EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p. 13,14.
[83] European Commission, Delipetrev, Tsinarakii, Kostić, AI Watch Historical Evolution of Artificial Intelligence, EUR 30221 EN, 2020, p.11.; *Chahal, Gulia*, Int. J. Innov. Technol. Explor. Eng., 2019, p.4914; *Zhang et al.*, AI and Ethics, 2024, 2447.
[84] EPRS, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020, p. 16.
[85] *Chang*, Wash. J. L. Tech. & Arts, 2024, p. 23,32.
[86] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4558; *Zhang et al.*, AI and Ethics, 2024, p. 2447.
[87] *Novelli et al,* Computer Law & Security Review, 2024, p. 1, 1.
[88] *Floridi, Chiriatti,* Mind. Mach., 2020, p. 681, 684.
[89] *Zhang et al.*, AI and Ethics, 2024, p. 2448.

including Google's Gemini[90], Meta's LLaMA[91], and Anthropic's Claude[92], alongside OpenAI's ChatGPT[93]. Moreover, these models have been embedded into search engines like Microsoft's Bing[94] and GitHub's Copilot[95], which is a critical interaction with search engines for this thesis. And some have been extended with plug-in tools, such as Gemini or ChatGPT.[96] These examples also reflects the growing popularity of chatbots and how they have become easily part of individuals' lives.

The widespread availability of big data has enabled LLMs to improve the users' lives significantly. For example, they can help journalists to write an article or assist students in summarising academic material. As a result, LLMs have been adopted in fields ranging from healthcare to banking.[97] However, their growing presence has raised serious concerns, especially around ethical issues, misinformation, and privacy.[98] This resembles the growing concerns with the widespread of usage of search engines in individuals everyday life. As LLMs are trained on real-world datasets which include sensitive and personal information to build an artificial universe, they may reproduce content with personal data, which raises concerns regarding privacy and the rights of data subjects whose information may be output by them. This increased awareness heightens the need to erase personal data from their training data as well as user chat histories. However, unlike search engines, which are relatively more transparent in how they index and display content, LLMs are often referred to as "black boxes" due to the opaque nature of their internal operations.[99] This lack of transparency, combined with

[90] *Pichai, Hassabis,* Introducing Gemini: Our Largest and Most Capable AI Model, https://blog.google/technology/ai/google-gemini-ai/#sundar-note, (last accessed on 16 May 2025).
[91] Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model, https://ai.meta.com/blog/large-language-model-llama-meta-ai/, (last accessed on 16 May 2025).
[92] Meet Claude, https://www.anthropic.com/claude, (last accessed on 16 May 2025).
[93] Introducing ChatGPT, https://openai.com/index/chatgpt/, (last accessed on 16 May 2025).
[94] *Mehdi*, Reinventing Search With a New AI-Powered Bing And Edge, Your Copilot For The Web, https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/, (last accessed on 16 May 2025).
[95] AI that builds with you, https://github.com/features/copilot, (last accessed on 21 June 2025)
[96] ChatGPT Plugins, https://openai.com/index/chatgpt-plugins/, (last accessed on 16 May 2025); Meet Gemini in Chrome, https://gemini.google/overview/gemini-in-chrome/?hl=en, (last accessed on 1 June 2025).
[97] *Bhasker et.al,* Tackling Healthcare's Biggest Burdens with Generative AI, https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai#/, (last accessed on 16 May 2025); Chui et al., McKinsey & Company, 2023, pp.1, 3, 18-24.
[98] *Lorenz, Perset, Berryhill,* OECD Artificial Intelligence Papers, 2023, p.1, 13; *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, pp.1, 1-2.
[99] *Council of the European Union*, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, p.16; *UNESCO*, Global toolkit on AI and the rule of law for the judiciary, 2023, CI/DIT/2023/AIRoL/01, p. 39; *ICO*, Big Data, Artificial Intelligence, Machine Learning and Data Protection, 2017, p. 86.

the capacity of LLMs to retain and reproduce personal information creates serious concerns regarding the right to be forgotten.[100]

Therefore, applying the RTBF to LLMs introduces complex challenges. It has been argued that comparing AI memory to human memory may be flawed because while human memory can forget or discard information, AI systems may retain data within their memory, even after removal from indexes.[101] This raises concerns about whether true erasure is possible under current technical and legal conditions. Since GDPR does not provide specific guidance on this point, it leaves AI providers uncertain about how to implement erasure requests in a legally compliant way. As previously discussed, the GDPR adopts a technology-neutral approach and continues to rely on legal reasoning set out in *Google Spain*. Consequently, it becomes necessary to seek guidance by comparing LLMs to search engines, the original subjects of RTBF jurisprudence. The following chapter undertakes this comparison from a technical perspective.

## I. Comparing Search Engines and LLMs

While Google Search has long been the dominant tool for online information access, with the introduction of LLMs such as OpenAI's ChatGPT, these models are emerging as alternative platforms in the evolving landscape of knowledge acquisition.[102] This change, along with the growing popularity of LLMs, has sparked debate about whether they might eventually replace search engines,[103] or, instead, enhance their capabilities.[104] This debate arises because both LLMs and search engines share similarities as tools for online information access. However, their operations are very different. While search engines retrieve and display links to third-party webpages based on user prompts, LLMs not only generate responses from their training data but also list links to relevant web content.[105] The similarities raise questions regarding whether

[100] *Chang*, Wash. J. L. Tech. & Arts, 2024, p.32.
[101] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p. 304, 305.
[102] *Escott*, Google Search Versus ChatGPT-ChatGPT was never meant to be a search engine, https://www.bostondigital.com/insights/google-search-versus-chatgpt-chatgpt-was-never-meant-be-search-engine, (last accessed on 17 May 2025).
[103] *Grant, Metz,* A New Chat Bot Is a 'Code Red' for Google's Search Business, https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html, (last accessed on 17 May 2025); *Caramancion*, Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, https://arxiv.org/abs/2401.05761, (last accessed on 17 May 2025); *Rowlands*, Goodbye Google? People Are Increasingly Swapping Google For The Likes Of Chatgpt, According To A Major Survey – Here's Why, https://www.techradar.com/tech/people-are-increasingly-swapping-google-for-the-likes-of-chatgpt-according-to-a-major-survey-heres-why, (last accessed on 17 May 2025).
[104] Xiong et al., IEEE Trans. Serv. Comput. 2024.
[105] *Kleinman, Antoinette,* ChatGPT Can Now Access Up To Date Information, https://www.bbc.com/news/technology-66940771, (last accessed on 17 May 2025).

LLMs should be subject to the same RTBF obligations as search engines, and the distinctions raise the question of whether it is technically feasible.

Given these legal and technical concerns, this chapter undertakes a comparative analysis of LLMs and search engines, particularly in light of their relevance to the RTBF. To guide this analysis, the chapter draws on the work of Zhang et al. (2024), who identify three key similarities and three core differences between LLMs and search engines.[106] Their framework provides a valuable foundation for assessing whether the functionalities of LLMs align with or diverge from those of search engines in ways that bear significance for the RTBF.

## 1. Similarities

### a) Use of Web-Sourced Data

Both search engines and LLMs rely heavily on data sourced from the web, which they process and structure to function effectively. As explained in earlier chapters, search engines use web crawlers to systematically browse the internet and collect content for indexing.[107] They transform scraped data into inverted indexes, which map keywords to their locations in the indexed documents.[108] This allows search engines to retrieve relevant results quickly when users submit queries.[109]

Similarly, LLMs are trained on vast datasets that include large volumes of publicly available web content. For example, OpenAI's GPT models use GPTBot[110], a web crawler that collects data from a wide range of publicly accessible websites, such as Common Crawl, social media platforms, and online forums like Reddit.[111] Therefore, it can be assumed that ChatGPT´s performance relies heavily on extensive web scraping practices. Reportedly, other LLM developers also have adopted a comparable data collection approach.[112] Accordingly, it is claimed that if individuals have ever posted anything even remotely personal on the internet, there is a high probability that this data may have been included in some of the world's most

[106] *Zhang et al.*, AI and Ethics, 2024, p. 2448.
[107] *Xiong et al*., IEEE Trans. Serv. Comput. 2024, p. 4560.
[108] Ibid.
[109] Ibid.
[110] Overview of OpenAI Crawlers, https://platform.openai.com/docs/gptbot, (last accessed on 17 May 2025).
[111] *Zhang et al.*, AI and Ethics, 2024, p. 2447.
[112] *Schaul, Chen, Tiku,* Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart, https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/, (last accessed on 17 May 2025).

popular LLMs.[113] And this data forms the foundation for training their neural networks, enabling them to understand language patterns and generate coherent responses.[114]

In conclusion, both systems process and scrape large scale of online content. Search engines structure this data for indexed retrieval, whereas LLMs use this data to train deep neural networks to generate the answers. Another similarity is how both systems transform raw, unstructured web data into organised formats tailored to their specific purposes. Search engines structure data for retrieval, while LLMs use it for generative language output.

**b) Facilitating Access to Online Information**

Both search engines and LLMs are widely used to access information online. They both have the purpose of making web-based data more accessible and usable for end users.[115]

Search engines provide users with access to information by presenting direct links to the sources that match their queries. These results are dynamically retrieved from the indexed web pages and ranked based on relevance and frequency.[116] This ability of the direct retrieval of information ensures search engines excel in connecting users to the most appropriate sources.[117]

LLMs, on the other hand, generate contextualised answers based on patterns learned during training. They synthesise relevant information and present it as a coherent, often conversational, reply.[118] And similar to search engines, LLMs are also able to list links to the relevant web pages.[119]

Therefore, both systems are designed to help individuals access relevant information with ease. For instance, a user searching for "best laptops 2024" would receive a ranked list of relevant websites from a search engine within seconds, whereas an LLM, might generate a summary of top models and purchasing advice in a single response or list the links to associated web content.

**c) Integration of Their Functionalities**

---

[113] *Heikkila*, What does GPT-3 "know" about me?, https://www.technologyreview.com/2022/08/31/1058800/what-does-gpt-3-know-about-me/, (last accessed on 17 May 2025).
[114] *Zhang et al.*, AI and Ethics, 2024, pp. 2447-2448.
[115] *Caramancion*, Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, https://arxiv.org/abs/2401.05761, (last accessed on 17 May 2025), p.1.
[116] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4560.
[117] Ibid.; *Caramancion*, Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, https://arxiv.org/abs/2401.05761, (last accessed on 17 May 2025), p.1
[118] *Zhang et al.*, AI and Ethics, 2024, p. 2449.
[119] *Kleinman, Antoinette,* ChatGPT Can Now Access Up To Date Information, https://www.bbc.com/news/technology-66940771, (last accessed on 17 May 2025).

In recent years, the functionalities of search engines and LLMs have become increasingly interconnected. This convergence reflects the evolving landscape of online information retrieval and the complementary roles that LLMs and search engines now play in that process.

Search engines have begun to integrate LLMs in order to enhance user interaction. A prominent example is Microsoft Bing, which have incorporated GPT-4 into its interface under the branding "Copilot."[120] This integration enables Bing to deliver more contextually rich, conversational responses alongside conventional lists of links.[121]

At the same time, modern LLMs are increasingly embedding search engine-like capabilities to improve their access to real-time information. Unlike earlier models, which lack the ability to access real-time information through internet searches, state-of-the-art LLMs can offer up-to-date information by harnessing search engines. For example, ChatGPT can browse the Internet to provide its answers,[122] while Google's Gemini incorporates search engine functionalities to pull the latest data from the web, enhancing the relevance and accuracy of its outputs.[123]

This integration of LLMs and search engines has led to a growing convergence in their functionalities, supporting both systems to enhance each other's capabilities. LLMs are increasingly used to improve the conversational capabilities of search engines by offering more intuitive, natural-language responses to user queries. In parallel, search engines provide LLMs with access to real-time information, enabling them to deliver more accurate and up-to-date outputs.[124] Thus, the distinction between these two technologies is becoming increasingly blurred, reinforcing their shared influence in shaping how individuals access and interact with information online.[125]

**2. Differences**

**a) Predictive Generation vs. Index-Based Retrieval:**

Although both LLMs and search engines process web data, they serve different purposes. LLMs are trained to predict the next word in a sequence, which enables them to generate coherent, human-like text by learning statistical patterns in language.[126] This process is based on

[120] *Mehdi*, Reinventing Search With a New AI-Powered Bing And Edge, Your Copilot For The Web, https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/, (last accessed on 16 May 2025).
[121] *Zhang et al.*, AI and Ethics, 2024, p. 2449.
[122] *Kleinman, Antoinette,* ChatGPT Can Now Access Up To Date Information, https://www.bbc.com/news/technology-66940771, (last accessed on 17 May 2025).
[123] *Zhang et al.*, AI and Ethics, 2024, p. 2449.
[124] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, pp. 4558-4571.
[125] Ibid.
[126] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4561.

probabilistic modelling rather than direct referencing of factual data.[127] Consequently, the relationships between words in LLM-generated responses do not necessarily reflect verified or current real-world information.[128]

Search engines, are primarily built for indexing and retrieving web content.[129] They scan the internet to create an index of information, and when users input queries, they sort and present the results.[130] Their architecture is optimised to direct users to specific sources rather than generate new content.

**b) Interaction Design: Conversational vs. Query-Based:**

Another difference lies in the way users interact with these systems. Typically LLMs perform through conversational interactions. Users engage with these systems in dynamic, multi-turn dialogues where they´re able to follow up, clarify, or change their requests in natural language.[131] This aligns with LLM´s goal of imitating human-like communication through interactions by prompts.[132]

Search engines, on the other hand, use a more traditional interaction model, where users type search queries into a search box. After that, search engines returns a list of web pages ranked by relevance to the input keywords.[133] While this is highly effective for document retrieval, it offers limited interactivity and contextual adaptability compared to the dialogic nature of LLMs.

This technical comparison illustrates why ongoing debates question whether LLMs might eventually replace or merely complement search engines. While they differ in user interaction and operational design, the similarities are striking, particularly in how both systems source vast quantities of publicly available web data and serve as tools to access information. These functional overlaps suggest that LLMs, much like search engines, significantly influence the visibility and dissemination of personal information online. Given that LLMs` training datasets may include personal data, and that LLMs can amplify the accessibility of such data, their role

---

[127] *Hacker et al*., Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, p.1112, 1113.
[128] *Zhang et al.*, AI and Ethics, 2024, p. 2449; *Kleinman, Antoinette,* ChatGPT Can Now Access Up To Date Information, https://www.bbc.com/news/technology-66940771, (last accessed on 17 May 2025), p. 1.
[129] *Zhang et al.*, AI and Ethics, 2024, p. 2449; *Caramancion*, Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, https://arxiv.org/abs/2401.05761, (last accessed on 17 May 2025), p. 2.
[130] *Caramancion*, Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, https://arxiv.org/abs/2401.05761, (last accessed on 17 May 2025), p. 3.
[131] *Zhang et al.*, AI and Ethics, 2024, p. 2449.
[132] *Xiong et al*., IEEE Trans. Serv. Comput. 2024, p. 4562
[133] *Zhang et al.*, AI and Ethics, 2024, p. 2449.

closely resembles that of search engines as considered in the RTBF jurisprudence. As a result, organisations training and operating LLMs may also fall within the scope of data controllers under the GDPR. The next chapter will examine whether, and to what extent, LLMs fall under the scope of the RTBF and how their role aligns with its obligations.

## II. Applicability of the RTBF to LLMs

As stated in Chapter B/IV, assessing the applicability of the RTBF to LLMs first requires examining whether they fall within the scope of the GDPR. Accordingly, this chapter applies the material, personal, and territorial scope criteria to LLMs to assess whether the RTBF should be extended from search engines to LLMs and enforced against them.

### 1. Material Scope

#### a) Personal Data

The GDPR applies where personal data is subject to processing by automated means, or by manual means if it forms part of a filing system (Article 2 GDPR). Accordingly, first this section will assess how LLMs involve personal data across different stages of their lifecycle.

LLMs compile data through various methods, much of which includes personal data. First, as established in earlier chapters, LLMs are trained on extensive and diverse datasets that include social media, blogs, web pages, articles, and other publicly available sources.[134] Given the ubiquitous presence of personal information online, the inclusion of such data in training sets is virtually unavoidable. Notably, there have been concerns regarding the lack of transparency in the sources used by LLM developers to compile training data.[135] Developers often justify their lack of disclosure based on competition and safety.[136] Nonetheless, OpenAI openly states in its privacy notice that it processes a range of personal data for ChatGPT, including account information, communication details (such as names, contacts and messages), and users' contact data from social media.[137] Similarly, it has been confirmed that Meta also use both public and

[134] *Novelli et al,* Computer Law & Security Review, 2024, p. 1, 5; CEDPO, Generative AI: The Data Protection Implications, 2023, p. 4; Congressional Research Service, Generative Artificial Intelligence and Data Privacy: A Primer, R47569, 2023, p. 4; Council of the European Union, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, p.14.
[135] *Hardinges, Simperl, Shadbolt,* Harvard Data Science Review, 2024; OpenAI, GPT-4 Technical Report, ArXiv, https://arxiv.org/abs/2303.08774, (last accessed on 18 May 2025).
[136] *Hardinges, Simperl, Shadbolt,* Harvard Data Science Review, 2024, p.2; OpenAI, GPT-4 Technical Report, ArXiv, https://arxiv.org/abs/2303.08774, (last accessed on 18 May 2025), p.2.
[137] *Naghiyev*, Baku State University Law Review, 2024, p.1, 4; How ChatGPT and Our Language Models Are Developed, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed, (last accessed on 18 May 2025).

non-public user data collected since 2007 for its AI development.[138] The composition of widely used training datasets further illustrates this issue: Common Crawl constitutes 60% of GPT-3's training data, social media conversations make up 50% of PaLM's, and Reddit content has been extensively used by OpenAI and Google.[139]

Second, LLMs improve their operations by leveraging data gathered from user interactions and feedback.[140] For instance, companies such as OpenAI may collect users' interactions with the model for future training.[141] These inputs may easily include personal information, for example, when users request help for drafting emails and provide the details about a specific event or task which can easily include a personal detail. Even when such prompts do not explicitly include identifiable details, if the information can be linked back to the user during interaction, they may still qualify as personal data under the GDPR. This interactive nature of LLMs also makes it easier for users to share more context hence more personal information over multiple rounds of conversation. For example, some users have reported using ChatGPT for medical consultations, which may involve personal data.[142] Moreover, as advanced LLMs accept inputs beyond text, such as images, videos, or voice recordings,[143] when these inputs involve recognisable elements, such as a person's voice or facial features, they can also be qualified as personal data, since they relate to identifiable individuals.

Third, the outputs generated by LLMs might contain personal data.[144] This can occur even when the model's training data does not explicitly include personal information, since users might introduce it during prompts.[145] Researchers have demonstrated that LLMs can regenerate details such as names, phone numbers, and email addresses if this information was present during training.[146] The model can memorise and reproduce such information and this may even

[138] Noyb Urges 11 DPAs to Immediately Stop Meta's Abuse of Personal Data for AI, https://noyb.eu/en/noyb-urges-11-dpas-immediately-stop-metas-abuse-personal-data-ai, (last accessed on 18 May 2025).
[139] *Zhang et al.,* AI and Ethics, 2024, p. 2447.
[140] *Zhang et al.,* AI and Ethics, 2024, p. 2449.
[141] New Ways To Manage Your Data in Chatgpt, https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt/, (last accessed on 18 May 2025).
[142] Reardon, AI Chatbots Can Diagnose Medical Conditions at Home. How Good Are They?, https://www.scientificamerican.com/article/ai-chatbots-can-diagnose-medical-conditions-at-home-how-good-are-they/, (last accessed on 18 May 2025).
[143] ChatGPT Can Now See, Hear, and Speak, https://openai.com/index/chatgpt-can-now-see-hear-and-speak/, (last accessed on 14 May 2025); Krawczyk, Bard's Latest Update: More Features, Languages and Countries, https://blog.google/products/gemini/google-bard-new-features-update-july-2023/, (last accessed on 14 May 2025); Dall-E 3 Is Now Available in ChatGPT Plus and Enterprise, https://openai.com/blog/dall-e-3-is-now-available-in-chatgpt-plus-and-enterprise, (last accessed on 14 May 2025).
[144] Lee et al., AI and Law: The Next Generation, https://blog.genlaw.org/explainers/, (last accessed on 18 May 2025), p. 5.
[145] *Novelli et al,* Computer Law & Security Review, 2024, p.6.
[146] *Carlini et al*., USENIX Security Symposium, 2021.

occur without a direct prompt.[147] Although developers take steps to remove personal data from training datasets, such measures are not entirely effective.[148] It should be noted that such outputs may include either accurately memorised personal data or hallucinated information.[149] The latter, known as hallucination, refers to the model generating factually incorrect or misleading information that was never present in the training data.[150] Even when provided with relevant context, the LLM-based generative search engines may produce inaccurate citations or flawed conclusions due to their reliance on probabilistic modelling.[151] These hallucinations can be minor factual errors or serious misrepresentations, such as fabricated statements about public figures, [152] or deepfake-like content.[153] The recent complaint filed by Noyb to the Austrian Data Protection Authority shows the growing concern about LLM-generated misinformation that involves personal data.[154]

Lastly, some scholars argue that LLMs themselves could be considered as personal data because they are vulnerable to certain security risks.[155] One of these risks is known as an inversion attack, which involves techniques used to extract or infer personal data embedded in the model's parameters during training.[156] Closely related is the memorisation issue, where LLMs reproduce fragments of training data containing personal information, which can occur during ordinary use or as a result of such attacks.[157] These vulnerabilities support the view that LLMs themselves may fall within the scope of personal data.

In conclusion, considering the GDPR's broad definition of personal data and its applicability to information that is publicly accessible, it is reasonable to argue that the data collected and used by LLMs in these scenarios may fall within the scope of personal data as defined by the Regulation.

**b) Processing Personal Data**

[147] *Zhang et al.,* AI and Ethics, 2024, p. 2448.
[148] Ibid; Our Approach to AI Safety, https://openai.com/index/our-approach-to-ai-safety/, (last accessed on 9 June 2025).
[149] *Zhang et al.,* AI and Ethics, 2024, p. 2449.
[150] *Zhang et al.,* AI and Ethics, 2024, p. 2448.
[151] Ibid.
[152] ChatGPT and Co: Are AI-driven search engines a threat to democratic elections?, https://algorithmwatch.org/en/bing-chat-election-2023/, (last accessed on 19 May 2025).
[153] *Novelli et al,* Computer Law & Security Review, 2024, p. 7.
[154] *Noyb*, Complaint Against OpenAI, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025).
[155] *CEDPO*, Generative AI: The Data Protection Implications, 2023, pp. 11-14; *Veale, Binns, Edwards.,* Phil. Trans. R. Soc. A., 2018, pp. 1, 6-8.
[156] *Novelli et al,* Computer Law & Security Review, 2024, p.7
[157] *Novelli et al,* Computer Law & Security Review, 2024, p. 7; *Carlini et al.,* USENIX Security Symposium, 2021, p. 2637.

The material scope of the GDPR also requires personal data to be subject to processing (Article 2 GDPR). LLMs fall under this scope in several stages.

The first stage is pre-training, in which models are fed large amounts of unlabelled data to learn and understand language. This data is collected by using automated tools that systematically browse web pages, identify the relevant information and extract content for use in training, commonly known as web scraping.[158] As the personal data includes indirect or incomplete information that can lead to identification, the scraping and use of such data during the pre-training already constitutes personal data processing, even before the model is fully trained or deployed.[159] This practice closely resembles the operations of search engines, in reference to which the CJEU confirmed that automated, constant, and systematic indexing of personal data from publicly available sources constitutes processing.[160] Given the similarity of the operations conducted by LLM operators on the information that is publicly available on the internet to train LLMs and the activities the CJEU has classified as processing in the context of search engines, the collection and preparation of such data for LLM training must equally be regarded as processing under the GDPR.[161]

The second stage in which LLMs process data is fine-tuning, where a pre-trained model is refined for a specific task using a smaller, more targeted dataset obtained through web crawling.[162] Fine-tuning involves supplementing the pre-trained model with task-specific data that provides the necessary examples and context to align the model's outputs.[163] Once the model is fine-tuned and validated, it is deployed for real-world use.

The third stage of processing occurs during the generation of output. At this point, LLMs may produce content that includes personal data. If the outputs contain names and bibliographical details of real individuals, it is considered as processing personal data, whether the information is accurate or not. In some cases, individuals may be identified not only by direct references but also through contextual clues in prompts or responses, especially when combined with search engines.[164] This risk is heightened in the case of public-facing LLMs, which are more likely to

[158] *Naghiyev*, Baku State University Law Review, 2024, p. 29; *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, pp. 4-5.
[159] *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, p. 5.
[160] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para 28.
[161] Kuru, International Data Privacy Law, 2024, pp. 236, 330-331.
[162] *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, pp. 4561-4562; Council of the European Union, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, pp. 4-5.
[163] Council of the European Union, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, pp. 4-5; *Xiong et al.,* IEEE Trans. Serv. Comput. 2024, p. 4561.
[164] *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, p. 5.

produce outputs referencing identifiable persons.[165] Importantly, the individuals mentioned in an LLM's output are not always the same as those whose data appeared in the training set, even if the same name is used, because LLMs can generate names of real people or create details that seem realistic but are not taken directly from the training data.[166] However, users might still link this output to real individuals. If the generated content makes it possible to identify someone, whether it is fully accurate or partly made up, it may still be considered as personal data under the GDPR.[167]

The final stage of data processing occurs via user inputs. Personal data may be included in the contents such as prompts, questions, or uploaded files that an individual provides when interacting with an LLM. This user-input is often retained and reused to further train or improve future models.[168] OpenAI has clarified in its public FAQs that data submitted by users is used to enhance its services.[169]

In conclusion, LLMs engage in multiple stages of data handling that qualify as processing under the GDPR. This includes the collection of vast datasets, the breakdown of that data into smaller units, and its subsequent organisation in a certain way. These activities, most of which are automated, clearly meet the definition of processing under Article 4(2) GDPR. As outlined, these processing stages may include personal data and, therefore, bring LLMs under the scope of the GDPR.

**2. Personal Scope**

The personal scope of the GDPR concerns who is responsible for complying with the regulation, namely, the roles of data controllers and processors. In the case of LLMs, as highlighted by CNIL, the qualification of the developer or deploying company as a controller must be assessed on a case-by-case basis, depending on their involvement in the different stages of data processing.[170]

Legal entities that develop and deploy LLMs act as data controllers, as they determine the purposes and means of processing personal data. Companies such as OpenAI and Google clearly meet this definition with respect to the processing operations involved in establishing

[165] Ibid.
[166] Ibid.
[167] Ibid.
[168] *Zhang et al.,* AI and Ethics, 2024, p. 2447.
[169] How Your Data Is Used to Improve Model Performance, https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance, (last accessed on 19 May 2025).
[170] *CNIL*, Determining the legal qualification of AI system providers, https://www.cnil.fr/en/determining-legal-qualification-ai-system-providers , (last accessed on 19 May 2025).

the parameters for foundational training and storing of the model, since they exclusively decide the modalities of data processing, for instance, choosing to release a freely accessible LLM.[171]

During the pre-training and training phases, these developers act as sole controllers, as they define both the data to be used and the purposes for which it is processed.[172] For example, CNIL has clarified that when the provider of a chatbot uses publicly available data to train an LLM, it qualifies as a data controller due to its decision-making power over both the purposes and means of processing.[173]

In later stages, the user input and output generation, if the company uses those inputs or outputs for purposes beyond mere service provision, such as fine-tuning or improving system performance, they are still qualified to be data controllers.[174] For example, if a developer processes both user prompts and generated content for its own distinct purposes such as training the model and protecting its systems, it may be qualified as data controller.

By the outlined comparisons and analysis of the use of personal data, it is evident that LLMs rely on similar data sources as search engines, often involving publicly available content that may contain personal data. ChatGPT, for instance, at the time of its initial release, became the fastest-growing application in history,[175] placing LLMs in a comparable position to search engines in terms of their ability to make personal information more ubiquitously available and interconnected. Just like search engines, LLMs amplify the reach of personal data, increasing its discoverability across platforms. As of recent data, ChatGPT serves around 400 million weekly users.[176] Another similarity with search engines is that LLM developers also do not share the same legitimate interests as the original publishers of the data they crawl. Instead, this data is repurposed by organisations for model training and the provision of generative services. This repurposing may also result in unforeseen consequences for the original publishers.

---

[171] *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, p. 12.
[172] *CNIL*, Determining the legal qualification of AI system providers, https://www.cnil.fr/en/determining-legal-qualification-ai-system-providers , (last accessed on 19 May 2025).
[173] *CNIL*, Determining the legal qualification of AI system providers, https://www.cnil.fr/en/determining-legal-qualification-ai-system-providers , (last accessed on 19 May 2025).
[174] *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, p. 12; Scholarly discussions have proposed that during fine tuning stage and user stage, users may also be regarded as joint controllers alongside developers, due to their influence over output generation. However, since this thesis focuses exclusively on the responsibilities of LLM developers, the potential role of users will not be analysed. See *Ruschemeier,* Cambridge Forum on AI: Law and Governance, 2025, pp. 12-13.
[175] Hu, Chatgpt Sets Record for Fastest-Growing User Base - Analyst Note, https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/, (last accessed on 19 May 2025).
[176] Duarte, Number of ChatGPT Users (March 2025), https://explodingtopics.com/blog/chatgpt-users, (last accessed on 21 June 2025).

Therefore, similar to search engine operators, undoubtedly, the organisations developing and deploying LLMs should be regarded as data controllers under the GDPR.[177]

### 3. Territorial Scope

Having established the material and personal scope of the GDPR for LLMs, this section will assess its territorial scope. Under Article 3 GDPR, territorial scope is based on two main criteria. First, whether the data controller or processor has an establishment within the EU. Second, if the company is located outside the EU, whether it offers the services to individuals in the EU or monitors their behaviour, such as tracking website visits, searches, or app usage.[178] The territorial applicability of the GDPR must be assessed on a case-by-case basis, depending on the structure and operations of the deploying company. For instance, OpenAI's ChatGPT falls under the GDPR as it has an establishment in Dublin and makes its services accessible to individuals in the EU.[179]

In conclusion, deployers of LLMs are subject to the GDPR where the conditions of territorial scope are met. Article 3 ensures that the Regulation applies to both EU-based and non-EU entities that target or monitor individuals within the EU. Accordingly, the processing of personal data by LLMs falls within both the material and territorial scope of the GDPR.

### 4. Legal Grounds for Activating the RTBF for LLMs

Article 17(1) of the GDPR outlines six conditions under which data subjects may request the erasure of personal data. For LLMs, four of these grounds are potentially relevant.

The first is the withdrawal of consent. LLM developers can rely on user consent as a legal basis for data processing. For instance, in OpenAI´s privacy policy, the company refers to consent and legitimate interest as a legal ground for the processing.[180] In this case, according to Article 7(3) of the GDPR, the individuals can withdraw that consent at any time and doing so must be as easy as granting it.[181] Once consent is withdrawn, the controller is obligated to erase the personal data concerned, unless another valid legal basis under the GDPR justifies the continued processing (Article 17(1)(b) GDPR).[182] If the developer cannot demonstrate necessity for

[177] *Zhang et al.,* AI and Ethics, 2024, p. 2449.
[178] *EDPB*, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3), version 2.1, 2020, p. 14.
[179] Introducing OpenAI Dublin, https://openai.com/index/introducing-openai-dublin/, (last accessed on 20 May 2025).
[180] Europe Privacy Policy, https://openai.com/policies/eu-privacy-policy/ , (last accessed on 20 May 2025).
[181] Recital 65 GDPR; *EDPB*, Guidelines 05/2020 on Consent under Regulation 2016/679, Version 1.1, 2020, pp. 23-24.
[182] *EDPB*, Guidelines 05/2020 on Consent under Regulation 2016/679, Version 1.1, 2020, pp. 23-24.

another lawful purpose, such as fulfilling a contractual obligation, the data must be erased without undue delay.

The second ground is where data subjects exercise their right to object to data processing pursuant to Article 21(1) GDPR. However, the developer may continue the processing if it can demonstrate compelling legitimate grounds that override the interests, rights, and freedoms of the data subject, or if the processing is necessary for the establishment, exercise, or defence of legal claims.[183] This balancing test becomes particularly relevant when LLM developers rely on the legitimate interest basis, such as research purposes or freedom of expression, to justify the processing of training or user data. For instance, OpenAI refers to ChatGPT as a "low key research preview"[184] and Google describes Bard as an "experiment"[185]. Nevertheless, these labels do not negate the commercial nature of these services. Even if such processing can be partially justified for research purposes, the balancing of interests enables data subjects to override those interests where their fundamental rights and freedoms are at greater risk.

The third ground is unlawful processing, which is a compliance issue that ChatGPT found itself in.[186] While consent under Article 6(1)(a) is a possible legal basis, it is generally impractical for LLM developers to obtain valid consent from all individuals whose personal data appears in large-scale web-scraped datasets, particularly when those individuals are unknown to the developers beforehand.[187] Also, the vast scope of web-scraped data, combined with the unpredictable applications of LLMs, makes it difficult to meet the GDPR's requirements for informed, specific, and freely given consent.[188] As a result, developers can rely on the legitimate interest basis under Article 6(1)(f) GDPR, which requires a balancing of interests between the developing entity and the persons whose data are used. While socially beneficial applications and reasonable expectations on the part of the data subject may weigh in favour of the controller,[189] these conditions can rarely be fulfilled in the context of LLM training.[190] Further,

[183] *EDPB*, Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR, Version 2.0, 2020, pp. 8-9.
[184] Weise et al., Inside the A.I. Arms Race That Changed Silicon Valley Forever, https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html, (last accessed on 20 May 2025).
[185] Hsiao, Collins, Try Bard and Share Your Feedback, https://blog.google/technology/ai/try-bard/, (last accessed on 20 May 2025).
[186] Lomas, Spanish Privacy Watchdog Says It's Probing ChatGPT Too, https://techcrunch.com/2023/04/13/chatgpt-spain-gdpr/, (last accessed on 20 May 2025); Germany Launches Data Protection Inquiry over ChatGPT, https://www.thelocal.de/20230425/germany-launches-data-protection-inquiry-over-chatgpt, (last accessed on 20 May 2025).
[187] *Novelli et al,* Computer Law & Security Review, 2024, p. 6; *CEDPO*, Generative AI: The Data Protection Implications, 2023, p.10.
[188] Ibid.
[189] Recital 47 GDPR.
[190] *Novelli et al,* Computer Law & Security Review, 2024, p. 6.

the use of training data in LLMs has come under regulatory attention due to the secondary use of personal data, often scraped from the web.[191] Since this processing departs from the original collection purpose, it creates doubt on the lawfulness of such actions.[192] The prevailing uncertainty about the outcome of the balancing test and whether the interests of LLM developers can override data subject rights supports the RTBF claims to be based on unlawful processing.

The fourth ground concerns inaccurate, irrelevant or outdated data. This legal basis was rooted in the *Google Spain* judgment, where the CJEU ruled that even lawfully collected data may become subject to erasure, provided that it is inaccurate, irrelevant, or outdated in light of the purposes for which it was collected.[193] This principle is relevant to LLMs, as these models may generate outputs containing such personal data.[194] Even if the collection and processing of the training data were initially lawful, the presence of such outdated or inaccurate data may activate the RTBF to prevent further dissemination of such information.

As the legal basis for the RTBF is established, it remains necessary to balance this right against freedom of expression and access to information. As it is also outlined in the *Google Spain* ruling, while the RTBF serves as a critical mechanism for protecting personal data, it must not disproportionately restrict the public's right to access information or the freedom to express it.[195] This balancing exercise becomes especially complex when the data subject is a public figure, as the rights to freedom of expression and access to information are more likely to take precedence.[196] Moreover, extending the RTBF beyond the territorial boundaries of the EU may also raise significant concerns about its impact on the freedom of expression and the right of access to information. Therefore, a proportional approach is needed to protect personal data

[191] ICO, Joint Statement on Data Scraping and the Protection of Privacy, 2023, p.1; *CEDPO*, Generative AI: The Data Protection Implications, 2023, p.10.
[192] Seinen, Walter, van Grondelle, in: Medina et al. (eds) p. 153, 153.; WP29, Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse, 1021/00/EN WP 207, 2013, p. 19; WP29, Opinion 03/2013 on Purpose Limitation, 00569/13/EN WP 203, 2013, p. 23
[193] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 93; *EDPB*, Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR, Version 2.0, 2020, p. 7.
[194] *Council of the European Union*, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, pp. 12, 16; *Hacker et al*., Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, p. 1113, 1120.
[195] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 99; pinion of AG Jääskinen, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos and Mario Costeja González, Case C-131/12, ECLI:EU:C:2013:424, paras 132-134.
[196] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 81; *EDPB*, Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR, Version 2.0, 2020, p. 9.

without unduly restricting the free flow of information globally.[197] In the context of LLMs, this limitation might be relevant when the individual concerned is well-known. Ultimately the balancing test requires weighing the data subject's right to privacy against the informational value of the content for the public, ensuring that decisions regarding erasure are made in a context-sensitive and rights-respecting manner.

## E. Challenges in applying the RTBF on LLMs

Having established that the RTBF applies to LLMs, this chapter examines the challenges that arise when attempting to implement data erasure in these systems. As outlined in earlier chapters, the GDPR remains ambiguous regarding the extent of erasure required under the RTBF, and regulatory bodies have offered different interpretations and recommendations on its implementation, which allows for varied implementations of the RTBF, particularly in relation to the technological architecture of LLMs.[198] The technical feasibility has also been questioned for search engines where deletion may not occur as immediately or completely as expected due to a large amount of data being stored across multiple servers and in caches.[199] But in the search engine´s case, the data is treated as an object stored in databases and with machine learning this debate becomes more complicated.[200] Given the unique data processing mechanisms of these models the appropriateness of erasure techniques must be assessed on a case-by-case basis. This kind of approach recognises the practical limitations of data removal in the digital age while striving to uphold the underlying principles of data protection.

Firstly, to be able to delete data, it must first to be able to be identified. However, LLMs are different than traditional databases, where data is stored in a certain and accessible location. They process information in a distributed way across their neural networks and it is not clear which parameters represent which data.[201] Therefore, this distributed and opaque nature of LLMs makes it extremely difficult not only to erase specific personal data but even to identify its presence within the model in the first place, when a data subject exercises their right to erasure under the GDPR.[202]

---

[197] Opinion of AG Szpunar, Case 507/17 Google LLC v the CNIL, Case C-507/17, ECLI:EU:C:2019:15, 2019, para 60-61.
[198] *Hawkins et al*., in: Rannenberg/Drogkaris/Lauradoux (eds), p.20, 22; *Zhao*, Cath. U. J. L. & Tech, 2022, pp.73, 85-86; *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p.309.
[199] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p.309, 310.
[200] Ibid.
[201] *Manab*, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p.2.
[202] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, 304, 309.; *Zhao*, Cath. U. J. L. & Tech, 2022, p. 94; Manab, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p. 2.

Secondly, during training LLMs form complex interdependencies between various data points by identifying patterns across vast datasets. As a result, the deletion of specific data can potentially disrupt the model's broader functionality.[203] Similar to a human brain, efforts to eliminate one piece of information may lead to collateral loss of associated knowledge. As these connections are often not fully understood, it is difficult to determine which other records may be affected.[204] In some cases, deleting the information of one individual has unintentionally affected outputs related to others with similar names.[205] Given this entanglement, neural networks have long been described as black boxes.[206] Therefore, if these systems are not simply collections of retrievable data but rather opaque architectures that perform intelligent tasks based on complex internal representations, targeted deletion becomes especially difficult.[207] Most machine unlearning techniques still suffer from limited accuracy, which makes it difficult to meet GDPR standards.[208] Moreover, LLMs have been shown to distort stored information when learning new inputs or to hallucinate outputs.[209] These hallucinations are particularly problematic, as they are not part of the training dataset and are therefore extremely difficult to correct or remove.[210] Furthermore, the removal of individual data points may reduce the model's ability to perform at its previous level, potentially introducing overfitting, bias, or reduced interpretability.[211]

This challenge raises an additional legal concern: can the economic interests of LLM developers override data subjects' right to erasure under GDPR? In *Google Spain*, the CJEU held that data subject rights generally prevail over a controller's economic interests, except where an

[203] *Lobo* et al., IEEE Conference on Artificial Intelligence, 2023, p. 179, 180; *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p. 315.
[204] Manab, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), pp. 2 - 3.
[205] Manab, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p. 3; *Xu et al*., IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, p. 2150, 2163.
[206] *Council of the European Union*, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 2023, p. 16; NESCO, Global toolkit on AI and the rule of law for the judiciary, 2023, CI/DIT/2023/AIRoL/01, p. 39; *ICO*, Big Data, Artificial Intelligence, Machine Learning and Data Protection, 2017, p. 86.
[207] Manab, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p. 3; *Zhao*, Cath. U. J. L. & Tech, 2022, p. 95.
[208] Manab, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p. 3; *Lobo* et al., IEEE Conference on Artificial Intelligence, 2023, p.180.
[209] Ibid.
[210] *Zhang et al.,* AI and Ethics, 2024, p. 2450.
[211] *Lobo et al.,* IEEE Conference on Artificial Intelligence, 2023, p.180.

overriding public interest justifies retention.[212] Although each case must be assessed individually through a balancing test, the default position is that the rights of data subjects prevail over economic interests.[213] Developers of LLMs may invoke the exemptions in Article 17(3) GDPR, however, in the absence of such exceptions, they cannot solely rely on claims of technical impossibility or economic burden.

Thirdly, even when personal data have been effectively removed from a training dataset, this removal does not retroactively affect models that have already been trained on that data.[214] Deletion becomes meaningful if the model is subsequently retrained without the affected data. Even then, beyond the assumption that the remaining training data are sufficient to produce a new model with desirable properties such as, performance, complexity, or interpretability,[215] retraining the model may take several months.[216] For instance, LLaMA was trained over a two-month period.[217] This timeline far exceeds the notion of "undue delay" under the GDPR, which is typically interpreted as approximately one month.[218]

Finally, LLMs are capable of continuously incorporating new information, which means that even if specific personal data is removed, it may quickly re-enter the system through other sources, for example via user inputs or interactions.[219] Additionally, the model's ability to infer and generate related content based on patterns in its training data increases the likelihood that previously erased information may reappear in generated outputs.[220] This potential of reintroduction undermines the effectiveness of the RTBF under Article 17 GDPR.

In conclusion, the technical and legal barriers to the application of the RTBF to LLMs are significant. The distributed, opaque and evolving architecture of these models makes the identification and removal of specific personal data exceptionally difficult. Therefore, it is crucial to develop effective mechanisms to operationalise the RTBF in LLMs, especially given that it generally takes precedence over the controller's competing interests.

---

[212] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, paras. 81, 97.
[213] CJEU, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment of 13 May 2014, Case C-131/12, para. 81, 98.
[214] *Lobo et al.*, IEEE Conference on Artificial Intelligence, 2023, p.180.
[215] Ibid.
[216] *Zhang et al.,* AI and Ethics, 2024, p. 2450, *Hawkins et al.*, in: Rannenberg/Drogkaris/Lauradoux (eds), p.30.
[217] *Zhang et al.,* AI and Ethics, 2024, p. 2450.
[218] Wolford, Everything you need to know about the "Right to be forgotten", https://gdpr.eu/right-to-be-forgotten/, (last accessed on 24 May 2025).
[219] *Zhang et al.,* AI and Ethics, 2024, pp. 2449-2450; *Zhao*, Cath. U. J. L. & Tech, 2022, p. 94.
[220] *Zhang et al.,* AI and Ethics, 2024, pp. 2449-2450.

## F. Towards solutions: technical and regulatory responses

This chapter begins by exploring potential technical strategies to address the challenges outlined above and then turns to the efforts of supervisory authorities in supporting the interpretation and implementation of the RTBF in the context of LLMs. The aim is to provide a structured understanding of how the right to be forgotten can be meaningfully applied to the architecture and operation of LLMs.

### I. Technical Solutions

When personal data is included in LLM training, ensuring the RTBF becomes particularly challenging. While the challenges outlined above remain unresolved, ongoing research continues to investigate potential technical solutions. This section highlights emerging approaches to enable the RTBF in LLMs. However, it should be noted that these methods require further research and this section does not aim to list every solution in the field, but instead to map out a range of strategies that could contribute to a more practical application of the RTBF in LLMs.

#### 1. Privacy by Design

Before the corrective measures that can be applied after the model deployment, data controllers may embed privacy safeguards into the model before training which enhance the practical enforcement of the RTBF. The GDPR supports this proactive approach through its data protection by design principle (Article 25 GDPR), that aims to ensure that individuals' rights are protected throughout both the design and processing stages.[221] This requires the implementation of suitable technical and organisational strategies during data processing, to ensure privacy-preserving designs are integrated by default (Article 25(1) GDPR).[222] Strategies such as data minimisation, anonymisation, encryption, transparency, and pseudonymisation, are recognised as appropriate measures to implement data protection by design.[223] This section will focus particularly on anonymisation.

Anonymisation involves altering data so that the identification of a data subject is no longer possible.[224] Once data is effectively anonymised and cannot be traced back to an individual, it

[221] Recital 78 GDPR; *EDPB*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p. 4.
[222] *EDPB*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p 6; *EDPS*, Opinion of the European Data Protection Supervisor on the Data Protection Reform Package, 2012, p. 29.
[223] Recital 78 GDPR; *EDPB*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, pp. 13, 15, 21-22, 26.
[224] *AEPD-EDPS*, 10 Misunderstandings Related to Anonymisation, 2021, p. 2; *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p. 5.

no longer qualifies as personal data according to the Article 4(1) of the GDPR and therefore falls outside its scope.[225] In this sense, anonymisation can be considered as an alternative to the deletion of the personal data.[226] However, to be considered effective, anonymisation must ensure that all means reasonably and likely to be used for identification are rendered unfeasible.[227] Yet, despite advances in anonymisation methods, there is always a risk of re-identification.[228] As exemplified by the Netflix Prize dataset disclosure, even when data has been anonymised, attackers may still re-identify individuals using sophisticated techniques, such as linking anonymised data to external or background information.[229] Applying anonymisation in the context of LLMs presents further challenges, as these systems are vulnerable to security breaches that can be result in re-identification.[230] In response to these limitations, randomisation techniques have been proposed as more robust alternatives for achieving data protection by design.[231]

The randomisation technique involves altering the accuracy of data to weaken the link between the information and the data subject.[232] Within this category, differential privacy has emerged as the leading approach, offering formal mathematical guarantee for the privacy of training data.[233] It introduces controlled randomness into datasets which makes it difficult to determine whether a specific individual's information was used.[234] This method has shown considerable promise for enhancing data protection.[235] However, challenges arise when aligning differential privacy with the GDPR's anonymisation standard, as the data controller typically retains access

[225] Recital 26 GDPR; *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p. 5; *Weitzenboeck* et al., International Data Privacy Law, 2022, pp.184, 189; *WP29*, Opinion 4/2007 on the Concept of Personal Data, 2007, 01248/07/EN WP 136, p. 21.
[226] *EDPB*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p.13.
[227] *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p. 5.; *EDPB*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p.13.
[228] *Weitzenboeck* et al., International Data Privacy Law, 2022, p. 203; *AEPD-EDPS,* 10 Misunderstandings Related to Anonymisation, 2021, p. 5.
[229] *AEPD-EDPS,* 10 Misunderstandings Related to Anonymisation, 2021, p. 7; Narayanan, Shmatikov, IEEE Symposium on Security and Privacy, 2008.
[230] Kovacs, Simple Attack Allowed Extraction of ChatGPT Training Data, https://www.securityweek.com/simple-attack-allowed-extraction-of-chatgpt-training-data/, (last accessed on 24 May 2025); Chilton, The New Risks ChatGPT Poses to Cybersecurity Data, https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity, (last accessed on 24 May 2025).
[231] *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p.12.
[232] Ibid.
[233] *Zhang et al.,* AI and Ethics, 2024, p. 2450.
[234] *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p.15; *Gandhi, Jayanti,* Technology Factsheet Series, 2021, p. 2.
[235] *Gandhi, Jayanti,* Technology Factsheet Series, 2021, p. 3.

to the original dataset.[236] Moreover, although differential privacy aims to prevent re-identification, the risk of misidentification through attacks cannot be entirely eliminated.[237]

**2. Machine Unlearning**

Machine unlearning has emerged as a targeted response to the limitations of erasure in LLMs, where deleting data from the training set alone does not undo its influence on the model's outputs. To ensure the deleted data no longer influences the model's behaviour, machine unlearning involves retraining a model without the erased data.[238] Therefore, the removal of the data from the training dataset can make machine unlearning to be considered as a very useful technique for the RTBF application on LLMs. Currently, there are two main approaches to machine unlearning: exact unlearning and approximate unlearning.[239]

Exact unlearning aims to fully eliminate the influence of specific personal data by retraining the model from scratch, using algorithms designed to reverse the effect of those data points on the model's parameters.[240] This approach ensures that the model performs as if the erased data had never been included.[241] Therefore, this approach could be considered as the most strict RTBF application.

On the other hand, approximate unlearning seeks to reduce, rather than fully remove, the impact of targeted data on an already trained model.[242] This is typically achieved by adjusting internal weights or introducing new data, with the goal of weakening the influence of the erased information without full retraining.[243]

Although both approaches are promising, they come with certain problems. Exact unlearning is impractical mainly for large-scale models such as ChatGPT. It requires full retraining for each individual erasure request, which demands significant computational resources, time, and

[236] *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, p.15.
[237] *WP29*, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, pp.15-16.
[238] *Uliussen, Rui, Johansen*, Computer Law & Security Review, 2023, p. 1, 7; *Floridi*, Philosophy & Technology, 2023, p.1, 7.
[239] *Zhang et al.,* AI and Ethics, 2024, p. 2451.
[240] *Hawkins et al*., in: Rannenberg/Drogkaris/Lauradoux (eds), p. 30; *Xu et al*., IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, p. 2150.
[241] *Hawkins et al*., in: Rannenberg/Drogkaris/Lauradoux (eds), p. 30; *Floridi*, Philosophy & Technology, 2023, p.6; *Zhang et al.,* AI and Ethics, 2024, p. 2451; *Xu et al.*, IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, p. 2150.
[242] *Zhang et al.,* AI and Ethics, 2024, p. 2451; *Xu et al*., IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, p. 2150.
[243] *Zhang et al.,* AI and Ethics, 2024, p. 2451; *Uliussen, Rui, Johansen*, Computer Law & Security Review, 2023, p.8.

financial cost.[244] Even then, discrepancies may arise between the revised dataset and the model's internal representations, which may affect overall coherence and consistency.[245] Approximate unlearning, while less intensive, cannot guarantee full removal, because the model may still be influenced by the targeted personal data.[246] Moreover, approximate techniques are open to over-unlearning, which inadvertently degrades the model's overall performance.[247] Despite these challenges, machine unlearning is regarded as more closely aligned with the RTBF obligations, as it seeks to remove data from a model's memory entirely rather than merely making it inaccessible, as in the *Google Spain* judgment.[248]

Given the technical limitations discussed, there is currently no universal solution for effective data erasure in LLMs in a way that can sufficiently fulfil the RTBF requirements in Article 17 of the GDPR. While some approaches show promise, an implementable and reliable method remains elusive due to the inherent complexity and opacity of these models. As a result, widely used LLMs such as ChatGPT may continue to operate within the EU under a degree of legal uncertainty, or even non-compliance.

### II. Regulatory and Enforcement Responses to LLMs in the EU

This section examines the regulatory responses, particularly the efforts of supervisory authorities to interpret and enforce the RTBF in the context of LLMs. While the benefits of AI are widely acknowledged, EU officials such as Industry Commissioner Thierry Breton have emphasized the urgent need for a strong regulatory framework to ensure data protection and privacy.[249]

On 30 March 2023, the Italian data protection authority (DPA) issued a temporary ban on OpenAI's use of ChatGPT for processing personal data in Italy, citing a violation of the GDPR following a data breach that exposed payment-related data (such as payment address and credit card information) of ChatGPT subscribers.[250] The ban was issued under Article 58(2)(f) GDPR,

---

[244] *Uliussen, Rui, Johansen*, Computer Law & Security Review, 2023, p. 8; *Manab*, Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and the Right to be Forgotten, Arxiv, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025), p. 3; *Hawkins et al*., in: Rannenberg/Drogkaris/Lauradoux (eds), p. 30.
[245] *Uliussen, Rui, Johansen*, Computer Law & Security Review, 2023, p. 8; *Xu et al*., IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, p. 2153.
[246] *Uliussen, Rui, Johansen*, Computer Law & Security Review, 2023, pp. 8-9.
[247] *Zhang et al.,* AI and Ethics, 2024, p. 2451.
[248] *Floridi*, Philosophy & Technology, 2023, pp. 6-8.
[249] Chee, Mukherjee, ChatGPT in spotlight as EU's Breton bats for tougher AI rules, https://www.reuters.com/technology/eus-breton-warns-chatgpt-risks-ai-rules-seek-tackle-concerns-2023-02-03/, (last accessed on 25 May 2025).
[250] *Garante Per La Protezione Dei Dati Personali (Italian DPA),* Artificial Intelligence: Stop to ChatGPT by the Italian SA Personal Data Is Collected Unlawfully, No Age Verification System Is in Place for Children,

which allows temporary limitations on data processing in response to GDPR violations. OpenAI was later found to have violated several GDPR provisions, including principles of lawfulness, fairness and transparency, failure to establish a valid legal basis for processing, insufficient protection of children's data, lack of adequate information provided to data subjects, and failure to implement data protection by design.[251] It was required to address the identified issues and implement measures to enable data subject rights, including data erasure and rectification.[252] Upon demonstrating efforts to comply, such as introducing a  mechanisms to enable data erasure, OpenAI was granted permission to reinstate ChatGPT in Italy on 2 May 2023.[253] In the application of data protection rights to LLMs, this compliance can be considered as an important regulatory milestone.

Importantly, the regulatory scrutiny has extended beyond Italy. The Polish DPA (UODO) initiated its own proceedings,[254] while German DPA submitted information requests,[255] and France's CNIL published an action plan to examine LLMs' compliance with data protection law.[256] These steps represent a significant effort of EU member states that try to interpret and address the challenges LLMs pose to existing data protection frameworks.

More recently, in April 2024, the Austrian DPA received a complaint from Noyb after its founder, Max Schrems, attempted to exercise his rights of access and erasure.[257] Schrems submitted a query to ChatGPT to seek information about himself and it returned an incorrect birthdate, which was not publicly available.[258] Therefore, he requested from OpenAI, full access

---

https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english, (last accessed on 25 May 2025); *Altomani*, Italian Garante bans Chat GPT from processing personal data of Italian data subjects, https://www.dataprotectionreport.com/2023/04/italian-garante-bans-chat-gpt-from-processing-personal-data-of-italian-data-subjects/, (last accessed on 25 May 2025).

[251] *Italian DPA,* Provvedimento del 30 marzo 2023, https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870832, (last accessed on 25 May 2025); *Altomani*, Italian Garante bans Chat GPT from processing personal data of Italian data subjects, https://www.dataprotectionreport.com/2023/04/italian-garante-bans-chat-gpt-from-processing-personal-data-of-italian-data-subjects/, (last accessed on 25 May 2025).

[252] Ibid.

[253] Italian DPA, ChatGPT: OpenAI reinstates service in Italy with enhanced transparency and rights for European users and non-users, https://www.gpdp.it:443/web/guest/home/docweb/-/docweb-display/docweb/9881490, (last accessed on 25 May 2025).

[254] Urząd Ochrony Danych Osobowych, Technologia musi być zgodna z RODO, https://uodo.gov.pl/pl/138/2823, (last accessed on 25 May 2025).

[255] Landesbeauftragte für Datenschutz und Informationssicherheit NRW, Anhörung im Verfahren zur Prüfung des Dienstes ChatGPT und der zugehörigen Sprachmodelle GPT bis GPT-4, https://www.datenschutzzentrum.de/uploads/chatgpt/20230419_Request-OpenAI_ULD-Schleswig-Holstein_IZG.pdf, (last accessed on 25 May 2025).

[256] CNIL, Artificial intelligence: the action plan of the CNIL, https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil, (last accessed on 25 May 2025).

[257] *Noyb*, Complaint Against OpenAI, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025).

[258] *Noyb*, Complaint Against OpenAI, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025), para 4.

to any personal data the company retained about him and the removal of the incorrect birthdate from ChatGPT's generated responses. OpenAI responded that it lacked the technical capability to prevent the model from generating the incorrect birthdate, noting that while filters exist to limit the display of personal data, they cannot selectively erase or block specific details without affecting other related information.[259] NOYB argued that such technical challenges do not relieve OpenAI of its obligations under GDPR and thus constitute a breach.[260] This complaint represents an important development in applying the RTBF to LLMs and serves as a relevant real life example for the legal grounds that data subjects can base the RTBF requests against LLMs.

On 13 April 2023, the EDPB established a task force to coordinate investigations and share information regarding complaints against OpenAI and ChatGPT across EU member states.[261] The task force aimed to address enforcement gaps created by the fact that OpenAI did not have a formal establishment within the EU, rendering the GDPR's One-Stop-Shop mechanism inapplicable.[262] A preliminary report, stressed the importance of a lawful data processing, called for a balance between the interests of data controllers and data subjects, and recommended data subjects to be clearly informed when their data is used for model training.[263] However, the report did not offer any practical guidance that can support the application of the right to be forgotten. Instead, they request OpenAI to clarify its practices under Article 17 GDPR.[264] It is also important to point that different DPAs interpret the legality of web scraping in conflicting ways. The Dutch DPA considers that web scraping by private entities almost always violates the GDPR as it lacks a lawful basis.[265] In contrast, the French DPA permits it under conditions such as excluding content from websites that prohibit scraping and limiting collection to data explicitly made public by users.[266] These conflicting approaches causes a broader regulatory

[259] *Noyb*, Complaint Against OpenAI, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025), paras 7-8.
[260] *Noyb*, Complaint Against OpenAI, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025), para 30.
[261] *EDPB*, EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT, https://www.edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en, (last accessed on 25 May 2025).
[262] *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, 2024, p. 4, paras. 3-4.
[263] *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, 2024, paras. 13, 16, 28.
[264] *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, 2024, Section V(c), p. 13.
[265] Bodewits, Dutch DPA issues guidelines on data scraping, https://www.hoganlovells.com/en/publications/dutch-dpa-issues-guidelines-on-data-scraping_1, (last accessed on 25 May 2025).
[266] CNIL, La réutilisation des données publiquement accessibles en ligne à des fins de démarchage commercial, https://www.cnil.fr/fr/la-reutilisation-des-donnees-publiquement-accessibles-en-ligne-des-fins-de-demarchage-commercial, (last accessed on 25 May 2025).

ambiguity, leaving LLM developers without consistent compliance guidance and increasing the risk of GDPR violations.

Notably, these enforcement actions have prompted developers like OpenAI to take some compliance steps, for example, OpenAI published a privacy notice and offered opt-out tools that allow data subjects to prevent ChatGPT from using their interactions for future training purposes.[267] Although these are important steps forward, the enforcement actions remain insufficient to ensure the effective exercise of data subject rights such as the RTBF. As outlined, supervisory authorities are increasingly acting as de facto AI regulators, yet operate within a fragmented landscape. Inconsistency on lawful processing combined with unclear RTBF obligations and the absence of technical standards for deletion in LLMs, reflects a serious mismatch between the AI development and the capacity of existing legal tools. The GDPR's technology-neutral design was intended to make it adaptable to innovation,[268] but in practice its abstract provisions create uncertainty around the enforceability of the RTBF when applied to systems that are not designed to support deletion.[269] LLMs are increasingly becoming part of individuals´ lives with their own complexities and this may cause data protection rights like RTBF to become more symbolic rather than substantive. These challenges highlight the urgent need for regulatory clarity and the development of practical mechanisms to ensure meaningful enforcement of the RTBF.

## III. Proposals

As demonstrated in the previous analysis, existing regulatory tools and technical solutions are not sufficient to fully address the unique challenges posed by LLMs. These limitations especially create issues in the context of the RTBF because it clearly lacks practical enforceability when data is deeply embedded in LLMs. To preserve the RTBF's core intent in the age of generative AI, both its legal framing and technical implementation must be reconsidered.

The RTBF is designed to allow individuals to request the deletion of personal data under defined conditions. However, the GDPR lacks a defined threshold for what constitutes sufficient erasure.[270] While "removal" suggests full deletion, this is often technically

[267] Europe Privacy Policy, https://openai.com/policies/eu-privacy-policy/ , (last accessed on 20 May 2025).; Data Controls FAQ, https://help.openai.com/en/articles/7730893-data-controls-faq, (last accessed on 25 May 2025).
[268] Recital 15 GDPR.
[269] *Zhao*, Cath. U. J. L. & Tech, 2022, p.112; *EDPS*, Opinion of the European Data Protection Supervisor on the Data Protection Reform Package, 2012, para 141.
[270] *Hawkins et al*., in: Rannenberg/Drogkaris/Lauradoux (eds), p.22, 15; *EDPS*, Opinion of the European Data Protection Supervisor on the Data Protection Reform Package, 2012, para.141.

unachievable, particularly in online settings.[271] The interpretation of "forgetting" can also be misleading, as it creates unrealistic expectations about the erasure of information from the internet.[272] These outcomes become particularly unfeasible in the LLM context.

Given these constraints, adapting the RTBF to reflect the technological realities of LLMs is necessary. As demonstrated in this thesis, LLMs and search engines have similarities, but their technical designs are different. While de-listing, a well-established remedy for search engines, is feasible in that context, it is not suitable for LLMs due to the integrated nature of training data. As such, alternative solutions such as machine unlearning and differential privacy should be considered viable mechanisms for implementing the RTBF in relation to LLMs.

To effectively introduce such approaches, it is essential to have regulatory guidelines. The EDPB and national data protection authorities should issue clear interpretive guidelines to clarify how the RTBF applies to LLMs. By collaborating with technical experts and standards bodies, regulators are in a very important position to help translate legal requirements into actionable procedures. Establishing technical standards, defining shared terminology, minimum requirements and assessment procedures, could provide developers with a framework for compliance and risk mitigation.[273]

The ISO/IEC 29134:2023 is a relevant example of such a standard which offers detailed guidance on conducting privacy impact assessments (PIAs). While Article 35 GDPR already requires PIAs in certain cases, ISO/IEC 29134:2023 enhances this process by offering a structured methodology and greater technical depth.[274] Although not legally binding, such standards can complement regulation in domains where legal norms struggle to keep pace with technological complexity. Soft legislation has proven effective in other complex sectors, such as machinery and medical devices, where directives evolved from standard-based frameworks.[275] These show how soft law can bridge the gap between fast-moving innovation and slower legal reform.[276]

---

[271] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p. 309, 310; *ENISA*, The Right to Be Forgotten - Between Expectations and Practice, 2012, p. 8, 13.
[272] Koops, International Data Privacy Law, 2014, p. 250, 258; Korenhof, in: Hansen et al. (eds), 2014, p. 114, 126.
[273] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p. 312.
[274] ISO/IEC, ISO/IEC 29134:2023, Information technology – Security techniques – Guidelines for privacy impact assessment, 2nd edn, International Organization for Standardization, 2023.
[275] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast), OJ L 157, 09/06/2006, p. 24; Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, 12/07/1993, p. 1., *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p.312.
[276] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p. 312.

While the principle of privacy must remain rooted in fundamental rights, the growing influence of LLMs may require careful trade-offs in applying RTBF rules. For instance, the GDPR's "undue delay" requirement, often interpreted as one month, may be difficult to meet when dealing with complex models like LLMs.[277] Furthermore, companies often label their LLMs as research previews or experiments to justify processing under legitimate interests. Yet these products often fall well outside the boundaries of what constitutes scientific research in the conventional sense.[278] This blurring of lines between experimentation and commercialisation raises concerns on how to balance innovation with legal compliance and it underscores the need for clearer, future-proof regulatory frameworks.

Ultimately, a hybrid approach that combines the flexibility and specificity of technical standards with the authority and enforceability of hard law offers the most realistic path forward.[279] Coordinated efforts among regulators, developers and legal scholars will be essential to ensure that the right to be forgotten continues to serve its purpose in the age of generative AI.

## G. Conclusion

This thesis aims to explore whether the RTBF should be extended to LLMs. Anchored in the landmark *Google Spain* case, RTBF was originally developed in response to search engines' role in amplifying the accessibility of personal data. However, the emergence of LLMs has disrupted the boundaries of traditional data processing models, which requires a re-evaluation of the existing legal framework of the RTBF.

Through a comparative analysis of search engines and LLMs, this thesis has demonstrated that LLMs exhibit functionalities similar to those of search engines, particularly in their method of sourcing, processing and disseminating data. While the purpose and mechanics differ, the impact on individual privacy rights can be equally significant, if not greater. The legal classification of LLM developers as data controllers, their global reach and processing of personal data, all point to the need for accountability within the RTBF.

From a legal standpoint, this thesis concludes that the RTBF can and should apply to LLMs in principle. Personal data appears at every stage of LLM lifecycle, from training datasets to real-time outputs, thus invoking the material, territorial, and personal scope of the GDPR. Several legal bases under Article 17 GDPR, including unlawful processing, withdrawal of consent, and

[277] *Zhang et al.,* AI and Ethics, 2024, p. 2451.
[278] Ibid.
[279] *Villaronga, Kieseberg, Li*, Computer Law & Security Review, 2018, p.312.

data subject objection, provide a legitimate foundation for data subjects to request erasure of their personal information from these systems.

Nonetheless, there are still significant obstacles to overcome in practice. The black-box nature of LLMs, combined with issues of memorisation, hallucination and dataset opacity presents challenges to the practical enforcement of the RTBF. At the same time, ambiguous right to be forgotten standards and lack of regulatory guidance further hinder its effective application. These unresolved complexities underscore the importance of technological and legislative innovation for the meaningful implementation of the RTBF to LLMs.

In conclusion, as the digital landscape continues to evolve, so must our interpretation and application of fundamental data protection rights. Although the RTBF was conceived in the era of search engines, its core aim which is empowering individuals to reclaim control over their personal information in the digital world, remains equally relevant for the LLMs. Adapting the RTBF to LLMs by clear guidelines and adopting a hybrid approach should be essential to preserve the privacy of data subjects in the age of AI.

**Bibliography**

*Agencia Española de Protección de Datos, European Data Protection Supervisor*, 10 Misunderstandings Related to Anonymisation, 2021.

*Alessi, Stefania*, Eternal Sunshine: The Right to Be Forgotten in the European Union after the 2016 General Data Protection Regulation, in: Emory International Law Review, 1/2017, p.145 - 171.

*AlgorithmWatch*, ChatGPT and Co: Are AI-driven search engines a threat to democratic elections?, AlgorithmWatch, 5 October 2023, https://algorithmwatch.org/en/bing-chat-election-2023/, (last accessed on 19 May 2025).

*Altomani, Pietro,* Italian Garante bans Chat GPT from processing personal data of Italian data subjects, Data Protection Report, 5 April 2023, https://www.dataprotectionreport.com/2023/04/italian-garante-bans-chat-gpt-from-processing-personal-data-of-italian-data-subjects/, (last accessed on 25 May 2025).

*Anthropic*, Meet Claude, Anthropic, https://www.anthropic.com/claude, (last accessed on 16 May 2025).

*Article 29 Data Protection Working Party,* Guidelines On the Implementation of the Court of Justice of the European Union Judgment On "Google Spain and Inc V. Agencia Española De Protección De Datos (AEPD) and Mario Costeja González" C-131/12, 14/EN WP 225, 2014.

*Article 29 Data Protection Working Party*, Opinion 4/2007 on the Concept of Personal Data, 2007, 01248/07/EN WP 136.

*Article 29 Data Protection Working Party*, Opinion 03/2013 on Purpose Limitation, 00569/13/EN WP 203, 2013.

*Article 29 Data Protection Working Party*, Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse, 1021/00/EN WP 207, 2013.

*Article 29 Data Protection Working Party,* Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN, WP216, 2014.

*Bhasker, Shashank/Bruce, Damien/Lamb, Jessica, Stein, George*, Tackling Healthcare's Biggest Burdens with Generative AI, McKinsey & Company, 10 July 2023, https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai#/, (last accessed on 16 May 2025).

*Bodewits, Joke*, Dutch DPA issues guidelines on data scraping, Hogan Lovells, 13 May 2024, https://www.hoganlovells.com/en/publications/dutch-dpa-issues-guidelines-on-data-scraping_1, (last accessed on 25 May 2025).

*Caramancion, Kevin Matthe,* Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios, Arxiv, 2024, https://arxiv.org/abs/2401.05761 , (last accessed on 17 May 2025).

*Chahal, Ayushi/Gulia, Preeti*, Machine Learning and Deep Learning, in: International Journal of Innovative Technology and Exploring Engineering, 2019, p. 4910-4914.

*Chilton, Jim,* The New Risks ChatGPT Poses to Cybersecurity Data, Harward Business Review, 21 April 2023, https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity, (last accessed on 24 May 2025).

*Chui, Michael/Hazan, Eric/Roberts, Roger/Singla, Alex/Smaje, Kate/Sukharevsky, Alex/Yee, Lareina/Zemmel, Rodney*, The Economic Potential of Generative AI: The Next Productivity Frontier, in: McKinsey & Company, 2023, p.1-65.

Confederation of European Data Protection Organisations, Generative AI: The Data Protection Implications, 2023.

Congressional Research Service, Generative Artificial Intelligence and Data Privacy: A Primer, R47569, 2023.

Court of Justice of the European Union, Judgment of 24 September 2019, Case C-136/17, *GC and Others v Commission nationale de l'informatique et des libertés (CNIL),* ECLI:EU:C:2019:773.

Court of Justice of the European Union, Judgment of 24 September 2019, Case C-507/17, *Google LLC v. Commission nationale de l'informatique et des libertés (CNIL)*, ECLI:EU:C:2019:772.

Court of Justice of the European Union, Judgment of 13 May 2014, Case C-131/12, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos and Mario Costeja González*, ECLI:EU:C:2014:317.

Court of Justice of the European Union, Press and Information, Press Release No 70/14, Luxembourg, 13 May 2014.

Court of Justice of the European Union; Judgment of 23 March 2010, Case C-236/08, *Google France SARL and Google Inc. v Louis Vuitton Malletier SA,* ECLI:EU:C:2010:159.

*Commission nationale de l'informatique et des libertés*, Artificial intelligence: the action plan of the CNIL, CNIL online, 16 May 2023, https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil, (last accessed on 25 May 2025).

*Commission nationale de l'informatique et des libertés*, Determining the legal qualification of AI system providers, CNIL online, 07 June 2024, https://www.cnil.fr/en/determining-legal-qualification-ai-system-providers , (last accessed on 19 May 2025).

*Commission nationale de l'informatique et des libertés*, La réutilisation des données publiquement accessibles en ligne à des fins de démarchage commercial, CNIL online, 30 April 2020, https://www.cnil.fr/fr/la-reutilisation-des-donnees-publiquement-accessibles-en-ligne-des-fins-de-demarchage-commercial (last accessed on 25 May 2025).

*Carlini, Nicholas/Tramèr, Florian/Wallace, Eric/Jagielski, Matthew/ Herbert-Voss, Ariel/Lee, Katherine/ Roberts, Adam/Brown, Tom/Song, Dawn/Erlingsson, Úlfar/Oprea, Alina/Raffel, Colin,* Extracting Training Data from Large Language Models, in: Proceedings of the 30th USENIX Security Symposium, 2021, p. 2633-2650.

*Chang, Cheng-chi,* When Ai Remembers Too Much: Reinventing The Right to Be Forgotten for The Generative Age*,* in: Washington Journal of Law, Technology & Arts, 2024, p.22-45.

*Chee, Foo Yun/Mukherjee, Supantha,* ChatGPT in spotlight as EU's Breton bats for tougher AI rules, Reuters, 3 February 2023, https://www.reuters.com/technology/eus-breton-warns-chatgpt-risks-ai-rules-seek-tackle-concerns-2023-02-03/, (last accessed on 25 May 2025).

*Council of Europe,* Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108, Strasbourg, 28 January 1981.

*Council of Europe,* European Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4 November 1950.

*Council of the European Union*, Analysis and Research Team, ChatGPT in the Public Sector – Overhyped or Overlooked?, 24 April 2023.

Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, 12.07.1993, p. 1.

*Duarte, Fabio*, Number of ChatGPT Users (March 2025), Exploding Topics, 20 June 2025, https://explodingtopics.com/blog/chatgpt-users, (last accessed on 21 June 2025).

*European Commission, Delipetrev, Blagoj/Tsinarakii, Chrisa/Kostić, Uroš,* Historical Evolution of Artificial Intelligence, 2020.

*European Data Protection Board,* EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT, EDPB online, 13 April 2023, https://www.edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en, (last accessed on 25 May 2025).

*European Data Protection Board*, Guidelines 3/2018 on the territorial scope of the GDPR (Article 3), Version 2.1, 2020.

*European Data Protection Board*, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020.

*European Data Protection Board*, Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR, Version 2.0, 2020.

*European Data Protection Board*, Guidelines 05/2020 on Consent under Regulation 2016/679, Version 1.1, 2020.

*European Data Protection Board*, Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR, Version 2.0, 2021.

*European Data Protection Board*, Report of the work undertaken by the ChatGPT Taskforce, 2024.

*European Data Protection Supervisor,* The History of the General Data Protection Regulation, *EDPS* online, https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en, (last accessed on 10 May 2025).

*European Data Protection Supervisor*, Opinion of the European Data Protection Supervisor on the Data Protection Reform Package, 2012.

*European Parliamentary Research Service,* The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Scientific Foresight Unit, 2020.

*Escott, Jaime,* Google Search Versus ChatGPT - ChatGPT was never meant to be a search engine, Boston Digital, https://www.bostondigital.com/insights/google-search-versus-chatgpt-chatgpt-was-never-meant-be-search-engine, (last accessed on 17 May 2025).

*European Union Agency for Fundamental Rights, Council of Europe,* Handbook on European Data Protection Law, 2018 edn, Luxembourg, 24 May 2018.

*Floridi, Luciano,* Machine Unlearning: Its nature, scope, and importance for a "delete culture", in: Philosophy & Technology, 2023, p.1-12.

*Floridi, Luciano/Chiriatti, Massimo*, GPT-3: Its Nature, Scope, Limits, and Consequences, in: Minds & Machines, 2020, p. 681-694.

*Gandhi, Raina/Jayanti, Amritha,* Differential Privacy, in: Technology Factsheet Series, 2021, p. 1-12.

*Garante Per La Protezione Dei Dati Personali,* Artificial Intelligence: Stop to ChatGPT by the Italian SA Personal Data Is Collected Unlawfully, No Age Verification System Is in Place for Children, GPDP, 31 March 2023, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english, (last accessed on 25 May 2025).

*Garante Per La Protezione Dei Dati Personali,* Provvedimento del 30 marzo 2023 [9870832], GPDP, 30 March 2023, https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870832, (last accessed on 25 May 2025).

*Garante Per La Protezione Dei Dati Personali,* ChatGPT: OpenAI reinstates service in Italy with enhanced transparency and rights for European users and non-users, GPDP, 28 April 2023, https://www.gpdp.it:443/web/guest/home/docweb/-/docweb-display/docweb/9881490, (last accessed on 25 May 2025).

*GitHub*, AI that builds with you, https://github.com/features/copilot, (last accessed on 21 June 2025).

*Google*, Requests to delist content under European privacy law, https://transparencyreport.google.com/eu-privacy/overview, (last accessed on 8 June 2025).

*Google*, Personal data removal request form, https://reportcontent.google.com/forms/rtbf, (last accessed on 29 May 2025).

*Gorzeman, Ludo/Korenhof, Paulan*, Escaping the Panopticon Over Time, in: Philosophy & Technology, 2017, p.73-92.

*Google*, Meet Gemini in Chrome, https://gemini.google/overview/gemini-in-chrome/?hl=en , Google, (last accessed on 1 June 2025).

*Grant, Nico/Metz, Cade,* A New Chat Bot Is a 'Code Red' for Google's Search Business, in: The New York Times, 2022, https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html, (last accessed on 17 May 2025).

*Grimmelmann, James,* The Structure of Search Engine Law, in: Iowa Law Review, 2007, p.1-63.

*Gürkaynak, Gönenc/Yılmaz, İlay /Durlu Gürzumar, Derya,* Understanding Search Engines: A Legal Perspective on Liability in the Internet Law Vista, in: Computer Law & Security Review, 1/2013, p. 40-47.

*Hacker, Philipp/Engel, Andreas/Mauer, Marco,* Regulating ChatGPT and other Large Generative AI Models, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, p. 1112-1123.

*Hardinges, Jack/ Simperl, Elena/Shadbolt, Nigel,* We Must Fix the Lack of Transparency Around the Data Used to Train Foundation Models, in: Harvard Data Science Review, 4/2024, p. 1-5.

*Hawkins, Katie/Alhuwaish, Nora /Belguith, Sana/Vranaki, Asma/Charlesworth, Andrew*, A Decision-Making Process to Implement the 'Right to Be Forgotten' in Machine Learning, in: Rannenberg, Kai/Drogkaris, Prokopios/Lauradoux, Cédric (eds.), Privacy Technologies and Policy - 11th Annual Privacy Forum, APF 2023, Proceedings, France, 2024, p. 20-38.

*Heikkila, Melissa*, What does GPT-3 "know" about me?, MIT Technology Review, 31 August 2022, https://www.technologyreview.com/2022/08/31/1058800/what-does-gpt-3-know-about-me/, (last accessed on 17 May 2025).

*Hsiao, Sissie/Collins, Eli,* Try Bard and Share Your Feedback, Google, 21 March 2023, https://blog.google/technology/ai/try-bard/, (last accessed on 20 May 2025).

*Hu, Krystal*, Chatgpt Sets Record For Fastest-Growing User Base - Analyst Note, Reuters, 2 February 2023, https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/, (last accessed on 19 May 2025).

*Information Commissioner's Office,* Big Data, Artificial Intelligence, Machine Learning and Data Protection, 2017, 20170904, Version: 2.2.

*Information Commissioner's Office*, Joint Statement on Data Scraping and the Protection of Privacy, 2023.

*ISO/IEC,* ISO/IEC 29134:2023, Information technology – Security techniques – Guidelines for privacy impact assessment, 2nd edn, International Organization for Standardization, 2023.

*Kerr, Julia*, What is a Search Engine? The Simple Question the Court of Justice of the European Union Forgot to Ask and What It Means for the Future of the Right to be Forgotten, in: Chicago Journal of International Law, 2016, p.217-243.

*Kleinman, Zoe/Antoinette, Radford,* ChatGPT Can Now Access Up To Date Information, in: BBC News, 2023, https://www.bbc.com/news/technology-66940771, (last accessed on 17 May 2025).

*Koops, Bert-Jaap,* The Trouble with European Data Protection Law, in: International Data Privacy Law, 4/2014, p. 250-261.

*Korenhof, Paulan,* Forgetting Bits and Pieces: An Exploration of the "Right to be Forgotten" as Implementation of "Forgetting" in Online Memory Processes, in: Hansen, Marit/Hoepman, Jaap-Henk/Leenes, Ronald/Whitehouse, Diane (eds), Privacy and Identity Management for Emerging Services and Technologies, Berlin, Heidelberg, 2014, p. 114-127.

*Kovacs, Eduard,* Simple Attack Allowed Extraction of ChatGPT Training Data, SecurityWeek, 1 December 2023, https://www.securityweek.com/simple-attack-allowed-extraction-of-chatgpt-training-data/, (last accessed on 24 May 2025).

*Krawczyk, Jack,* Bard's Latest Update: More Features, Languages and Countries, Google, 13 July 2023, https://blog.google/products/gemini/google-bard-new-features-update-july-2023/, (last accessed on 14 May 2025).

*Kuru, Taner,* Lawfulness of The Mass Processing of Publicly Accessible Online Data to Train Large Language Models, in: International Data Privacy Law, 4/2024, p. 326-351.

*Landesbeauftragte für Datenschutz und Informationssicherheit NRW,* Anhörung im Verfahren zur Prüfung des Dienstes ChatGPT und der zugehörigen Sprachmodelle GPT bis GPT-4, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, 19 April 2023, https://www.datenschutzzentrum.de/uploads/chatgpt/20230419_Request-OpenAI_ULD-Schleswig-Holstein_IZG.pdf, (last accessed on 25 May 2025).

*Lee, Katherine/Cooper, A. Feder/Grimmelmann, James/Ippolito, Daphne,* AI and Law: The Next Generation, in: GenLaw Blog, 2023, https://blog.genlaw.org/explainers/, (last accessed on 18 May 2025).

*Lobo, Jesus L/Lopez, Sergio Gil/Del Ser Javier,* The Right to Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges, in: IEEE Conference on Artificial Intelligence, 2023, p. 179-180.

*Lomas, Natasha,* Spanish Privacy Watchdog Says It's Probing ChatGPT Too, TechCrunch, 13 April 2023, https://techcrunch.com/2023/04/13/chatgpt-spain-gdpr/, (last accessed on 20 May 2025).

*Lorenz, Philippe/Perset, Karine/Berryhill, Jamie,* Initial Policy Considerations for Generative Artificial Intelligence, OECD Artificial Intelligence Papers, 2023, p.1-40.

*Manab, Meem, Arafat,* Eternal Sunshine of the Mechanical Mind: The Irreconcilability of Machine Learning and The Right to Be Forgotten, Arxiv, 2024, https://arxiv.org/abs/2403.05592v1, (last accessed on 24 May 2025).

*Meta*, Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model, Meta, 24 February 2023, https://ai.meta.com/blog/large-language-model-llama-meta-ai/, (last accessed on 16 May 2025).

*Moffat, Viva R.*, Regulating Search, in: Harvard Journal of Law and Technology, 2009, p. 475-513.

*Naghiyev, Kanan*, ChatGPT from a Data Protection Perspective, in: Baku State University Law Review, 2024, p.1-34.

*Narayanan, Arvind/Shmatikov, Vitaly,* Robust De-anonymization of Large Sparse Datasets, in: IEEE Symposium on Security and Privacy, 2008, p. 111-125.

*No Author*, Germany Launches Data Protection Inquiry over ChatGPT, in: The Local Germany, 2023, https://www.thelocal.de/20230425/germany-launches-data-protection-inquiry-over-chatgpt, (last accessed on 20 May 2025).

*Novelli, Claudio/Casolari, Federico/Hacker, Philipp/Spedicato, Giorgio/Floridi, Luciano,* Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity, in: Computer Law & Security Review, 2024, p.1-16

*Noyb*, Complaint Against OpenAI, Noyb, 29 April 2024, https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf, (last accessed on 25 May 2025).

*Noyb,* Noyb Urges 11 DPAs to Immediately Stop Meta's Abuse of Personal Data for AI, Noyb, 6 June 2024, https://noyb.eu/en/noyb-urges-11-dpas-immediately-stop-metas-abuse-personal-data-ai, (last accessed on 18 May 2025).

*Mehdi, Yusuf,* Reinventing Search With a New AI-Powered Bing And Edge, Your Copilot For The Web, Microsoft, 7 February 2023, https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/, (last accessed on 16 May 2025).

*OpenAI*, ChatGPT Plugins, https://openai.com/index/chatgpt-plugins/, OpenAI, 23 March 2023 (last accessed on 16 May 2025).

*OpenAI*, Europe Privacy Policy, https://openai.com/policies/eu-privacy-policy/, (last accessed on 20 May 2025).

*OpenAI,* ChatGPT Can Now See, Hear, and Speak, https://openai.com/index/chatgpt-can-now-see-hear-and-speak/, OpenAI, 25 September 2023, (last accessed on 14 May 2025).

*OpenAI*, Dall-E 3 Is Now Available in ChatGPT Plus and Enterprise, OpenAI, 19 October 2023, https://openai.com/blog/dall-e-3-is-now-available-in-chatgpt-plus-and-enterprise, (last accessed on 14 May 2025).

*OpenAI*, Data Controls FAQ, OpenAI, https://help.openai.com/en/articles/7730893-data-controls-faq, (last accessed on 25 May 2025).

*OpenAI*, GPT-4 Technical Report, ArXiv, 2024, https://arxiv.org/abs/2303.08774, (last accessed on 18 May 2025).

*OpenAI*, How ChatGPT and Our Language Models Are Developed, OpenAI, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed, (last accessed on 18 May 2025).

*OpenAI*, How Your Data Is Used to Improve Model Performance, OpenAI, https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance, (last accessed on 19 May 2025).

*OpenAI*, Introducing ChatGPT, OpenAI, 30 November 2022, https://openai.com/index/chatgpt/, (last accessed on 16 May 2025).

*OpenAI,* Introducing OpenAI Dublin, OpenAI, 13 September 2023, https://openai.com/index/introducing-openai-dublin/, (last accessed on 20 May 2025).

*OpenAI*, New Ways To Manage Your Data in Chatgpt, OpenAI, 25 April 2023, https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt/, (last accessed on 18 May 2025).

*OpenAI*, Our Approach to AI Safety, OpenAI, 5 April 2023, https://openai.com/index/our-approach-to-ai-safety/, (last accessed on 9 June 2025).

*OpenAI*, Overview of OpenAI Crawlers, OpenAI, https://platform.openai.com/docs/gptbot, (last accessed on 17 May 2025).

*Pichai, Sundar/Hassabis, Demis,* Introducing Gemini: Our Largest and Most Capable AI Model, Google, 6 December 2023, https://blog.google/technology/ai/google-gemini-ai/#sundar-note, (last accessed on 16 May 2025).

*Politou, Eugenia/Alepis, Efthimios/Patsakis, Constantinos*, Forgetting Personal Data and Revoking Consent Under the GDPR: Challenges and Proposed Solutions, in: Journal of Cybersecurity, 2018, p.1-20.

*Reardon, Sara,* AI Chatbots Can Diagnose Medical Conditions at Home. How Good Are They?, Scientific American, 31 March 2023, https://www.scientificamerican.com/article/ai-chatbots-can-diagnose-medical-conditions-at-home-how-good-are-they/, (last accessed on 18 May 2025).

*Reding, Viviane,* The EU Data Protection Reform 2012: Making Europe the Standard Setter for Modern Data Protection Rules in the Digital Age, SPEECH/12/26, 22 January 2012, https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_12_26, (last accessed 10 May 2025).

Regulation (EU) 2016/679 Of The European Parliament and of The Council of 27 April 2016 On The Protection of Natural Persons with Regard to The Processing of Personal Data and On the Free Movement of Such Data, and Repealing Directive 95/46/EC (GDPR), OJ L 119, 4.5.2016, p. 1.

*Rowlands, Chris*, Goodbye Google? People Are Increasingly Swapping Google For The Likes Of ChatGPT, According To A Major Survey – Here's Why, TechRadar, 20 February 2025, https://www.techradar.com/tech/people-are-increasingly-swapping-google-for-the-likes-of-chatgpt-according-to-a-major-survey-heres-why, (last accessed on 17 May 2025).

*Ruschemeier, Hannah,* Generative AI and Data Protection, in: Cambridge Forum on AI: Law and Governance, 2025, p.1-16.

*Schaul, Kevin/Chen, Szu Yu/Tiku, Nitasha,* Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart, in: The Washington Post, 2023, https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/, (last accessed on 17 May 2025).

*Seinen, Wouter/Walter, Andre/van Grondelle, Sari,* Compatibility as a Mechanism for Responsible Further Processing of Personal Data, in: Medina, Manel/ Mitrakas, Andreas/Rannenberg, Kai/ Schweighofer, Erich/Tsouroulas, Nikolaos (eds.), Privacy Technologies and Policy, Cham, 2018, p. 153-171.

*The European Network and Information Security Agency*, The Right to Be Forgotten - Between Expectations and Practice, 2012.

*Turing Alan M.,* Computing Machinery and Intelligence, Mind, 236/1950, p. 433-460

*Turing, Alan M.*, Intelligent Machinery, A Heretical Theory, in: Philosophia Mathematica, 3/1996, p. 256-260.

*Uliussen, Bjørn Aslak/Rui, Jon Petter/Johansen, Dag,* Algorithms That Forget: Machine Unlearning and The Right to Erasure, in: Computer Law & Security Review, 2023; p.1-12.

*UNESCO*, Global toolkit on AI and the rule of law for the judiciary, 2023, CI/DIT/2023/AIRoL/01.

*Urząd Ochrony Danych Osobowych,* Technologia musi być zgodna z RODO, UODO, 20 September 2023, https://uodo.gov.pl/pl/138/2823, (last accessed on 25 May 2025).

*Veale, Michael/Binns, Reuben/Edwards, Lilian,* Algorithms That Remember: Model Inversion Attacks and Data Protection Law, in: Philosophical Transactions Royal Society A, 2018, p.1-15

*Villaronga, Eduard Fosch/Kieseberg, Peter/Li, Tiffany,* Humans Forget, Machines Remember: Artificial Intelligence and The Right to Be Forgotten, in: Computer Law & Security Review, 2018, p. 304, 313.

*Wechsler, Simon,* The Right to Remember: The European Convention on Human Rights and the Right to Be Forgotten, Columbia Journal of Law & Social Problems, 2015, p. 135-165.

*Weise, Karen/Metz, Cade/Grant, Nico,* Inside the A.I. Arms Race That Changed Silicon Valley Forever, in: the New York Times, 2025, https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html, (last accessed on 20 May 2025).

*Weitzenboeck, Emily M./Lison, Pierre/Cyndecka, Malgorzata/Langford, Malcolm,* The GDPR and unstructured data: is anonymization possible?, in: International Data Privacy Law, 2022, p.184-206.

*Wolford, Ben,* Everything You Need to Know About The "Right to Be Forgotten", GDPR EU, https://gdpr.eu/right-to-be-forgotten/, (last accessed on 24 May 2025).

*Xiong Haoyi/Li, Yuchen/Li, Xuhong/Du, Mengnan/Yin, Dawei/Helal, Sumi,* When Search Engine Services Meet Large Language Models: Visions and Challenges, in: IEEE Transactions on Services Computing, 2024, p. 4558-4577.

*Xu, Jie/Wu, Zihan/Wang, Cong/Jia, Xiaohua,* Machine Unlearning: Solutions and Challenges, in: IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, p. 2150-2168.

*Zhang, Dawen/Finckenberg-Broman, Pamela/Hoang Thong/Pan, Shidong/Xing Zhenchang/Staples, Mark/Xu, Xiwei,* Right to Be Forgotten In The Era Of Large Language Models: Implications, Challenges, And Solutions, in: AI and Ethics, 2024, p. 2445-2454.

*Zhao, Zeyu,* The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws, in: Catholic University Journal of Law and Technology, 1/2022, p. 73-112.