MASTER THESIS

Titel der Master Thesis / Title of the Master‘s Thesis

„Navigating EU Data Protection Law: The Challenge of the Right to Be Forgotten in AI-Driven Text Producers“

verfasst von / submitted by

Busra Bilsin

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Laws (LL.M.)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / Postgraduate programme code as it appears on the student record sheet:	UA 992 548
Universitätslehrgang lt. Studienblatt / Postgraduate programme as it appears on the student record sheet:	Europäisches und Internationales Wirtschaftsrecht / European and International Business Law
Betreut von / Supervisor:	ao. Univ.-Prof. Mag. Dr. Siegfried Fina

Abstract: The second half of the 20th century witnessed a pivotal moment with the integration of machines into daily life, catalysing the emergence of a data-driven ecosystem. This symbiotic relationship between big data and AI sparked significant advancements, notably in generative AI, elevating the intrinsic value of data. However, this exponential surge in data collection, processing, and storing to train text producers precipitated apprehensions concerning data protection, particularly the right to be forgotten, which aims at purging data from databases. The legal uncertainty surrounding the execution of this right exacerbated concerns regarding the extent to which text producers must erase data. Although certain technical solutions show promise in facilitating erasure, their efficacy alone falls short of meeting GDPR compliance standards. Crucially, comprehensive legal adaptions are imperative, ensuring that data subjects grasp the intricate interplay between legal and technical aspects of this right.

Zusammenfassung: Die zweite Hälfte des 20. Jahrhunderts erlebte mit der Integration von Maschinen in das tägliche Leben einen entscheidenden Moment, der das Entstehen eines datengesteuerten Ökosystems katalysierte. Diese symbiotische Beziehung zwischen Big Data und KI führte zu bedeutenden Fortschritten, insbesondere im Bereich der generativen KI, und steigerte den Eigenwert von Daten. Dieser exponentielle Anstieg der Datenerfassung, -verarbeitung und -speicherung zur Schulung von Texterzeugern hat jedoch zu Befürchtungen hinsichtlich des Datenschutzes geführt, insbesondere hinsichtlich des Rechts auf Vergessenwerden, das darauf abzielt, Daten aus Datenbanken zu löschen. Die Rechtsunsicherheit in Bezug auf die Ausübung dieses Rechts verschärfte die Bedenken hinsichtlich des Ausmaßes, in dem Textproduzenten Daten löschen müssen. Obwohl bestimmte technische Lösungen vielversprechend sind, um die Löschung von Daten zu erleichtern, reicht ihre Wirksamkeit allein nicht aus, um die Standards der DSGVO zu erfüllen. Entscheidend ist, dass umfassende rechtliche Anpassungen vorgenommen werden, um sicherzustellen, dass die betroffenen Personen das komplizierte Zusammenspiel zwischen rechtlichen und technischen Aspekten dieses Rechts begreifen.

Table of Contents

Navigating EU Data Protection Law: The Challenge of the Right to Be Forgotten in AI-Driven Text Producers

1. INTRODUCTION

1.1. BACKGROUND AND PROBLEM STATEMENT

The evolution of artificial intelligence (”AI”) is currently hailed as a monumental breakthrough, although its foundations can be traced back to the 1950s. Technological advancements, particularly the widespread use of the internet, have provided easy access to vast amounts of data, leading to the rapid growth of AI. While initial works in the field of AI focused on discovering the functioning of the machine, AI has now evolved to recognise the vital role of data, prompting AI development companies to facilitate data as the core component of machine learning. In the current era of the internet, commonly known as Web 2.0, tech companies have been granted the opportunity to collect, store, process, and combine vast amounts of data to train their machine-learning algorithms. This integration of AI and Web 2.0 has paved the way for transformative advancements in leveraging the potential of data.

Although the concept of AI has been known for years, the term was first coined in the 1950s and tackled with the question raised by Alan Turing: Can machines think?.[1] In various industries, human beings serve as models for work, and the development of machines presents no exception. While such works cannot replicate human traits identically, Alan Turing introduced an imitation game, referred to as the Turing test, to determine whether a machine can exhibit human intelligence. In this test, a human evaluator engages in dialogue and cannot perceive whether it is conducted by a machine or a human being.[2] Although the Turing test has faced criticism from various authors, it still holds credibility, particularly considering the intelligence displayed by AI in well-defined tasks, surpassing human capacities.[3] At the end of this test, the machine is considered to have passed the test and displayed intelligence. Consequently, though different

[1] A. M. Turing, ‘Computing Machinery and Intelligence’ (1950) 236 Mind 433.

[2] ibid 434.

[3] David Dowe and Oppy Graham, ‘The Turing Test’ Winter 2021 The Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/archives/win2021/entries/turing-test/> accessed 1 September 2023; Gregory Scopino, *Algo Bots and the Law* (CUP 2020) 20; Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (OUP 2016) 4.

interpretations of AI are suggested, the common definition is regarded as a machine that has the capability of transcribing human intelligence and behaviour.[4]

In the pursuit of building humanlike intelligence in machines, the field of AI has undergone a remarkable evolution. AI was subjected to a shift towards symbolic AI in which in the 1980s expert knowledge in computer programs was articulated in symbolic language.[5] In the ever-evolving facet of technology, where progress depends on overcoming the constraints of current systems, advancements in machine learning, a subfield of AI, have rendered the paradigm change in AI inevitable. As a result, this progression has led to a transition from symbolic AI to machine learning, equipping machines with the capacity to produce meaningful insights and solutions for intricate problems using data. Since 2011, the era of big data and developments in computational power have given rise to a pivotal development in the AI domain, which is known as deep learning, a facet of machine learning. This advancement has allowed AI systems to achieve a deeper understanding of complex problems, marking a significant milestone in the field.

Deep learning employs multiple-layered artificial neural networks to imitate the model that biological neurons operate in the human brain, forming a stack of layers in the hierarchical system in which each layer examines the different parts of the input data.[6] Through the operation of neural networks, deep learning analyses the input to extract necessary patterns to place it into a category that is relevant to the patterns and produces the output based on its classification.[7] Unlike old systems that rely on human interventions for the pattern extraction process, deep learning automatically acquires the ability to extract patterns from extensive datasets, rather than being

[4] Commission, 'AI Watch Defining Artificial Intelligence' 2020 (EUR 30117) 7; European Parliament (hereinafter "EP"), 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' June 2020 (PE 641.530); Kaplan, (n 3) 1; Chris Lewis, 'The Need for a Legal Framework to Regulate the Use of Artificial Intelligence' (2022) 47 University of Dayton Law Review 285, 290; Scopino (n 3) 19; A. M. Turing, 'Intelligent Machinery, A Heretical Theory' (1996) 4 Philosophia Mathematica 256, 257.

[5] Commission, 'Historical Evolution of Artificial Intelligence' 2020 (EUR 30221) 9; Kaplan (n 3) 23.

[6] Scopino (n 3) 35; Kaplan (n 3) 29.

[7] Lewis 'The Need for a Legal Framework to Regulate the Use of Artificial Intelligence' (n 4) 290; Jose Luis Bermudez, *Machine Learning: From Expert Systems to Deep Learning* (3rd edn, CUP 2020) 320; Ayushi Chahal and Preeti Gulia, 'Machine Learning and Deep Learning' [2019] International Journal of Innovative Technology and Exploring Engineering 4910, 4912.

explicitly programmed.[8] Deep learning systems are refined through feedback, allowing for iterative advancements within this interaction loop.[9]

This departure from earlier methods was shaped when it was noticed that increasing the volume of training data enables AI to yield robust results.[10] This recognition has underscored that the performance and the accuracy of deep learning improved in accordance with the quantity of data.[11] As big data fuels the progression of AI systems, the development of the big data concept is contingent upon the growth of AI technology, owing to the state-of-the-art computing power and storage compared to conventional applications.[12] This symbiotic relationship between big data and AI has paved the way for further developments in AI.

As deep learning gained prominence, the remit of AI systems extended to the realm of generative AI, notably in the development of sophisticated chatbots. Although the origins of text-based chatbots date back to the 1960s, primarily aiming to pass the Turing test, generative AI has revolutionised the nature of human-computer interactions through natural language processing, leveraging neural networks for the achievement of cutting-edge achievements.[13] As the enlargement of AI models enhancing the capabilities, in particular, outcomes derived from machine learning, large language models (”LLM”), a type of generative AI, have recently spearheaded ground-breaking innovations in AI.[14] Chatbots, utilising LLM can generate synthetic content across various domains, including images, sounds, videos, and text, primarily prompted by text input, employing neural network-based generation techniques. This capacity to engage in

[8] Commission, ‘Historical Evolution of Artificial Intelligence’ (n 5) 11-12; EP, ‘The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence’ (n 4) 8-9; Scopino (n 3) 36; Kaplan (n 3) 30.

[9] Scopino (n 3) 36.

[10] EC, ‘Historical Evolution of Artificial Intelligence’ (n 5) 11.

[11] ibid; Chahal and Gulia (n 7) 4914.

[12] The concept ‘big data’ encompasses three key components: volume, variety, and velocity. These three Vs enable ‘big data’ to accumulate enormous volumes of data, featuring variety of data, at a rapid velocity. See John D Kelleher and Brendan Tierney, *Data Science* (MIT Press 2018) 9.

[13] OECD, ‘AI Language Models: Technological, Socio-Economic and Policy Considerations’ (April 2023, OECD Digital Economy Papers No 352) 22; Peter Nagy and Gina Neff, ‘Talking to Bots: Symbiotic Agency and the Case of Tay’ (2016) 10 International Journal of Communication 4915, 4918.

[14] Deep Ganguli and others, ‘Predictability and Surprise in Large Generative Models’ (Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, 2022) 1747, 1748.

lifelike conversations with human users by producing text sets this type of chatbot apart from its conventional counterparts.[15]

The commoditisation of big data has enabled text producers to boost the quality of life for consumers (*e.g.* physicians could use them in organising patients' health information and students can utilise them to extract information from aggregate literature), impacting various sectors from banking to healthcare while significantly improving productivity and efficiency in creating economic value across all industries owing to their ability to analyse vast amounts of data expeditiously.[16] Despite the promising prospects this technology offers, the accompanied challenges, including ethical dilemmas, gender-biased algorithms, the spread of misinformation and disinformation, discrimination, environmental issues, and privacy infringements should not be overlooked.[17] Considering that AI harnesses data which includes sensitive and personal information to construct an artificial universe mirroring the physical world, these text producers can potentially produce outcomes with personal data, given their presence in the real world. Along with the growing awareness and significance of data protection and privacy, personal data inherent in their datasets and outcomes eased the way for data protection and privacy, *inter alia*, to emerge as a critical area of focus for text producers.[18]

Albeit the protection afforded by technology-agnostic regulations, recent developments have spurred legislators to adopt proactive and dynamic legal measures. In the EU, the first attempt to

[15] Although there are different generative AI chatbots that are able to generate units of text, video, image, and audio, in this research, the term 'text producer' refers to a chatbot that generates text-based outcomes in response to text prompts from users.

[16] Shashank Bhasker and others, '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> accessed 8 November 2023; Micheal Chui and others, 'The Economic Potential of Generative AI: The Next Productivity Frontier' (*McKinsey & Company*, 14 June 2023) 18 and 24 <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/> accessed 15 December 2023; OECD, 'Harnessing the Power of AI and Emerging Technologies' 15 November 2022 (DSTI/CDEP(2022)14/FINAL) 7.

[17] OECD, 'Initial Policy Considerations for Generative Artificial Intelligence' September 2023 (No 1) 13; Norwegian Consumer Council, 'Ghost in the Machine' (June 2023) 15; OECD, 'AI Language Models: Technological, Socio-Economic and Policy Considerations' (n 13) 26.

[18] While the terms 'privacy' and 'data protection' are often used interchangeably, it must be emphasized that they are distinguishable in context and that they do not serve as substitutes for each other, according to the EU legal framework and the jurisprudence of the European courts. See Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222.

regulate AI was initiated in 2021 with the introduction of the AI Act.[19] This Act aims to tackle both the existing and emerging challenges posed by AI technologies. However, it is worth considering that strict obligations (*e.g.* mandatory tests regarding security and robustness and transparency requirements), may potentially stifle competition in the AI market, steering market players towards anti-competitive practices and forcing some to exit the market.[20] Consequently, the effectiveness of legislative tools in addressing these issues and proper ways to balance the costs of both intervention and non-intervention is still questionable.

In this regard, the collection and use of an enormous amount of data have specifically come under the purview of data protection legislation.[21] Since the enactment of the General Data Protection Regulation ("GDPR") in 2018, a robust protective shield has been cast over personal data across the European Union ("EU").[22] As the exponential growth in data processing by these text producers has ignited intense debates on the effective implementation of the GDPR, this processing has attracted the scrutiny of data protection authorities. Following a data breach on March 20, the Italian Data Protection Authority (”DPA") suspended the operation of OpenAI's text producer, ChatGPT, issuing an interim order and launched a thorough investigation, asserting that ChatGPT was not in compliance with the GDPR.[23] The DPA held three main GDPR violations: (i) the unlawful collection of personal data for algorithm training, (ii) lack of an age verification system, and (iii) inaccurate personal data. In light of this pioneering decision, personal

[19] Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (AI Act)' COM (2021) 206 final.

[20] 'OpenAI May Leave the EU If Regulations Bite' *Reuters* (25 May 2023) <https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/> accessed 13 November 2023.

[21] Confederation of European Data Protection Organizations (hereinafter ”CEDPO"), 'Generative AI: The Data Protection Implications' (2023) CEDPO AI Working Group 7; Dawen Zhang and others, 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' (*Arxiv*, 2023) <https://arxiv.org/pdf/2307.03941.pdf> accessed 10 November 2023; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 45.

[22] 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 (General Data Protection Regulation) [2016] OJ L 119/1 (hereinafter "the GDPR").

[23] Garante Per La Protezione Dei Dati Personali (Italian Data Protection Authority) (hereinafter ”GPDP"), '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> accessed 11 November 2023.

data processed by text producers may appear in the form of training materials, generated outputs, and the model itself due to the security lapses which may cause security attacks.[24]

Although the DPA's decision may seem to shed light on the main challenges posed by text producers, several deeper issues within the scope of the GDPR remain unresolved and demand further analysis. The surge in data protection concerns in modern society has amplified the call for the erasure of personal data from training data used by text producers, as well as user chat histories.[25] This emerging discourse brings to the forefront the right to be forgotten ("RTBF"), a novel aspect of the GDPR.[26] The rise and establishment of this right find their roots in the landmark *Google Spain* case ruled by the Court of Justice of the European Union ("CJEU") where a Spanish citizen requested Google Spain to delete his association with news articles in a Spanish newspaper that appeared when searching for his name.[27] The CJEU navigated the equilibrium between the individual's RTBF and Google Spain's interests, ruling that the news articles in question should be de-listed from Google search results. Although this case seems to address the RTBF, it is crucial to note that the data at issue was erased from the search engine's active memory, not entirely deleted from its records.[28]

While the debate over the feasibility of erasing personal data from the internet is still open to dispute, the advent of sophisticated modern technologies has intensified and complicated these discussions. The complex nature of AI systems introduced unique challenges to the enforcement of this right. One can argue that the analogy between human memory and AI does not align with the reality of AI, primarily due to cost and efficiency concerns, thus, AI removes the data requested

[24] CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 14; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 45; Philipp Hacker and others, 'Regulating ChatGPT and other Large Generative AI Models' (ACM Conference on Fairness, Accountability, and Transparency, June 2023) 13.

[25] The rise in AI technology has resulted in a significant increase in request for content delisting. In recent years, Google has received millions of such requests, spanning a diverse range of content. See 'Requests to Delist Content Under European Privacy Law' (*Google*, 19 November 2023) <https://transparencyreport.google.com/eu-privacy/overview> accessed 19 November 2023.

[26] The terms 'right to be forgotten' and 'right to erasure' are commonly used interchangeably. In Article 17 of the GDPR, the 'right to be forgotten' is specifically mentioned in brackets. However, it is considered that the deletion of the data should be carried out through the process of erasure. See Paul Lambert, *The Right To Be Forgotten* (2nd edn, Bloomsbury Professional 2022) 162.

[27] Case C-131/12 *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* [2014].

[28] Herke Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 479.

for deletion from its index, rather than completely erasing it from its database, with the data in question is still retained in its memory.[29] Due to the GDPR's ambiguity on this matter, debates fiercely continue, as the GDPR leaves AI providers uncertain about how to effectively delete personal data while staying within legal boundaries.[30]

While text producers and search engines share similarities as tools for online information access, they diverge fundamentally in operation. Search engines offer webpages related to user prompts, whereas text producers not only generate answers based on training data but also list links to associated web content.[31] This distinction raises pertinent questions regarding the degree to which the GDPR stipulates that text generators delete personal data from their database and the feasibility of implementing this right to text producers. Similarly, the ambiguity of the extraterritorial scope of this right persists, as it remained unaddressed in the *Google Spain* case.[32]

This uncertainty is also derived from opaque policies of text producer providers regarding training data which includes any personal data contained within, the purposes of collecting such personal data, third parties with whom they shared it, and the pathway for users to use their data protection rights.[33] Similar to the AI Act's transparency requirements for AI systems, including text producers, the GDPR also upholds the principle of transparency, entailing information directed towards the public or data subjects to be brief, easily accessible, and readily comprehensible.[34] The DPA's decision on ChatGPT highlighted a fundamental contradiction: despite the transparency

[29] Eduardo Fosch Villaronga and others, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten' (2018) 34 Computer Law & Security Review 304.

[30] Katie Hawkins and others, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (Annual Privacy Forum, France, June 2023); Zeyu Zhao, 'The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws' (2022) 31 Catholic University Journal of Law and Technology 73, 88.

[31] While old text producers lack the ability to access real-time information through internet searches, state-of-the-art text producers can offer up-to-date information by harnessing search engines. See Zoe Kleinman and Antoinette Radford, 'ChatGPT Can Now Access Up To Date Information' *BBC* (27 September 2023) <https://www.bbc.com/news/technology-66940771> accessed 19 November 2023. Yet, it is essential to acknowledge the differences between text producers and search engines. See Zhang, 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' (n 21) 6 and 8.

[32] In the Google CNIL case, the CJEU has drawn the scope of this right's within the EU. See Case 507/17 *Google LLC v the Commission Nationale ne l'informatique et des Libertés* (CNIL) [2019].

[33] Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 20; Information Commissioner's Office (hereinafter "ICO"), 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (2017) 10.

[34] GDPR Recital 58; Philipp Hacker, 'AI Regulation in Europe: From the AI Act to Future Regulatory Challenges' in Ifeoma Ajunwa and Jeremias Adams-Prassl (eds), *Oxford Handbook of Algorithmic Governance and the Law* (OUP 2023) 4-5; Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 6 and 18.

obligations of text producers like OpenAI, there is a reluctance to disclose information, stating the protection of trade secrets and competition concerns.[35]

1.2. RESEARCH QUESTION AND SUB-QUESTIONS

Against this background, this research seeks to answer the question '*How to exercise the right to be forgotten in text producers*'. The main question is answered through the following sub-questions:

- *In what ways do text producers impact the legal aspects of the GDPR in exercising the right to be forgotten?*
- *To what extent are text producers obliged to erase personal data under the GDPR, considering their technological infrastructure and capabilities?*
- *What design principles and practices should be implemented by text producers to promote the effective use of the right to be forgotten?*

1.3. METHODOLOGY AND STRUCTURE

This study employed a doctrinal research methodology to examine the legal framework, the CJEU and the authorities' decision-making practice.[36] To explore the correlation between data protection and technology in the context of the RTBF, academic literature on these subjects was analysed.[37] The main research question was addressed through sub-questions in three separate chapters. These chapters utilized a range of sources, including EU regulations, decisions from national data protection authorities, CJEU judgments, and additional legal instruments concerning AI and data protection. Primary sources encompass EU regulations and statutory provisions, decisions made by national data protection authorities, rulings of the CJEU, and various legal tools concerning AI and data protection. Secondary sources include reports, guidelines of the European Data Protection

[35] GPDP (n 23); Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 19-20; Heike Felzmann and others, 'Transparency You Can Trust: Transparency Requirement for Artificial Intelligence Between Legal Norms and Contextual Concerns' (2019) 6 Big Data & Society 13.

[36] Terry Hutchinson, *Research and Writing in Law* (4th edn, Lawbook Co 2018) 51; Terry Hutchinson, 'Doctrinal Research: Researching the Jury' in Dawn Watkins and Mandy Burton (eds), *Research Methods in Law* (Routledge 2017) 13; Nigle Duncan and Terry Hutchinson, 'Defining and Describing What We Do: Doctrinal Legal Research' (2012) 17 Deakin Law Review 101.

[37] ibid.

Board (”EDPB”) and other EU institutions and national authorities, books, legal commentaries, and scholarly articles. The analysis was mandated through comprehensive literature searches, including legal databases and academic journals.

Chapter One analysed how text producers came to the attraction of the GDPR and what legal challenges these text producers pose in the use of the RTBF. The legal grounds and data protection principles envisaged in the GDPR were reviewed by touching upon decisions of national data protection authorities. Chapter Two investigated the degree to which the GDPR entails erasing personal data produced and used by text producers and how text producers are able to implement this right, considering their technological infrastructure and capabilities. Chapter Three examined potential solutions for the effective use of this right in text producers by presenting recommendations for future policies and research.

2. IN WHAT WAYS DO TEXT PRODUCERS IMPACT THE LEGAL ASPECTS OF THE GDPR IN EXERCISING THE RIGHT TO BE FORGOTTEN?

2.1. INTRODUCTION

This Chapter is dedicated to addressing the first sub-question: *In what ways do text producers impact the legal aspects of the GDPR in exercising the right to be forgotten?* It will delve into the methods by which text producers collect and process personal data and the main data protection concerns under the GDPR. Following this, the Chapter will analyse a structured approach towards implementing the RTBF in text producers.

2.2. HOW DO TEXT PRODUCERS COME UNDER THE SCRUTINY OF THE GDPR

Big data serves as a vital catalyst for AI, particularly in text producers, which are trained using comprehensive text data that includes books, web pages, articles, and other sources.[38] Considering their symbiotic relationship, big data necessitates advanced computing capabilities, especially AI,

[38] UNESCO, ‘Global Toolkit on AI and the Rule of Law for the Judiciary’ 2023 (CI/DIT/2023) 35.

to effectively analyse and integrate vast volumes of data.[39] Given its variable nature, big data encompasses various types of data,[40] including personal data.[41] Under the GDPR, the definition of personal data is broad, focusing more on the ability to identify a natural person, rather than the accuracy or subjectivity of the information.[42] Therefore, any information capable of identifying qualifies as personal data under the GDPR.[43] Similarly, GDPR provides a comprehensive framework for the processing of personal data, acknowledging the fundamental rights characteristic of data protection.[44] This framework goes on to exemplify processing, such as collection, use, organization, and storage, encompassing an ever-expanding array of operations which handle personal data.[45] The incorporation of personal data and its processing in the operations of text producers triggers the application of the GDPR. However, the broad definition of personal data can lead to challenges in compliance with the GDPR, especially in the era of big data, where the vast amount and complexity of data can make adherence to the GDPR more complex.[46]

[39] Mikkel Flyverbom and others, 'The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business' (2019) 58 Business and Society 3, 6.

[40] Maeve McDonagh and Moira Paterson, 'Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data' (2018) 44 Monash University Law Review 1, 3; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 14.

[41] Article 4(1) of the GDPR defines personal data as "any information relating to an identified or identifiable natural person" and explains that "an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". To decide whether information is able to identify a natural person, it is recommended to examine on a case-by-case basis. IP addresses and cookies may be regarded as personal data relating to an identifiable person. See Article 29 Data Protection Working Party (hereinafter "WP 29"), 'Opinion 1/2008 on Data Protection Issues Related to Search Engines' (4 April 2008, WP 148 00737/EN) 8-9; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (20 June 2007, 01248/07/EN WP 136) 13, 16, and 17.

[42] Lee A. Bygrave and Luca Tosoni, 'Article 4(1). Personal Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 109; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (n 41) 6.

[43] GDPR Recital 30; ibid. Given the broad definition of personal data, the WP 29 classifies cookie and IP addresses as personal data. See WP 29, 'Opinion 1/2008 on Data Protection Issues Related to Search Engines' (n 41) 6 and 7.

[44] The GDPR Article 4(2) and Recital 15; Glorin Sebastian, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (2023) 15 International Journal of Security and Privacy in Pervasive Computing 6; Kate Jones, 'AI Governance and Human Rights: Resetting the Relationship' (*Chatnam House*, January 2023) 24 <https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights> accessed 15 December 2023; Lee A. Bygrave and Luca Tosoni, 'Article 4(2). Processing' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 119.

[45] Bygrave and Tosoni, 'Article 4(2). Processing' (n 44) 120.

[46] ibid 113.

In modern technologies, the quality of outputs generated by AI models improves proportionally with the volume of data used for training.[47] Echoing this trend, AI models, in particular text producers, increasingly rely on vast datasets for training, which often results in diminished control over the content of these datasets.[48] This lack of oversight makes it challenging to ascertain whether they contain personal data. Under the influence of this trend, text producers compile data which often includes personal data, through different methods. Firstly, these producers rely heavily on extensive training data sourced from the internet.[49] This approach inherently incorporates personal data into their datasets, given the ubiquitous presence of such information online.[50] Second, they enhance their operations by leveraging data gathered from user interactions and feedback.[51] Third, the outputs generated by text producers might contain personal data.[52] This inclusion can occur even when their databases lack personal information, as users may introduce it through their prompts.[53] Lastly, the text producer model itself could be regarded as personal data due to its vulnerability to security attacks arising from security weaknesses.[54]

While it cannot be considered with certainty that text producers always contain personal information, the broad definitions of personal data and processing under the GDPR suggest a high probability of encountering personal data in the abovementioned scenarios for text producers.

[47] Ganguli, 'Predictability and Suprise in Large Generative Models' (n 14) 1748.

[48] Katherine Lee and others, 'AI and Law: The Next Generation' 2023 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4580739> accessed 24 November 2023 5.

[49] Xiaodong Wu and others, 'Unveiling Security, Privacy, and Ethical Concerns of ChatGPT' [2023] Journal of Information and Intelligence 8; Lee, 'AI and Law: The Next Generation' (n 48) 7; Sebastian, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (n 44) 4; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 8; OECD, 'AI Language Models: Technological, Socio-Economic and Policy Considerations' (n 13) 21.

[50] OECD, 'Harnessing the Power of AI and Emerging Technologies' (n 16) 7; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 14.

[51] ibid; Zhang, 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' (n 21) 5.

[52] Lee, 'AI and Law: The Next Generation' (n 48) 5.

[53] Sebastian, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (n 44) 6. New text producer models have the capability to access real-time information through internet searches. Therefore, outputs generated by these new models may import personal data from the internet, similar to the functionality of search engines. See 'All about Google Bard' (*Medium*, 10 July 2023) <https://ip-specialist.medium.com/all-about-google-bard-aae73b5534f3> accessed 24 November 2023; 'ChatGPT Can Now Browse the Internet for Updated Information' *Aljazeera* (28 September 2023) <https://www.aljazeera.com/news/2023/9/28/chatgpt-can-now-browse-the-internet-for-updated-information> accessed 24 November 2023.

[54] ibid; CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 14; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 45; Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 13.

Having established the material scope of the GDPR for text producers, the next step should be to investigate the territorial scope of the GDPR in this context. Legislators have adopted a similarly broad approach in defining this scope, underpinned by the objective of safeguarding fundamental rights.[55] The GDPR sets out three distinct conditions for its territorial applicability, among which the most relevant for text producers is the ” targeting condition”.[56] This condition involves offering goods or services to data subjects in the EU by controllers[57] or processors not established in the EU.[58] To apply this principle to controllers based outside the EU, EDPB recommends a two-stage examination.[59] The first stage investigates whether there is processing of personal data concerning data subjects in the EU. It is important to note that the material scope of the GDPR encompasses natural persons, regardless of their nationality or place of residence.[60] Given that the majority of text-producer providers on the market, companies are based in the United States (”US”), their lack of an establishment in the EU becomes particularly relevant.[61] In the second stage, evaluating the intention of controllers or processors to offer goods or services to data subjects in the EU necessitates a case-by-case analysis, considering relevant factors.[62] The provision of services in multiple languages, particularly those official in EU Member States and the preparation of policies specifically tailored to people located in the EU can be indicators of such intent.[63] Another indicator is the prompt compliance of ChatGPT with requirements set forth by the DPA, as its

[55] Dan Jerker B. Svantesson, ‘Article 3. Territorial Scope’ in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 76.
[56] The GDPR Article 3(2)(a); ibid 82; EDPB, ‘Guidelines 3/2018 on the Territorial Scope of the GDPR’ (12 November 2019) 12.
[57] The term ‘controller’ refers to the definition in Article 4(7) of the GDPR. Article 4(7) defines ‘controller’ as “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law”.
[58] EDPB, ‘Guidelines 3/2018 on the Territorial Scope of the GDPR’ (n 56) 14.
[59] ibid.
[60] GDPR Recital 14.
[61] The well-known text producers on the market, such as ChatGPT, Bard, and Llama, are based in the US. However, there are other competing text producers. See Alex York, ’12 Best ChatGPT Alternatives & Competitors in 2023’ (*ClickUp*, 18 October 2023) <https://clickup.com/blog/chatgpt-alternatives/ - 57-12-claude> accessed 24 November 2023. See also ‘Europe Terms of Use’ (*OpenAI*, 14 November 2023) <https://openai.com/policies/eu-terms-of-use> accessed 24 November 2023; ‘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/> accessed 24 November 2023.
[62] The GDPR Recital 23; Svantesson, ‘Article 3. Territorial Scope’ (n 55) 82-83; EDPB, ‘Guidelines 3/2018 on the Territorial Scope of the GDPR’ (n 56) 15-16.
[63] For instance, Bard provides services in over 40 languages, including official languages of Member States. See ‘Bard FAQ’ <https://bard.google.com/faq?hl=en> accessed 24 November 2023; ‘Europe Terms of Use’ (n 61).

swift implementation of certain measures demonstrates its intention to carry out its services within the EU.[64] Consequently, the data processing conducted by text producers fall within the material and territorial scope of the GDPR.

2.3. GDPR IN ACTION: COMPLIANCE STRATEGIES

After setting the scene for GDPR applicability, the focus shifts to determining whether the processing activities are based on a legal basis set out in Article 6 of the GDPR, a critical step in the deeper investigation of GDPR compliance. Following the Italian DPA's investigation, the Spanish data protection authority also launched a probe into OpenAI while data protection watchdogs in France and Germany commenced inquiries into OpenAI's GDPR compliance.[65] In response to the initiatives by different national data protection authorities, the EDPB established a dedicated task force on ChatGPT to strengthen cooperation and to enable the exchange of information between data protection authorities.[66] Given these developments, it becomes crucial to examine the legal grounds on a case-by-case basis, as they may vary depending on different legal justifications.

While the legal ambiguity surrounding training data, stemming from companies opaque policies, it is evident that they utilize online sources available on the internet. The rise of big data has empowered text producers to ingest large quantities of diverse content which are scraped from numerous online sources for training purposes.[67] This exploitation of data raises the question of whether such practices are grounded on a legal basis, even though data is sourced from publicly available domains.[68] Tech companies that engaged in data scraping often contend that they use

[64] Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 14.

[65] Natasha Lomas, 'Spanish Privacy Watchdog Says It's Probing ChatGPT Too' (*TechCrunch*, 13 April 2023) <https://techcrunch.com/2023/04/13/chatgpt-spain-gdpr/> accessed 5 December 2023; 'Germany Launches Data Protection Inquiry over ChatGPT' (*The Local de*, 24 April 2023) <https://www.thelocal.de/20230425/germany-launches-data-protection-inquiry-over-chatgpt> accessed 5 December 2023.

[66] 'EDPB Resolves Dispute on Transfers by Meta and Creates Task Force on ChatGPT' (EDPB, 13 April 2023) <https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en> accessed 5 December 2023.

[67] CEDPO 'Generative AI: The Data Protection Implications' (n 21) 4; Congressional Research Service, 'Generative Artificial Intelligence and Data Privacy: A Primer' (23 May 2023) 4; Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (24 April 2023) 14.

[68] GPDP (n 23). The GDPR Article 6 stipulates six legal grounds for the sake of lawful processing: (i) consent, (ii) performance of a contract, (iii) compliance with a legal obligation, (iv) vital interest, (v) performance of a public task,

personal data scraped from publicly available sources. However, since publicly accessible data maintains its personal data characteristic, various data protection authorities have examined the legality of data scraping practices.[69] Recent decisions by national data protection authorities regarding data scraping conducted by Clearview AI (”Clearview”) have established that data scraping is unlawful if it lacks a valid legal basis.[70] In the Clearview case, the company amassed a large collection of images of individuals from the internet, particularly from social media platforms, for use in their facial recognition service.[71] It is argued by Clearview that their processing activities are exempted under the provision permitting the processing of personal data made publicly available by data subjects.[72] Nevertheless, this claim has not gained recognition by data protection authorities, asserting that the simple fact of being accessible on social media did not constitute an implicit permit for scraping.[73] In light of these decisions, it is considered that the processing of personal data through scraping requires a separate legal ground, apart from that of the original processing, if the purpose of the further processing differs from the original one.[74]

Although the original website from which personal data is scraped may justify its data processing on a lawful basis, this justification does not automatically extend this basis to any subsequent

and (vi) legitimate interest. However, if incompatible further processing takes place as in the Clearview case, requirements in Article 6(4) of the GDPR must be satisfied.

[69] ICO, ‘Joint Statement on Data Scraping and the Protection of Privacy’ (24 August 2023) 1 <https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf> accessed 15 December 2023; CEDPO ‘Generative AI: The Data Protection Implications’ (n 21) 10; Janos Meszaros and others, ‘ChatGPT: How Many Data Protection Principles Do You Comply with?’ (2023) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4647569> accessed 4 December 2023 9; ICO, ‘Guide to the General Data Protection Regulation (GDPR)’ 14 October 2022 167; WP29, ‘Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC’ (WP 217, 844/14/EN, 9 April 2014) 39.

[70] CNIL, SAN-2022-019, 17 October 2022 <https://www.cnil.fr/en/facial-recognition-20-million-euros-penalty-against-clearview-ai> accessed 25 November 2023; EDPB, ‘Facial Recognition: Italian SA Fines Clearview AI EUR 20 Million’ (*EDPB*, 10 May 2023) <https://edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en> accessed 25 November 2023; EDPB, ‘Decision by the Austrian SA against Clearview AI Infringements of Article 5, 6, 9, 27’ (*EDPB*, 12 May 2023) <https://edpb.europa.eu/news/national-news/2023/decision-austrian-sa-against-clearview-ai-infringements-articles-5-6-9-27_en> accessed 25 November 2023.

[71] James Clayton and Ben Derico, ‘Clearview AI Used Nearly 1m Times by US Police, It Tells the BBC’ *BBC* (San Francisco, 27 March 2023) <https://www.bbc.com/news/technology-65057011> accessed 25 November 2023.

[72] GDPR Article 9(2)(e); Gaurav Pathak, ‘Manifestly Made Public: Clearview and GDPR’ (2022) 8 European Data Protection Law Review 419, 420.

[73] Pathak, ‘Manifestly Made Public: Clearview and GDPR’ (n 72) 421.

[74] Catherine Altobelli and others, ‘To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR’ in Jean Herveg (eds), *Deep Driving into Data Protection* (Larcier 2021) 157-158.

processing.[75] In line with the purpose limitation principle, processing must adhere to specified, explicit, and legitimate purposes, prohibiting any incompatible further processing.[76] However, in principle, the GDPR does not permit further processing unless (i) the purpose of further processing is compatible with the initial collection and processing purpose, (ii) the controller secures consent from the data subject for the new purpose, or (iii) a European or Member State law which represents a necessary and proportionate measure is in place.[77] In the context of legal ambiguity surrounding the reuse of data for AI training, a debate persists regarding whether scraping data that constitutes further processing necessitates the application of Articles 6(1) of the GDPR.[78] On the other hand, there remains the question of whether Article 6(4) of the GDPR should be regarded as a distinct legal basis for further processing.[79] Should the requirement arise to provide a legal basis as specified in Article 6(1) of the GDPR, the most appropriate option would be to rely on the legitimate interest of the controller.[80]

At the outset, a compatibility test that assesses whether the later purpose aligns with the primary purpose should be conducted.[81] A key factor is whether data scraping can be reasonably expected

[75] Andrew M. Parks, 'Unfair Collection: Reclaiming Control of Public Available Personal Information from Data Scrapers' (2022) 120 Michigan Law Review 913, 924.

[76] GDPR Article 5(1)(b); Philipp Hacker, 'A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act' (2021) 13 Law, Innovation and Technology 257, 276; Altobelli, 'To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR' (n 74) 166 and 167.

[77] GDPR Article 6(4); Altobelli, 'To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR' (n 74) 167.

[78] Hacker, 'A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act' (n 76) 276; Altobelli, 'To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR' (n 74) 167; Waltraut Kotschy, 'Article 6. Lawfulness of Processing' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 338; Wouter Seinen and others, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' (6th Annual Privacy Forum, Barcelona, June 2018) 155; Bart Custers and Helena Ursic, 'Legal Barriers and Enablers to Big Data Reuse' (2016) 2 European Data Protection Law Review 208, 213; WP 29, 'Opinion 03/2013 on Purpose Limitation' (2 April 2013, 00569/13/EN WP 203) 36.

[79] ibid.

[80] Hacker, 'A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act' (n 76) 276; Custers and Ursic, 'Legal Barriers and Enablers to Big Data Reuse' (n 78) 213; WP 29, 'Opinion 03/2013 on Purpose Limitation' (n 78) 36.

[81] GDPR Article 6(4) outlines a set of factors which must be considered in the evaluation of the compatibility of further processing: (a) any link between original and intended processing purposes, (b) the context under which data was collected, especially the reasonable expectations of data subjects, considering relationship between data subjects and the data controller, (c) nature of personal data, (d) potential impact of further processing on data subjects, and (e) presence of protective measures.

by data subjects.[82] People share personal information online, both intentionally and unintentionally.[83] Despite public accessibility of this information, it is vital to consider whether scraping activities surpass what a reasonable person would expect.[84] Whereas companies may detail such practices in their privacy policies, complex language often leads to a lack of attention and uninformed processing by data subjects.[85] Regarding text producers' data scraping practices, collecting data from various internet sources makes it unfeasible to expect to include such details in the privacy policies of all scraped websites and to notify all the data subjects.[86] Even if data scraping notifications are in place, they likely exceed the reasonable expectations of data subjects, as the purpose of further processing is not compatible with the original due to the differences in processing contexts.[87] Considering their non-transparent policies, it is unclear whether they implement sufficient protective measures for data security in scraping practices. As a result, the subsequent processing by text producers is generally deemed incompatible with the original purpose.[88]

In this situation, a specific European or Member State law which represents a necessary and proportionate measure or consent of the data subject must be sought by the controller for any incompatible further processing.[89] If further processing relies on the consent of the data subject,

[82] Paul De Hert and Irene Kamara, 'Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach' (2018) 4 Brussels Privacy Hub 321, 339; ICO, 'Lawful Basis for Processing: Legitimate Interests' (22 March 2018) 18.

[83] Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 14.

[84] ICO, 'Lawful Basis for Processing: Legitimate Interests' (n 82) 18.

[85] However, under the transparency principle, the information provided must be presented in a manner that is straightforward, easily accessible and uses language that is clear and easy to comprehend. See GDPR Recital 39; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 69; Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 13; Custers and Ursic, 'Legal Barriers and Enablers to Big Data Reuse' (n 78) 214; Graydon Hayes, 'Have You Read Your Privacy Policies' (Privacy Commissioner, 16 August 2019) <https://www.privacy.org.nz/blog/have-you-read-your-privacy-policies/> accessed 26 November 2023.

[86] ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 69; Custers and Ursic, 'Legal Barriers and Enablers to Big Data Reuse' (n 78) 214; Bart Custers and Helena Ursula, 'Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection' (2016) 6 International Data Privacy Law 4, 11.

[87] Seinen, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' (n 78) 163; WP 29, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse' (5 June 2013, 1021/00/EN WP 207) 19; WP 29, 'Opinion 03/2013 on Purpose Limitation' (n 78) 23.

[88] The fact that text producers acquire data from the internet and various other sources complicates the illustration that the data subject can reasonably expect such processing. See GDPR Recital 50; ICO, 'Lawful Basis for Processing: Legitimate Interests' (n 82) 18.

[89] ibid; Hacker, 'A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act' (n 76) 293; Kotschy, 'Article 6. Lawfulness of Processing' (n 78) 343.

such consent will be the legal basis for further processing and should be in compliance with the GDPR's Article 4(11) requirements.[90] Notifying data subjects of any further processing is in line with their reasonable expectations, as they are empowered with control over their data, allowing them to actively express consent or dissent regarding the processing.[91]

In any case, data subjects hold the right to be informed, compelling controllers to furnish them with essential information about processing activities.[92] The concept of processing is intertwined with informed data subjects, as lawful processing must include notifying data subjects of the presence and purposes of the processing operation, adhering to the principles of fair and transparent processing.[93] When personal data is not directly acquired from the data subject, the controller is also obliged to provide a detailed explanation, including the purposes of collecting the personal data, the source the personal data have been collected, documentation of the way the personal data have been sourced, and the pathway for users to exercise their data protection rights.[94] Should further processing ensue, the data subject ought to be informed beforehand, avoiding retroactive notifications.[95] Yet, informing data subjects whose data undergo processing for training purposes might prove impossible or demand disproportionate effort.[96] Despite text

[90] These requirements specify four critical elements of valid consent: (i) free, ensuring the data subject has a genuine choice and control over their data without negative repercussions for refusal; (ii) specific, where the consent is tied to clear and particular purposes; (iii) informed, mandating that the data controller provides detailed information about processing purposes; and (iv) unambiguous, necessitating active, explicit consent from the data subject, which cannot be presumed from silence or lack of action. See EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (4 May 2020); WP 29, 'Opinion 15/2011 on the Definition of Consent' (13 July 2011, 01197/11/EN WP187); Altobelli, 'To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR' (n 74) 167.

[91] Altobelli, 'To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR' (n 74) 173; Kotschy, 'Article 6. Lawfulness of Processing' (n 78) 343; Seinen, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' (n 78) 156 and 159; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (11 April 2018, 17/EN WP260 rev.01) 23.

[92] GDPR Recital 50; Mario Egbe Mpame and Robert Niedermeier, 'Processing Personal Data under Article 6(f) of the GDPR: The Concept of Legitimate Interest' (2019) 3 International Journal for the Data Protection Officer, Privacy Officer and Privacy Counsel 18, 27.

[93] GDPR Article 14 in conjunction with Recital 60.

[94] Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 20; Gabriela Zanfir-Fortuna, 'Article 14. Information To Be Provided Where Personal Data Have Not Been Obtained from the Data Subject' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 444; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 18; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 10.

[95] GDPR Article 14(4); WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 17.

[96] GDPR Article 14(5)(b) in conjunction with Recital 61; Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 14; Zanfir-Fortuna, 'Article 14. Information To Be Provided Where Personal Data Have Not Been Obtained from the Data Subject' (n 94) 446; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 28.

producers collecting data from various sources, the challenge of efficiently identifying all data origins persists. They cannot claim exemption from this requirement unless demonstrate the impossibility.[97] The excessive number of data subjects and protective measures employed may add a positive weight on the side of the controller to the disproportionate effort to inform the data subject.[98] In any situation, the controller should uphold the accountability principle,[99] conducting a balancing test to gauge the impact on data subjects if the information is withheld.[100] Furthermore, protective steps, such as performing a data protection impact assessment and issuing public explanations about the use of diverse data sources align with data subjects' reasonable expectations, ensuring the safeguarding of their rights.[101]

The legitimate interest basis, exemplified in cases like Clearview, presents a potential foundation.[102] Nevertheless, conducting the "balancing test" – which weighs the controller's interest against the data subject's fundamental rights, freedoms, and interests – remains a significant challenge, even when the controller or a third party's interests are explicitly defined and lawful.[103] Given the fact that tech companies indicate that they released text producers for research purposes without charge, this discourse may favour the controller's side during this balancing test, as there is no mere commercial interest.[104] On the other hand, apart from the erosion

[97] WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 29.

[98] GDPR Recital 62.

[99] According to the accountability principle, processing activities must adhere to the rules and principles and adopts appropriate strategies envisaged in the GDPR. These strategies ought to be customized based on the specific details and context of the situation. This principle strongly correlates with the transparency principle, as being transparent is a fundamental aspect of being an accountable controller. See GDPR Article 5(2) in conjunction with Recital 39; Cecile De Terwangne, 'Article 5. Principles relating to Processing of Personal Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 318-319; WP 29, 'Opinion 3/2010 on the Principle of Accountability' (13 July 2010, 00062/10/EN WP 173) 9, 11, 13, and 14.

[100] Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 14; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 31.

[101] ICO, 'Guide to the General Data Protection Regulation (GDPR)' (n 69) 167; ibid.

[102] CNIL, 'Restricted Committee Deliberation No. SAN-2022-019 of 17 October 2022 concerning Clearview AI' 10; Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 10; Parks, 'Unfair Collection: Reclaiming Control of Public Available Personal Information from Data Scrapers' (n 75) 935; Michael Schade, 'How ChatGPT and Our Language Models Are Developed (*OpenAI*, November 2023) <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed> accessed 5 December 2023.

[103] Kotschy, 'Article 6. Lawfulness of Processing' (n 78) 338; WP29, 'Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC' (n 69) 24, 25, and 29.

[104] Pablo Trigo, 'Can Legitimate Interest Be An Appropriate Lawful Basis for Processing Artificial Intelligence Training Datasets' (2023) 48 Computer Law & Security Review 7 and 8; ICO, 'Guide to the General Data Protection Regulation (GDPR)' (n 69) 76; ICO, 'Lawful Basis for Processing: Legitimate Interests' (n 82) 14; WP29, 'Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC' (n 69) 35.

of self-determination over personal data, the probability of processing inaccurate or irrelevant information by text producers could a disproportionately impact the data subject, depending on the severity of the consequences.[105] Additionally, establishing reasonable expectations of data subjects becomes more difficult due to the absence of a pre-existing relationship between text producers and data subjects, given that the former acquires data from third parties.[106] In striving for equilibrium between processing interests and their impact on the data subjects, employing protective measures – such as transparently explaining the controller's legitimate interests to data subjects – may help alleviate tensions.[107]

Under the accuracy principle, controllers are obliged to maintain the accuracy of data and ensure it is kept up to date.[108] For text producers utilizing training data, there is an inherent risk of including outdated information since data scraped from the internet, while possibly current at the time of collection, can quickly become outdated.[109] Beyond just outdated data, there is an increasing risk that text producers are prone to generate inaccurate outputs which might arise from different scenarios. The prevalence of misinformation and disinformation[110] online results in the

Contrary to the WP 29's reference to public interests or the interests of the wider community in evaluating the controller's legitimate interest, this group may also comprise the controller or processor. Therefore, it is considered that the concept of the 'third party' should warrant a narrower scope. See De Hert and Kamara, 'Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach'(n 82) 333-334.

[105] GDPR Recital 47; Mpame and Niedermeier, 'Processing Personal Data under Article 6(f) of the GDPR: The Concept of Legitimate Interest' (n 92) 23; De Hert and Kamara, 'Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach'(n 82) 335-336; WP29, 'Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC' (n 69) 38.

[106] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 10; ICO, 'Lawful Basis for Processing: Legitimate Interests' (n 82) 18.

[107] Mpame and Niedermeier, 'Processing Personal Data under Article 6(f) of the GDPR: The Concept of Legitimate Interest' (n 92) 23; WP29, 'Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC' (n 69) 41-42; Kotschy, 'Article 6. Lawfulness of Processing' (n 78) 339; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 34. For instance, following the DPA's decision, OpenAI updated its privacy policy to include consent and legitimate interest as its legal grounds, specifying the nature of its interests. See 'Europe Privacy Policy' (*OpenAI*, 22 June 2023) <https://openai.com/policies/eu-privacy-policy> accessed 26 November 2023.

[108] GDPR Article 5(1)(e) in conjunction with Recital 39; Cecile De Terwangne, 'Article 5. Principles relating to Processing of Personal Data' (n 99); Sanjay Sharma, *Data Privacy and GDPR Handbook* (Wiley 2020) 130.

[109] Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 16. For instance, the training data used for ChatGPT 3.5 is confined to information available up to September 2021. See Raymond Niles, 'GPT-3.5 Turbo Updates' (*OpenAI*, November 2023) <https://help.openai.com/en/articles/8555514-gpt-3-5-turbo-updates> accessed 26 November 2023.

[110] 'Misinformation' refers to the production and spread of inaccurate information without intent to harm, whereas 'disinformation' denotes the deliberate production and dissemination of false content. The common thread between these concepts is their potential to mislead individuals. See OECD, 'AI Language Models: Technological, Socio-Economic and Policy Considerations' (n 13) 33; 'Misinformation vs Disinformation' (*Taylor & Francis*)

collection and processing factually inaccurate data by text producers, thus producing inaccurate responses, as pointed out by the DPA.[111] Given that text producers are not designed to capture the sense of the information they generate, there is a notable risk of producing outputs that are non-existing yet persuasive outputs.[112] While providers of text producers issue disclaimers about inaccuracies in results, a critical aspect is whether an average user can verify these outputs' accuracy.[113] Potential solutions might involve requiring text producers to corroborate their outputs by citing relevant sources, thereby enhancing reliability.[114] However, these models might refer to sources which do not exist or incorrectly reflect the sources they cite.[115]

Considering the abovementioned concerns, tech companies offering text producers must maintain transparency about their data sources, especially as their models tend to mislead users.[116] The ”black box” concept has become a buzzword for AI, highlighting the lack of transparency in their operation systems.[117] Under transparency, controllers are required to ensure that data subjects are

<https://insights.taylorandfrancis.com/social-justice/misinformation-vs-disinformation/> accessed 1 December 2023; 'Factsheet 4: Types of Misinformation and Disinformation' (*UNHCR*) <https://www.unhcr.org/innovation/wp-content/uploads/2022/02/Factsheet-4.pdf> accessed 1 December 2023.

[111] GPDP (n 23); A. Feder Cooper and others, 'Report of the 1st Workshop on Generative AI and Law' (*Arxiv*, 2023) 11 <https://arxiv.org/pdf/2311.06477.pdf> accessed 30 November 2023; Krzysztof Wach and others, 'The Dark Side of Generative Artificial Intelligence: A Critical Analysis of Controversies and Risks of ChatGPT' (2023) 11 Entrepreneurial Business and Economics Review 7, 12 and 15; UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 38) 37.

[112] OECD, 'Initial Policy Considerations for Generative Artificial Intelligence' (n 17) 14; Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 14; Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 12; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 22; Elena Esposito, 'Algorithmic Memory and the Right to Be Forgotten' (2017) 4 Big Data & Society 1, 4 and 5. For instance, a case of defamation arose from an inaccurate output generated by ChatGPT. See Byron Kaye, 'Australian Mayor Readies World's First Defamation Lawsuit over ChatGPT Content' *Reuters* (5 April 2023) <https://www.reuters.com/technology/australian-mayor-readies-worlds-first-defamation-lawsuit-over-chatgpt-content-2023-04-05/> accessed 7 December 2023.

[113] OECD, 'Initial Policy Considerations for Generative Artificial Intelligence' (n 17) 14; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 22; Ruochen Zhao and others, 'Can ChatGPT-like Generative Models Guarantee Factual Accuracy? On the Mistakes of New Generation Search Engines' (*Arxiv*, 2023) 1 <https://arxiv.org/pdf/2304.11076.pdf> accessed 1 December 2023; Irene Solaiman and others, 'Evaluating the Social Impact of Generative AI Systems in Systems and Society' (*Arxiv*, 2023) 11-12 <https://arxiv.org/pdf/2306.05949.pdf> accessed 1 December 2023; CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 4.

[114] OECD, 'Initial Policy Considerations for Generative Artificial Intelligence' (n 17) 15-17.

[115] Zhao, 'Can ChatGPT-like Generative Models Guarantee Factual Accuracy? On the Mistakes of New Generation Search Engines' (n 113) 2; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 22; Aaron Welborn, 'ChatGPT and Fake Citations' (*Duke University*, 9 March 2023) <https://blogs.library.duke.edu/blog/2023/03/09/chatgpt-and-fake-citations/> accessed 7 December 2023.

[116] GPDP (n 23); Zhao, 'Can ChatGPT-like Generative Models Guarantee Factual Accuracy? On the Mistakes of New Generation Search Engines' (n 113) 5.

[117] Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 16; UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 38) 39; McDonagh and Paterson, 'Data Protection in

fully informed about potential risks, protective measures, and their rights concerning data processing.[118] The fundamental consideration of the transparency principle is to provide clarity and openness in how personal data is processed, making certain that data subjects are fully aware of the nature, purpose, and the extent of the data processing through such information articulated in straightforward language.[119] The rationale for this approach is to guarantee that average data subjects can fully understand how their data is processed.[120] By improving transparency and accountability, a middle ground can be achieved between the fallible decision-making process of text producers and the protection of data subjects' rights and interests, at least providing them with clear insight into the risks they face. On the other hand, the operation of text producers exemplifies the black box problem, as they often keep their training data and data processing techniques hidden from public scrutiny.[121] This secrecy is largely ascribed to their intricate technical architecture,[122] which, in return, fosters a propensity for operating as closed systems.[123] Conveying the intricate and multifaceted systems of text producers to data subjects in a GDPR-compliant manner poses a significant challenge.[124] Providers might defend the opacity and the restriction on external examinations by asserting the need to safeguard trade secrets and avoid safety hazards.[125] To address this dilemma, a recommended approach is to provide access to assembled information for

an Era of Big Data: The Challenges Posed by Big Personal Data' (n 40) 3; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 86.

[118] GDPR Article 5(1)(a) in conjunction with Recital 39; OECD, 'AI Language Models: Technological, Socio-Economic and Policy Considerations' (n 13) 35; De Terwangne, 'Article 5. Principles relating to Processing of Personal Data' (n 99) 314-315; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 6; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 27-28.

[119] GDPR Article 12(1); ibid.

[120] WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 7.

[121] Sebastian, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (n 44) 2; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 20; EP, 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 4) 44-45.

[122] Complexity poses a significant challenge in systems utilizing deep neural networks, as this method limits the ability to monitor processing. See Tansib Hosain and others, 'Path to Gain Functional Transparency in Artificial Intelligence with Meaningful Explainability' (2023) 3 Journal of Metaverse 166, 167 and 169.

[123] Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 19; Boris P. Paal, 'Artificial Intelligence as A Challenge for Data Protection Law and Vice Versa' in Silja Voeneky and others (eds), *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (CUP 2022) 293; Helena U. Vrabec, *Data Subjects Rights under the GDPR* (Oxford 2021) 5; Radim Polcak, 'Article 12. Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 407.

[124] ibid.

[125] Recital 63; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 20; Polcak, 'Article 12. Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject' (n 123) 407.

review, rather than disclosing the entire operational strategy of the system.[126] This approach aims to reconcile the rights and interests of data subjects with the challenges of non-transparent policies, while still safeguarding trade secrets. Furthermore, despite publishing disclaimers for compliance with the transparency obligation, they cannot foresee or account for the processing of personal data that occurs through various manners, such as user-initiated prompts.[127] The task becomes even more burdensome when users input personal data into text producers.

Although training data might be adjustable by filtering out harmful content, managing the outputs and inputs is far more challenging as these outputs cannot be controlled once generated.[128] When processing involves user inputs containing personal information, that when combined with an existing dataset, could reveal sensitive data, it raises concerns about compliance with Article 6 of the GDPR.[129] Justification for processing user data might be based on consent or the performance of a contract.[130] As evolving training systems leverage user feedback to fine-tune models to improve accuracy, often relying on the legitimate interest of the controller, an opt-out system becomes crucial to allow users to prevent their data from being used for training.[131] While controllers might argue that data subjects can reasonably anticipate their data being used to enhance services, it is equally important for them to implement an opt-out system to prevent text

[126] Recital 63; Polcak, 'Article 12. Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject' (n 123) 407.

[127] This consideration reveals that the transparency obligation encompasses two aspects: beforehand and subsequent transparency. See EP, 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 4) 53.

[128] Lee, 'AI and Law: The Next Generation' (n 48) 10-12; Ganguli, 'Predictability and Surprise in Large Generative Models' (n 14) 1750, 1752, 1754, and 1755; Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 16.

[129] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 19.

[130] 'Europe Privacy Policy' (n 107). To rely on the performance of a contract, the contract in question must be concluded between the controller and the data subject. This legal ground is applicable if the processing is necessary for fulfilling the contract. Such necessity should be objectively interpreted, considering the viewpoints of both the data subject and controller. In the context of text producers, essential information such as user names, contact details, and other pertinent data may be collected based on the need to perform the contract. See GDPR 6(1)(b) in conjunction with Recital 40; Kotschy, 'Article 6. Lawfulness of Processing' (n 78) 331-332; EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (n 90) 10; ICO, 'Lawful Basis for Processing' (19 May 2023) 16 and 17. See n 87 for information regarding consent.

[131] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 7 and 12; Ghost in the Machine' (n 17) 18; OECD, 'AI Language Models: Technological, Socio-Economic and Policy Considerations' (n 13) 21; OECD, 'Initial Policy Considerations for Generative Artificial Intelligence' (n 17) 6; Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 5; Michale Schade, 'How Your Data Is Used to Improve Model Performance' (*OpenAI*, November 2023) <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance> accessed 5 December 2023.

producers from using their data for training purposes.[132] Introducing and implementing such measures demonstrates the controller's commitment to respecting individual data preferences and reinforces their position in the balancing test.[133]

However, processing sensitive data faces stricter limitations, as the processing is prohibited in principle, and a non-exhaustive list of exceptions must be demonstrated by controllers to process sensitive data.[134] Controllers must demonstrate compliance by showcasing their adherence to specific exemptions outlined in the GDPR.[135] In text producers, their upgraded versions, evolving to accept inputs beyond text, now receive inputs in the form of images or videos.[136] They capture necessary patterns from text-based or image-based inputs to generate results in accordance with the description of the image or text in the input.[137] While a person's voice and facial structure could qualify as biometric data, the GDPR interprets this data as sensitive only if it is used to uniquely identify individuals.[138] While photo and voices preserve their personal data nature if they are not used for direct recognition or identification, they might not fall under sensitive data provisions.[139] Although text producers process images and voices to execute tasks, such as crafting humorous captions for a photo, their primary goal is not user identification or recognition.

[132] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 12; Ghost in the Machine' (n 17) 46. For instance, ChatGPT introduced an opt-out mechanism after the DPA's decision. See 'Europe Privacy Policy' (n 107).

[133] ibid.

[134] The concept of sensitive data can be represented in four categories: (i) personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership; (ii) genetic and biometric to uniquely identify a natural person, (iii) data concerning health; (iv) data concerning a natural person's sex life or sexual orientation. See GDPR Article 9 in conjunction with Recital 51; ICO, 'Lawful Basis for Processing' (n 130) 37 and 38; WP 29, 'Advice Paper on Special Categories of Data' (20 April 2011) 8.

[135] The most appropriate exemption may be explicit consent which must be an active reaction to a proposal regarding the processing with affirmative action, either in oral or written form in the context of text producers. See ICO, 'Lawful Basis for Processing' (n 130) 39; WP 29, 'Advice Paper on Special Categories of Data' (20 April 2011) 8; EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (n 90) 20; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (n 41) 25-26.

[136] 'ChatGPT Can Now See, Hear, and Speak' (*OpenAI*, 25 September 2023) <https://openai.com/blog/chatgpt-can-now-see-hear-and-speak> accessed 3 December 2023; Jack Krawczyk, 'Bard's Latest Update: More Features, Languages and Countries' (*Google*, 13 July 2023) <https://blog.google/products/bard/google-bard-new-features-update-july-2023/> accessed 3 December 2023. New versions are able to respond with generated pictures via their image generators. See '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> accessed 3 December 2023.

[137] ibid.

[138] GDPR Articles 4(14) and 9 in conjunction with Recital 51; Lee A. Bygrave and Luca Tosoni, 'Article 14(4). Biometric Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 212; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (n 41) 8.

[139] ICO, 'Lawful Basis for Processing: Special Category Data' 17 October 2022 9; WP 29, 'Opinion 3/2012 on Developments in Biometric Technologies' (27 April 2012, WP 193, 00720/12/EN) 5.

However, these processes might inadvertently disclose such data in other contexts due to their presence in the database, aimed at enhancing performance. Moreover, the utilization of other sensitive data, including health information,[140] mandates compliance with Article 9 of the GDPR.[141]

Processing user data typically rests on consent or the performance of the contract, yet complexities emerge when users input information pertaining to another natural person.[142] For instance, a user utilized a text producer to write a letter and add a friend's contact details to the content. This scenario presents a challenge in determining the lawful basis for processing under the GDPR. On the other hand, when personal data concerning another individual is processed without direct acquisition from the data subject, it triggers the necessity for notification as stipulated by Article 13 of the GDPR.[143] An additional area of concern arises regarding the user's responsibility when sharing the personal data of individuals with text producers.[144]

Security attacks on text producers also introduce a major concern in data protection. The inherent security vulnerabilities of text producers introduce a significant risk, as they may be susceptible to attacks that could lead to the inadvertent disclosure of personal data.[145] One type of these incursions is designed to outwit the system, inducing text producers to deliver unforeseen

[140] GDPR Article 4(15).

[141] Text producers have the ability to elicit sensitive data from non-sensitive data. See Michael Veale and others, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' [2018] 376 Philosophical Transactions of the Royal Society 3.

[142] Uri Gal, 'ChatGPT Is A Data Privacy Nightmare. If You've Ever Posted Online, You Ought To Be Concerned' (*The University of Sydney*, 8 February 2023) <https://www.sydney.edu.au/news-opinion/news/2023/02/08/chatgpt-is-a-data-privacy-nightmare.html> accessed 3 December 2023.

[143] Zanfir-Fortuna, 'Article 14. Information To Be Provided Where Personal Data Have Not Been Obtained from the Data Subject'(n 94) 444; WP 29, 'Guidelines on Transparency under Regulation 2016/679' (n 91) 18; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 10.

[144] UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 38) 50; Megan E. Kern, 'ChatGPT Generates More Than Data Outputs; Data Security and Privacy Concerns Grow as Artificial Intelligence Technology Rapidly Advances' (*Lexology*, 17 May 2023) <https://www.lexology.com/library/detail.aspx?g=1523494d-4f33-4602-b399-65be00d96d64> accessed 3 December 2023.

[145] CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 11-13; UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 38) 46; Wu, 'Unveiling Security, Privacy, and Ethical Concerns of ChatGPT' (n 49) 6 and 7; Veale, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (n 141) 6-8; Eduard Kovacs, 'Simple Attack Allowed Extraction of ChatGPT Training Data' (*Security Week*, 1 December 2023) <https://www.securityweek.com/simple-attack-allowed-extraction-of-chatgpt-training-data/> accessed 3 December 2023; Jim Chilton, 'The New Risks ChatGPT Poses to Cybersecurity Data' (*Harvard Business Review*, 21 April 2023) < https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity> accessed 3 December 2023.

responses or disclose personal data.[146] This form of manipulation leverages text producers' own processing mechanics through deceptive prompts, bypassing the need for a direct attack on the system's infrastructure.[147] Under the GDPR, data breaches require prompt notification to data subjects.[148] Yet, this requirement becomes particularly challenging due to the complex nature of text producers, where training data is sourced from a vast array of origins. Notifying data subjects often involves an impractical level of effort, sometimes bordering on the impossible[149] This conundrum brings to the forefront the unresolved issue of transparency and accountability obligations incumbent upon providers. To manage this security risk, a strategic shift is needed, in which providers should embrace external audits and foster transparency and accountability, offering a clearer window into their operations and methodologies.[150]

2.4. INTERIM CONCLUSION

In the context of text producers, defining the GDPR's material and territorial scopes is just the initial step to trigger its application. The crux of adherence lies in securing a lawful basis for data processing, followed by a strict observance of the GDPR's data processing principles. Text

[146] ibid. These security attacks lead to breaches of personal data due to the consequent revelation of personal data. See Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 20; Cedric Burton, 'Article 34. Communication of A Personal Data Breach to the Data Subject' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 659; WP 29, 'Guidelines on Personal Data Breach Notification under Regulation 2016/679' (6 February 2018, 18/EN WP250rev.01) 7.

[147] ibid.

[148] GDPR Articles 4(12) and 34 in conjunction with Recitals 86-88. However, the duty to inform data subjects about breaches requires high-level risks to the rights and freedom of data subjects, considering several factors, such as the characteristic of the breach, type, sensitivity, and quantity of personal data, and the severity of impact on individuals. See WP 29, 'Guidelines on Personal Data Breach Notification under Regulation 2016/679' (n 146) 22-26; Burton, 'Article 34. Communication of A Personal Data Breach to the Data Subject' (n 146) 659.

[149] An exemption exists for notifying data subjects of a data breach, akin to the one in processing notification. If informing data subjects involves disproportionate effort, controllers may be exempt from this requirement. This exemption is conditional upon providing a public notification or employing similar methods that effectively inform data subjects. See GDPR Article 34(2)(c); Burton, 'Article 34. Communication of A Personal Data Breach to the Data Subject' (n 146) 660; WP 29, 'Guidelines on Personal Data Breach Notification under Regulation 2016/679' (n 146) 22.

[150] Wu, 'Unveiling Security, Privacy, and Ethical Concerns of ChatGPT' (n 49) 9; CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 13; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 32-33. For example, in a move towards transparency, OpenAI released an update regarding the security attack occurred in March 2023. This action involved making a public notification available, thereby it may be seen as a step in fulfilling the transparency and notification obligations. See 'March 20 ChatGPT Outage: Here's What Happened' (*OpenAI*, 24 March 2023) <https://openai.com/blog/march-20-chatgpt-outage> accessed 7 December 2023.

producers face a particularly intricate task in this regard, as they must establish a legal basis not only for user data but also for the data used in training and data about other natural persons. After the establishment of legal grounds, the compliance journey shifts to the controller's duty to notify data subjects about the processing of their data even if the data has not been acquired from data subjects. Exemptions exist under certain conditions, such as impossibility or disproportionate effort, especially when protective measures are in place to uphold the fundamental rights and freedoms of data subjects. Furthermore, given its vulnerability to security attacks, the requirement to inform data subjects of personal data breaches in high-risk situations arises. However, should notifying each data subject prove to be an excessively onerous task, text producers might find themselves exempt from this obligation, providing the delicate balance between regulatory compliance and operational practicality.

3. TO WHAT EXTENT DOES THE GDPR MANDATE TEXT PRODUCERS TO ERASE PERSONAL DATA, CONSIDERING THEIR TECHNOLOGICAL INFRASTRUCTURE AND CAPABILITIES?

3.1. INTRODUCTION

After addressing the concerns with GDPR compliance for text producers, this Chapter will carry out an in-depth analysis of how the RTBF applies to text producers. This involves a critical examination of the technical feasibility for text producers to integrate this right into their systems. central to this analysis is the pursuit of answering the second sub-question: *To what extent are text producers obliged to erase personal data under the GDPR, considering their technological infrastructure and capabilities?* This exploration aims to elucidate the balance between legal mandates and the practical realities of advanced AI systems, offering insights into how these producers can align with the GDPR's requirements while navigating their technological landscapes.

3.2. THE REMIT OF THE RTBF UNDER THE GDPR

3.2.1. The Rise of the RTBF

In the latter half of the 20th century, as machines became increasingly integrated into daily life, they began to play a more prominent role in individuals' private lives.[151] This rise in AI use coincided with the recognition of the value of data, leading to the emergence of the data-driven ecosystem.[152] Tech companies, recognizing the potential of data, started collecting and utilizing vast amounts of it to offer personalized services, and especially with the rise of machine learning, they have started integrating this system into their operations to provide more personalised services.[153] As a result, they shifted people's habits towards spending more time online, thereby sharing more data which they can utilize for their services. However, many people remained unaware of the worth of their data to tech companies and how their data was being compiled and processed, in return, they have begun to incrementally lose self-determination over their personal data. This gradual loss of control over personal data prompted legislators to introduce reinforced instruments, including but not limited to the RTBF,[154] with the enactment of the GDPR.[155] To restore control over personal data, which was weakened through the imbalance of power and the advancement of technology, the RTBF was thoroughly regulated by the GDPR.[156]

The origins of the RTBF can be traced back to the *Google Spain* case, in which the CJEU ruled that Google must delist the name of the data subject upon request.[157] Notably, the *Google Spain* judgment did not directly apply the RTBF as outlined in the GDPR due to the availability of the news on the internet through alternative search keywords, aside from the individual's name.[158]

[151] Jef Ausloos, *The Right To Erasure in EU Data Protection Law* (OUP 2020) 39.

[152] Commission, 'Towards a Thriving Data-driven Economy' COM (2014) 442 final, 2 and 4.

[153] ibid; Sandra Wachter, 'Data Protection in the Age of Big Data' (2019) 2 Nature Electronics 6.

[154] Particularly, to reinstate the control over personal data in AI systems, the right of access, and the right to object which is closely related to the right to be forgotten are important. See GDPR Articles 15, 17 and 21 in conjunction with Recitals 59 and 63.

[155] Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 52 and 53; Vrabec, *Data Subjects Rights under the GDPR* (n 123) 131; Wachter, 'Data Protection in the Age of Big Data' (n 153) 6; Indra Splecker genannt Dohmann, 'A New Framework for Information Markets: Google Spain' (2015) 52 Common Market Law Review 1033, 1040; Commission, 'Towards a Thriving Data-driven Economy' (n 152) 11.

[156] Lambert, *The Right To Be Forgotten* (n 26) 103 and 153; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 55-57; Mira Burri and Rahel Schar, 'The Reform of the EU Data Protection Framework: Outlining Key Changes and Assessing Their Fitness for A Data-Driven Economy' Journal of Information Policy (2016) 6 479, 490; Commission, 'A Comprehensive Approach on Personal Data Protection in the European Union' COM (2010) 609 final 4.

[157] *Google Spain* (n 27).

[158] Vrabec, *Data Subjects Rights under the GDPR* (n 123) 136; Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' (n 28) 479; David Lindsay, 'The 'Right To Be Forgotten' by Search Engines under Data Privacy Law: A Legal and Policy Analysis of the Costeja Decision' in Andrew T. Keyton (ed), *Comparative Defamation and Privacy Law* (CUP 2016) 201; WP 29, 'Guidelines on the Implementation of the Court of the Justice of the European

However, the GDPR interprets the RTBF differently, emphasizing the actual erasure of personal data rather than delisting it from search engine results. The GDPR allows data subjects to invoke the RTBF if their claims align with the grounds specified in Article 17.[159] However, given the non-absolute nature of the RTBF, when bestowing data subjects with autonomy to decide how their data is utilized and processed, a balance must be struck between the rights and freedoms of data subjects, the interests of controllers, and the public's interests.[160] To achieve this balance, the GDPR includes a set of exemptions that allow controllers to be exempted from the obligation to apply the RTBF.[161]

Following the *Google Spain* decision, the French data protection authority ("CNIL") fined Google for not delisting the data from all the domain names, including those outside the EU.[162] CNIL argued that the delisting should apply universally, as Google's search engine domains were accessible in France and shared common databases and indexing across domains, although the search is aimed at the domain name that matches the country based on the online user's network address, Google's search engine domain names are reachable in France.[163] This raised questions about the territorial scope of the RTBF, with concerns that it should not extend beyond EU jurisdiction to interference in other jurisdictions.[164] According to a strict interpretation of the

Union Judgment on "Google Spain and Inc V. Agencia Espanola De Proteccion De Datos (AEPD) and Mario Costeja Gonzalez" C-131/12' (26 November 2014, 14/EN WP 225) 6 and 9.

[159] There are six grounds which prompt the application of the RTBF: (a) data no longer needed for the original purpose; (b) withdrawal of consent which is the basis for the processing and no alternative legal ground for the processing; (c) data subject's objection under Article 21(1) and there are no overriding legitimate grounds for the processing, or objection under Article 21(2); (d) unlawful processing of personal data; (e) legal obligation which requires data erasure; (f) data collected relating information society services. See GDPR Article 17(1).

[160] Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 71-72; Jef Ausloos and Aleksndra Kuczerawy, 'From Notice-and-Takedown to Notice-and Delist: Implementing Google Spain' (2016) 14 Colorado Technology Law Journal 291, 225; Commission, 'A Comprehensive Approach on Personal Data Protection in the European Union' (n 156) 4.

[161] *Google CNIL* (n 32) 17; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 248; Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' (n 28) 482. This exemptions are outlined as follows: if the processing is essential: (a) to uphold freedom of expression and information; (b) to fulfil a legal obligation, carry out a public task, or exercise official authority; (c) for public health interest; (d) public interest archiving, scientific and historical research, or statistical purposes; (e) for legal claims. See GDPR Article 17(3) in conjunction with Recital 65.

[162] *Google CNIL* (n 32) 13.

[163] ibid 13-14.

[164] ibid 17; Jessica Friesen, 'The Impossible Right To Be Forgotten' (2021) 47 Rutgers Computer & Technology Law Journal 173, 175 and 188; Kenneth Propp, 'Introductory Note To Google LLC V. Commission Nationale de l'informatique et des Libertés (CNIL) and Eva Glawischnig-Piesczek V. Facebook Ireland LTD. (C.J.E.U.)' (2020) 59 The Maerican Society of International Law 428, 429.

GDPR, the territorial scope is determined by the processing that impacts data subjects within the EU or is carried out by controllers or processors established within the EU.[165] On the other hand, to ensure the RTBF is an effective tool vis-à-vis data protection concerns, delisting is required to occur on pertinent domains.[166] However, Google introduced a geo-blocking option that restricts access to delisted information based on the IP addresses of internet users connecting from countries where delisting applies, and it is considered a promising formula for CNIL's concerns.[167] There remains a possibility of circumvention using VPNs or other tools enabling bypassing the filters.[168]

3.2.2. The Legal Grounds for Activating the RTBF

In the delineation of this right in text producers, the four legal grounds unfold in different contexts. Central to this is consent that text producers lean on user consent for data processing.[169] Users, holding the reins of control, can retract their consent at any moment, and text producers must acknowledge it as not just a mere reversal but a powerful tool.[170] Therefore, text producers are required to offer a way for data subjects to withdraw their consent that is as straightforward as the method used to provide it initially.[171] Upon the withdrawal, controllers are obliged to erase the personal data in question, should no alternative legal basis to legitimize the processing.[172] The revocation of consent leads to the implementation of the RTBF where text producers cannot justify their processing of user data as a necessary part of the performance of a contract. In cases where

[165] GDPR Article 3 in conjunction with Recitals 22-24; Case 507/17 *Google LLC v the Commission Nationale de l'informatique et des Libertés* (CNIL) [2019], Opinion of AG Szpunar, paras 48-49; Vrabec, *Data Subjects Rights under the GDPR* (n 123) 138; Svantesson, 'Article 3. Territorial Scope' (n 55) 83.

[166] Edoardo Celeste and Federico Fabbrini, 'The Right To Be Forgotten In the Digital Age: The Challenges of Data Protection Beyond Borders' (2020) 21 German Law Journal 55, 64; WP 29, 'Guidelines on the Implementation of the Court of the Justice of the European Union Judgment on "Google Spain and Inc V. Agencia Espanola De Proteccion De Datos (AEPD) and Mario Costeja Gonzalez" C-131/12' (n 158) 9.

[167] Yann Padova, 'Is the Right To Be Forgotten A Universal Regional, or 'Glocal' Right?' (2019) 9 International Data Privacy Law 15, 26.

[168] ibid 28; Friesen, 'The Impossible Right To Be Forgotten' (n 164) 191-192; Dohmann, 'A New Framework for Information Markets: Google Spain' (n 155) 1043; WP 29, 'Guidelines on the Implementation of the Court of the Justice of the European Union Judgment on "Google Spain and Inc V. Agencia Espanola De Proteccion De Datos (AEPD) and Mario Costeja Gonzalez" C-131/12' (n 158) 9.

[169] 'Europe Privacy Policy' (n 107).

[170] GDPR Article 7(3) in conjunction with Recital 65; Europe Privacy Policy' (n 107); Vrabec, *Data Subjects Rights under the GDPR* (n 123) 142-143; EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (n 90) 23-24; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 30 and 41.

[171] ibid; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 46; Schade 'How Your Data Is Used to Improve Model Performance' (n 131); 'Europe Privacy Policy' (n 107).

[172] GDPR Article 17(1)(b); ibid.

consent fades, data subjects step onto the stage empowered to object to processing predicated on the legitimate interest ground.[173] Yet, it is not a straightforward path since the conditions require conducting a balancing test where text producers must establish that their compelling legitimate grounds outweigh the rights and freedoms of data subjects or are indispensable for the weave of legal claims.[174] This would be the case when text producers rely on the legitimate interest basis for processing both user data and training data. While navigating this path may seem more arduous, as the grounds for exemption must be not just legitimate but compelling, it casts a wider scope, embracing a broader spectrum of justifiable grounds.[175]

Amidst a chorus of scrutiny from national data protection authorities, ChatGPT finds itself at the heart of a GDPR compliance debate, particularly concerning the legality of its data processing practices.[176] In this complex realm of legality, the instance where the processing of training data falls short of lawful standards emerges as a potential catalyst, establishing the invocation of the RTBF. The use of the training data in text producers faces scrutiny due to the secondary use of personal data, often scraped from the internet.[177] The processing of training data, which diverges from its original intended purpose casts doubts on the legality of such actions.[178] Although the legitimate interest ground is commonly considered the most appropriate legal basis for justifying the use of the training data by text producers, it is doubtful that controllers would prevail in the

[173] GDPR Articles 17(1)(c) and 21(1); Vrabec, *Data Subjects Rights under the GDPR* (n 123) 143; Gabriela Zanfir-Fortuna, 'Article 21. Right To Object' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 517; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 211. On the other hand, the application scope of the right to object in direct marketing purposes is not limited solely to these legal grounds.

[174] GDPR Article 21(1); Zanfir-Fortuna, 'Article 21. Right To Object' (n 173) 517; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 213; EDPB, 'Guidelines 5/2019 on the Criteria of the Right To Be Forgotten in the Search Engines Cases under the GDPR' (2 December 2019) 9.

[175] Zanfir-Fortuna, 'Article 21. Right To Object' (n 173) 517; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 211.

[176] See n 65.

[177] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 9; ICO, 'Joint Statement on Data Scraping and the Protection of Privacy' (n 69) 1; CEDPO 'Generative AI: The Data Protection Implications' (n 21) 10.

[178] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 15; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 29 and 31; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 202 and 203; Seinen, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' (n 78) 163; WP 29, 'Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse' (5 June 2013, 1021/00/EN WP 207) 19; WP 29, 'Opinion 03/2013 on Purpose Limitation' (n 78) 23.

necessary balancing test.[179] This uncertainty surrounding the lawfulness of the processing paves the way for the activation of the RTBF on the grounds of unlawful processing.[180]

The intense debate in the *Google Spain* case centred on whether the presence of inaccurate, irrelevant, or outdated personal data warrants invoking the RTBF, sets a significant presence.[181] Even when the collection and processing of the training data have a legal ground listed in Article 6 of the GDPR, the presence of inaccurate, irrelevant, or outdated personal data, in light of the purpose of the processing could trigger the application of the RTBF.[182] This discussion is crucial in laying the groundwork for the implementation of the RTBF in cases where text producers handle such data or generate results that include these types of data.[183] For inaccurate personal data, one may argue that the right to rectification should be exercised to correct such data.[184] Given the complex technical background within which text producers operate, rectification may not be feasible.[185] In such scenarios, invoking the RTBF becomes a suitable alternative to ensure discontinuation of the processing of inaccurate data.

After the establishment of the ground for the RTBF, particularly following the criticism towards the CJEU's tendency to favour the right to data protection over the freedom of expression and the

[179] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 10; Hacker, 'A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act' (n 76) 276; ICO, 'Guide to the General Data Protection Regulation (GDPR)' (n 69) 76; ICO, 'Lawful Basis for Processing: Legitimate Interests' (n 82) 14; Custers and Ursic, 'Legal Barriers and Enablers to Big Data Reuse' (n 78) 213; WP 29, 'Opinion 03/2013 on Purpose Limitation' (n 78) 36; WP29, 'Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC' (n 69) 35.

[180] GDPR Article 17(1)(d); Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 215; EDPB, 'Guidelines 5/2019 on the Criteria of the Right To Be Forgotten in the Search Engines Cases under the GDPR' (n 174) 9 and 10.

[181] *Google Spain* (n 27) 16 and 19.

[182] Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 201; Vrabec, *Data Subjects Rights under the GDPR* (n 123) 134 and 141; Sharma, *Data Privacy and GDPR Handbook* (n 108) 205 and 206; EDPB, 'Guidelines 5/2019 on the Criteria of the Right To Be Forgotten in the Search Engines Cases under the GDPR' (n 174) 7.

[183] Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (n 67) 12 and 16; Hacker, 'Regulating ChatGPT and other Large Generative AI Models' (n 24) 14; Norwegian Consumer Council, 'Ghost in the Machine' (n 17) 22; Esposito, 'Algorithmic Memory and the Right to Be Forgotten' (n 112) 4-5.

[184] GDPR Article 16; Alexander A Wodi, 'The EU General Data Protection Regulation (GDPR): Five Years After and the Future of Data Privacy Protection in Review' [2023] 5 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4601142> accessed 10 December 2023; Laura Lagone, 'The Right To Be Forgotten: A Comparative Analysis' [2012] 8 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2229361> accessed 10 December 2023.

[185] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 7. For example, OpenAI declared that they may not be able to rectify an incorrect data in some instances. See 'Europe Privacy Policy' (n 107).

right of access to information in the *Google Spain* decision, it becomes imperative to balance these rights.[186] This balance is a crucial aspect, ensuring that while the RTBF provides a shield for data protection, it does not unduly infringe upon the freedom of expression and the right of access to information.[187] The challenge lies in finding a harmonious equilibrium where the RTBF is upheld without compromising public access to information and the freedom of expression.[188] This balancing act becomes even more complex when the data subject is a well-known personality, as the rights to freedom of expression and access to information are often deemed to take precedence over the RTBF.[189] Furthermore, extending the application scope of the RTBF beyond the borders of the EU could have adverse implications for the freedom of expression and the right of access to information, thereby an approach that respects the balance between data protection and the free flow of information on an international scale should be adopted.[190] This restriction may appear in text producers when the data subject is famous.[191] In conducting the necessary balancing test for the RTBF, considerations extend beyond mere fame.[192] Other pertinent factors, such as the nature of the data and the public's interest in accessing this information must be also considered.[193] This examination involves assessing the right to data protection of the individual against the value of

[186] *Google Spain* (n 27) 19-20; Case C-131/12 *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* [2014], Opinion of AG Jääskinen, paras 132-134; Vrabec, *Data Subjects Rights under the GDPR* (n 123) 137; Oksana V. Kiriiak, 'The Right To Be Forgotten: Emerging Legal Issues' (2021) 3 Review of European and Comparative Law 27, 32; Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' (n 28) 480; Robert C. Post, 'Data Privacy and Dignitary Privacy: Google Spain, the Right To Be Forgotten, and the Construction of the Public Sphere' (2018) 67 Duke Law Journal 981, 1015 and 1016; Dohmann, 'A New Framework for Information Markets: Google Spain' (n 155) 1046; Herke Kranenborg, 'Google and the Right To Be Forgotten' (2015) 1 European Data Protection Law Review 70, 78; Christopher Wolf, 'Impact of the CJEU's Right To Be Forgotten' (2014) 21 Maastricht Journal of European and Comparative Law 547, 553-554; Lagone, 'The Right To Be Forgotten: A Comparative Analysis' (n 184) 7.

[187] ibid; Sharma, *Data Privacy and GDPR Handbook* (n 108) 207.

[188] Opinion of AG Szpunar (n 165) paras 60-63; Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 255; Vrabec, *Data Subjects Rights under the GDPR* (n 123) 146.

[189] *Google Spain* (n 27) 17; Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 18; Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' (n 28) 480.

[190] Opinion of AG Szpunar (n 165) para 61; Opinion of AG Jääskinen (n 186) para 121;

[191] Meszaros, 'ChatGPT: How Many Data Protection Principles Do You Comply with?' (n 69) 18; Post, 'Data Privacy and Dignitary Privacy: Google Spain, the Right To Be Forgotten, and the Construction of the Public Sphere' (n 186) 1015; Dohmann, 'A New Framework for Information Markets: Google Spain' (n 155) 1050; Natasha Lomas, 'How to Ask OpenAI for Your Personal Data to Be Deleted or Not Used To Train Its AIs' (*TechCrunch*, 2 May 2023) <https://techcrunch.com/2023/05/02/chatgpt-delete-data/> accessed 11 December 2023; 'OpenAI Personal Data Removal Request' (*OpenAI*) <https://share.hsforms.com/1UPy6xqxZSEqTrGDh4ywo_g4sk30> accessed 11 December 2023.

[192] *Google Spain* (n 27) 17 and 18; Kranenborg, 'Google and the Right To Be Forgotten' (n 186) 78; Kranenborg, 'Article 17. Right to Erasure ('Right to Be Forgotten')' (n 28) 480; Dohmann, 'A New Framework for Information Markets: Google Spain' (n 155) 1051.

[193] ibid.

the information to the public, ensuring that decisions around the RTBF are made in a context-sensitive manner that respects both the rights of the individual and the public interest.

The GDPR mandates that the controller who initially made personal data public must notify other controllers handling the data slated for erasure.[194] While the GDPR regards technological feasibility and cost for this obligation, the challenge extends beyond financial and technical constraints. It is compounded by the inability to control data importation and system and copies made from publicly disclosed data.[195] But the responsibility to inform third parties involves a commitment to making reasonable efforts rather than guaranteeing specific outcomes.[196] Text producers, therefore, might have the capability to notify only relevant controllers whom they share such personal data.[197] Nevertheless, text producers should adopt technical measures to exert control over personal data importation.[198] This demonstrates their commitment to fulfilling this obligation diligently.

3.3. AN ONEROUS (IF NOT IMPOSSIBLE) TASK TO EXERCISE THE RTBF IN TEXT PRODUCERS

The common adage that ”the internet never forgets” is often attributed to the challenges and costs associated with forgetting information online compared to remembering it that once information has been shared online, the internet never forgets it.[199] Combining this tendency with the advanced

[194] GDPR Article 17(2).

[195] Vrabec, *Data Subjects Rights under the GDPR* (n 123) 148 and 149; David Lindsay, 'The 'Right To Be Forgotten' in European Data Protection Law' in Normann Witzleb and others (eds), *Emerging Challenges in Privacy Law* (CUP 2014) 298-299; European Data Protection Supervisor (hereinafter ”EDPS”), 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (2012) 25; European Network and Information Security Agency (hereinafter ”ENISA”), ' The Right To Be Forgotten – Between Expectations and Practice' (18 October 2011) 8.

[196] Vrabec, *Data Subjects Rights under the GDPR* (n 123) 148; EDPS, 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (n 195) 24.

[197] Eugenia Politou and others, 'Forgetting Personal Data and Revoking Consent under the GDPR: Challenges and Proposed Solutions' (2018) 4 Journal of Cybersecurity 11; EDPS, 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (n 195) 24-25. For instance, OpenAI indicated that they share personal information with their affiliates. They are also obliged to delete the personal data at issue upon the notification by OpenAI. See 'Europe Privacy Policy' (n 107).

[198] Vrabec, *Data Subjects Rights under the GDPR* (n 123) 149; WP 29, 'Guidelines on the Implementation of the Court of the Justice of the European Union Judgment on ”Google Spain and Inc V. Agencia Espanola De Proteccion De Datos (AEPD) and Mario Costeja Gonzalez” C-131/12' (n 158) 9.

[199] CEDPO, 'Generative AI: The Data Protection Implications' (n 21) 16; Esposito, 'Algorithmic Memory and the Right to Be Forgotten' (n 112) 5; Post, 'Data Privacy and Dignitary Privacy: Google Spain, the Right To Be Forgotten, and the Construction of the Public Sphere' (n 186) 1047; Villaronga, 'Humans Forget, Machines Remember: Artificial

technology underpinning AI systems, questions about the feasibility of exact erasure in this area are raised. Even if a controller can completely remove personal data from the system, the widespread accessibility of information and difficulty in controlling unauthorized copies or retention means that duplications may still circulate.[200] The GDPR's ambiguity regarding the scope of the RTBF leaves open questions about the extent of data erasure required by the regulation. This uncertainty allows for varied implementations of the RTBF, depending on the technological infrastructure of AI systems.[201] In the landmark *Google Spain* and *Google CNIL* cases, the CJEU interpreted the RTBF as a removal from search engine results, acknowledging the possibility of implementing exact erasure in certain cases.[202] Even if the GDPR stipulates exact erasure under the RTBF, it seems reasonable to focus on restricting access to personal data rather than complete erasure, especially in instances where erasing the data is unfeasible.[203] Given the data processing mechanisms of text producers, the appropriateness of erasure techniques should be examined through individual assessment. This approach recognizes the practical limitations of data removal in the digital age while striving to uphold the principles of data protection.

As AI models expand, their ability to collect data for training their algorithms also grows, and concurrently, their storage capacities are enhanced to hoover up unlimited amounts of data that render exact erasure a complex task.[204] Considering the intricate memory structures within AI models, erasing a single data point may threaten the model's stability, as discerning the significance of each data in the training process may prove elusive, given the complicated

Intelligence and the Right To Be Forgotten' (n 29) 304-305; ENISA, 'The Right To Be Forgotten – Between Expectations and Practice' (n 195) 13; Lindsay, 'The 'Right To Be Forgotten' in European Data Protection Law' (n 195) 294; Viktor Mayer-Schonberger, *Delete: The Virtue of Forgetting in the Digital Age* (Princeton University Press 2009) 92.

[200] Lindsay, 'The 'Right To Be Forgotten' in European Data Protection Law' (n 195) 298; ENISA, 'The Right To Be Forgotten – Between Expectations and Practice' (n 195) 8-9.

[201] Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 3; Zhao, 'The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws' (n 30) 86; Villaronga, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten' (n 29) 309; ENISA, 'The Right To Be Forgotten – Between Expectations and Practice' (n 195) 7.

[202] *Google CNIL* (n 32); *Google Spain* (n 27); Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 7.

[203] Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 8.

[204] Ganguli, 'Predictability and Surprise in Large Generative Models' (n 14) 1748; Zhao, 'The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws' (n 30) 88.

characteristics of neural networks.[205] This complexity means that data removal is not just a simple task of deletion but also involves considering the potential impact on the overall system.

While there might be a temptation to equate text producers with search engines due to their occasional similarity in function, a closer look reveals their inherent differences.[206] Text producers generate responses through their internal databases, contrasting sharply with search engines that sift through the vast web for relevant information in response to user queries.[207] Despite apparent similarities, their technical designs diverge significantly.[208] Text producers prioritize providing structured responses rather than serving as comprehensive search tools. Therefore, integrating the delisting solution commonly used in search engines is not technically viable for text producers.[209]

3.4. INTERIM CONCLUSION

The realization of data's potential value sparked excessive collection and processing of data, consequently amplifying the demand for the RTBF. Although commonly associated with the *Google Spain* case, the implementation of this right did not imply making personal data inaccessible through delisting, on the contrary, it points to removing personal data from the system. Given that the RTBF is not an absolute right, a balance must be struck between individuals' right to data protection and public access to information and the freedom of expression in the context of text producers. While different grounds stipulated in the GDPR allow for the application of the RTBF in text producers, intricate technicalities make it challenging, potentially rendering

[205] Jesus L. Lobo and others, 'The Right To Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges' (IEEE Conference on Artificial Intelligence, Santa Clara, 2023) 180; Zhao, 'The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws' (n 30) 95; Villaronga, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten' (n 29) 315.

[206] For instance, ChatGPT 4 can connect to internet to receive real-time information and search a specific query of the user like search engines. See Kleinman and Radford, 'ChatGPT Can Now Access Up To Date Information' (n 31).

[207] Zhang, 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' (n 21) 6 and 8; Tristan Greene, 'ChatGPT Will Not Replace Google Search' (*The London School of Economics and Political Science*, 27 January 2023) <https://blogs.lse.ac.uk/impactofsocialsciences/2023/01/27/chatgpt-will-not-replace-google-search/> accessed 12 December 2023; Jaime Escott, '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> accessed 12 December 2023.

[208] ibid.

[209] Zhang, 'Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' (n 21) 10.

implementation difficult or even unfeasible. Therefore, both technological and legal solutions must be considered to address these limitations.

4. DESIGN PRINCIPLES AND PRACTICES THAT SHOULD BE IMPLEMENTED BY TEXT PRODUCERS TO PROMOTE THE EFFECTIVE USE OF THE RTBF

4.1. INTRODUCTION

After examining the grounds that trigger the application of the RTBF and the extent of personal data erasure mandated by the GDPR in the previous Chapter, this Chapter will analyse design principles and practices for effectively implementing the RTBF in text producers. It will weave through potential technical solutions that are discussed currently as a way out and policy proposals to refine the legal framework of the GDPR concerning the RTBF. At the heart of this analysis is the endeavour to answer the third sub-question: *What design principles and practices should be implemented by text producers to promote the effective use of the right to be forgotten?* This exploration aims to address the challenges of the current technical solutions and the need for a clearer and more pragmatic legal framework. The goal is to shed light on pathways that enable text producers to align with emerging solutions, thereby ensuring a harmonious balance between compliance and the practical application of the RTBF.

4.2.TECHNICAL SOLUTIONS

4.2.1. Machine Unlearning

Machine unlearning presents a promising step towards the precise removal of personal data from AI systems.[210] Machine unlearning serves to eliminate problematic data points from the model while preserving its overall performance and functionality.[211] Beyond the problematic approach of retraining models from the ground up to delete data, machine unlearning introduces a unique

[210] Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 12; Luciano Floridi, 'Machine Unlearning: Its Nature, Scope, and Importance for A "Delete Culture"' (2023) 36 Philosophy and Technology 5; Lobo, 'The Right To Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges' (n 205) 181; Mohammad Al-Rubaie and J. Morris Chang, 'Privacy-Preserving Machine Learning: Threats and Solutions' (2018) 17 IEEE Security and Privacy 49, 57; Veale, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (n 141) 9.

[211] ibid.

capability: the potential to selectively unlearn a data point without the need for model training.[212] This method not only enables the elimination of risks associated with the prior approach but also offers a streamlined solution.[213] Retraining a model demands extensive time, effort, and financial resources, proving to be inefficient.[214] Furthermore, the energy consumption entailed in this retraining process poses a significant environmental concern, potentially causing substantial damage to the ecosystem.[215] Given the expansive scope of data collection and processing by text producers, the potential for an overwhelming influx of RTBF requests poses a significant challenge, particularly considering the exorbitant costs and time, and adverse environmental impact associated with retraining the model for each request. Fulfilling this obligation becomes an insurmountable task, given these practical constraints. Therefore, machine unlearning circumvents these drawbacks, presenting a more targeted and environmentally mindful approach to removing data from AI systems.[216]

Yet, this technique has shortcomings due to its nascent stage; the lack of ample evidence substantiating its functionality stands as a significant drawback.[217] The complexity and unpredictable nature inherent in these training methods might hinder a clear understanding of how

[212] ibid.
[213] Luciano Floridi, 'Supporting Turstworthy AI Through Machine Unlearning' [2023] 3 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4643518> accessed 13 December 2023.
[214] Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 12; Philipp Hacker, 'Sustainable AI Regulation' (*Arxiv*, 2023) 16 <https://arxiv.org/abs/2306.00292> accessed 13 December 2023; Floridi, 'Machine Unlearning: Its Nature, Scope, and Importance for A "Delete Culture"' (n 210) 5; Thanveer Shaik and others, 'Exploring the Landscape of Machine Unlearning: A Comprehensive Survey and Taxonomy' (*Arxiv*, 2023) 2 <https://arxiv.org/abs/2305.06360> accessed 14 December 2023; Ayush Sekhari and others, 'Remember What You Want To Forget: Algorithm for Machine Unlearning' (35th Conference on Neural Information Processing Systems, December 2021) 5; Lucas Bourtoule and others, 'Machine Unlearning' (IEEE Symposium of Security and Privacy, December 2020) 1; Veale, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (n 141) 9; Yinzhi Cao and Junfeng Yang, 'Towards Making Systems Forget with Machine Unlearning' (IEEE Symposium of Security and Privacy, 2015) 463, 464.
[215] Floridi, 'Supporting Trustworthy AI Through Machine Unlearning' (n 213) 3; Karen Hao, 'Training A Single AI Model Can Emit As Much Carbon As Five Cars in Their Life Time' (*MIT Technology Review*, 6 June 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4643518> accessed 13 December 2023.
[216] OECD, 'Recommendation of the Council on Artificial Intelligence' (2023, OECD/LEGAL/0449) 7; Josh Cowls and Luciano Floridi, 'A Unified Framework of Five Principles for AI in Society' [2019] Harvard Data Science Review 5-6.
[217] Hawkins, 'A Decision-Making Process to Implement the 'Right to be Forgotten' in Machine Learning' (n 30) 12; Floridi, 'Machine Unt learning: Its Nature, Scope, and Importance for A "Delete Culture"' (n 210) 6; Bourtoule and others, 'Machine Unlearning' (n 214) 3; Fabian Pedregosa and Eleni Triantafillou, 'Announcing the First Machine Unlearning Challenge' (*Google Research*, 29 June 2023) <https://blog.research.google/2023/06/announcing-first-machine-unlearning.html> accessed 13 December 2023; Veale, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (n 141) 9.

individual data points impact the intricate web of model parameters, especially mapping the effect of a solitary data point on complex models like deep neural networks.[218] Moreover, the step-by-step evolution of training where each update builds upon the previous, compounds the difficulty, creating a complex interdependence between data and model evolution.[219] Nevertheless, it is considered that the implementation of machine unlearning to text producers could align better with the RTBF mandated by the GDPR, rather than merely rendering data inaccessible as in the *Google Spain* case.[220] Considering the limitations of this technique, an argument arises favouring embedding solutions within models during their initial design phase as potentially the most viable method.[221]

4.2.2. Privacy by Design

Beyond remedies applied after designing models, controllers may consider nesting solutions in models prior to training as a means for the efficient execution of the RTBF. To guarantee adherence to data protection principles and safeguard the rights and freedoms of data subjects at model design and processing phases, the GDPR introduces the principles of data protection by design.[222] Accordingly, controllers must enact appropriate technical and organizational measures during processing, ensuring incorporation of privacy-preserving designs to meet GDPR compliance.[223] The GDPR gives hints about what constitutes appropriate measures that are, among others, pseudonymisation, data minimisation, and transparency, encryption, and anonymisation.[224]

[218] Lobo, 'The Right To Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges' (n 205) 181; Pedregosa and Triantafillou, 'Announcing the First Machine Unlearning Challenge' (n 217); Bourtoule, 'Machine Unlearning' (n 214) 3-4.

[219] ibid.

[220] Floridi, 'Machine Unlearning: Its Nature, Scope, and Importance for A "Delete Culture"' (n 210) 6.

[221] Tea Mustac, 'An Elephant Never Forgets and Neither Does ChatGPT' (JD Supra, 13 April 2023) <https://www.jdsupra.com/legalnews/an-elephant-never-forgets-and-neither-7307484/> accessed 13 December 2023.

[222] GDPR Article 25 in conjunction with Recital 78; ICO, 'Data Processing by Design and Default' (19 May 2023) 22; Lee A. Bygrave, 'Article 25. Data Protection by Design and by Default' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 573; WP 29, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (20 October 2020, Version 2.0) 4; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 72.

[223] WP 29, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (n 222) 5; EDPS, 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (n 195) 29.

[224] GDPR Article 25(1) in conjunction with Recital 78; ICO, 'Data Processing by Design and Default' (n 222) 27; Bygrave, 'Article 25. Data Protection by Design and by Default' (n 222) 577-578.

Apart from transparency which must be in place at every stage of the processing in text producers, currently discussed methods are pseudonymisation and anonymisation.[225]

The anonymisation technique renders the identification of a data subject through the data or alongside other data impossible.[226] As per the GDPR's definition of personal data, information must pertain to an identified or identifiable natural person to qualify as personal data.[227] Therefore, once data is anonymised and no longer traceable to any individual, it loses the attribute of personal data, aligning with its inability to be linked to any specific person.[228] Anonymisation is considered that anonymisation stands as an alternative to the RTBF; nonetheless, it bears the inherent risk of potential re-identification.[229] The threshold of the effectiveness of the anonymisation method should be to ensure that all feasible and reasonably expected means to identify the data subject become unfeasible.[230] There is no foolproof anonymisation technique that ensures complete certainty of preventing re-identification.[231] Re-identification might occur through attacks where attackers could use existing data and background information to identify individuals, akin to the de-anonymisation attack seen in the Netflix Prize case.[232] Implementing this method in text

[225] WP 29, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (n 222) 13; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 72.

[226] ibid 58; Agencia Espanola Proteccion Datos and EDPS, '10 Misunderstanding Related To Anonymisation' 2 <https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf> accessed 14 December 2023; WP 29, 'Opinion 05/2014 on Anonymisation Techniques' (10 April 2014, 0829/14/EN WP 216) 5.

[227] GDPR Article 4(1) in conjunction with Recital 26.

[228] Emily M. Weitzenboeck and others, 'The GDPR and Unstructured Data: Is Anonymization Is Possible' (2022) 12 International Data Privacy Law 184, 189; Bygrave and Tosoni, 'Article 4(1). Personal Data' (n 42) 105; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (n 41) 21.

[229] WP 29, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (n 222) 13; Agencia Espanola Proteccion Datos and EDPS, '10 Misunderstanding Related To Anonymisation' (n 226) 4; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 59; ENISA, 'Privacy and Data Protection by Design – from Policy to Engineering' (December 2014) 37; WP 29, 'Opinion 4/2007 on the Concept of Personal Data' (n 41) 21.

[230] Weitzenboeck, 'The GDPR and Unstructured Data: Is Anonymization Is Possible' (n 228) 192; WP 29, 'Opinion 05/2014 on Anonymisation Techniques' (n 226) 5.

[231] Weitzenboeck, 'The GDPR and Unstructured Data: Is Anonymization Is Possible' (n 228) 191; Agencia Espanola Proteccion Datos and EDPS, '10 Misunderstanding Related To Anonymisation' (n 226) 7; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 59.

[232] Agencia Espanola Proteccion Datos and EDPS, '10 Misunderstanding Related To Anonymisation' (n 226) 4; ENISA, 'Privacy and Data Protection by Design – from Policy to Engineering' (n 229) 37; Ines Ortega-Fernandez and others, 'Large Scale Data Anonymisation for GDPR Compliance' in Dimosthenis Kyriazis and John Soldatos (eds), *Big Data and Artificial Intelligence in Digital Finance* (Springer 2022) 328; ICO, 'Anonymisation: Managing Data Protection Risk Code of Practice' (November 2012) 16; 'Facing Lawsuit and Investigation, Netflix Cancels Contest' (*Lexology*, 7 April 2010) <https://www.lexology.com/library/detail.aspx?g=21ae0d4e-30d7-40f5-b753-efa7a9cb6d5b> accessed 14 December 2023. For example, in the Netflix Prize case, one argued that the disclosure of her rental history, particularly the viewing of gay-themed movies, inadvertently uncovered her sexual orientation. The disclosure of sensitive information under Article 9 of the GDPR compounds the inherent risk.

producers may be challenging due to their vulnerability to security breaches, potentially resulting in re-identification.[233] In this regard, the randomisation method comes to the forefront as a remedy for the implementation of more effective implementation.[234]

In the randomisation technique, the manipulation of the veracity of the data is conducted to diminish the direct association between the data and the data subject.[235] Under randomisation, differential privacy stands out as the prominent approach, designing algorithms to yield outcomes without compromising the privacy of natural persons.[236] The aim is to safeguard personal data by introducing randomness into query responses, thereby making it difficult to identify any specific data point's influence on the results.[237] This method is regarded as demonstrating substantial success compared with other methods, thereby a promising way to guarantee data protection.[238] Nevertheless, a concern regarding integrating differential privacy into the GDPR's version of anonymisation emerges as the data controller holds the original, non-anonymised data.[239] Furthermore, while differential privacy aims to prevent the identification of data points, there remains an inherent risk of misidentification through attacks.[240]

4.3. POLICY PROPOSAL

[233] Kovacs, 'Simple Attack Allowed Extraction of ChatGPT Training Data' (n 145); Chilton, 'The New Risks ChatGPT Poses to Cybersecurity Data' (n 145).

[234] WP 29, 'Opinion 05/2014 on Anonymisation Techniques' (n 226) 12.

[235] ibid.

[236] Villaronga, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten' (n 29) 318; Rachel Cummings and Deven Desai, 'The Role of Differential Privacy in GDPR Compliance' [2018] Proceedings of the Conference on Fairness Accountability, and Transparency 20, 21; Mike Hintze, 'Viewing the GDPR through A De-identification Lens: A Tool for Compliance, Clarification, and Consistency' (2018) 8 International Data Privacy Law 86, 87; ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 33) 72; ENISA, 'Privacy and Data Protection by Design – from Policy to Engineering' (n 229) 37; 'Differential Privacy' (*Harvard University*) <https://privacytools.seas.harvard.edu/differential-privacy> accessed 14 December 2023.

[237] WP 29, 'Opinion 05/2014 on Anonymisation Techniques' (n 226) 15; Abigail Goldsteen and others, 'Anonymizing Machine Learning Models' (ESORICS 2021 International Workshops, Germany, October 2021) 125; Raina Gandhi and Amritha Jayanti, 'Differential Privacy' (2021, Harvard Kennedy School) 2; Cynthia Dwork and others, 'Differential Privacy in Practice: Expose Your Epsilons!' (2019) 9 Journal of Privacy and Confidentiality 4.

[238] Gandhi and Jayanti, 'Differential Privacy' (n 237) 3; Cummings and Desai, 'The Role of Differential Privacy in GDPR Compliance' (n 236) 21.

[239] Julian Hotzel, 'Differential Privacy and the GDPR' (2019) 5 European Data Protection Law Review 184, 194-195.

[240] ibid 195.

While technical solutions bring a slight relief for data protection concerns, they do not provide full protection in terms of the GDPR, presenting inherent limitations. This, combined with the GDPR's vague directives, creates uncertainty around effectively implementing the RTBF.[241] It is clear that legal adjustments are necessary in this context, specifically within the RTBF framework of the GDPR. The RTBF concept revolves around removing or deleting personal data under specific conditions.[242] However, it lacks a defined threshold for the extent of data erasure.[243] While ”removal” suggests clearing data, the online landscape poses a significant challenge; once information is shared online, completely removing it from the internet becomes nearly impossible.[244] This complexity is particularly true in the realm of machine learning. Moreover, the literal meaning of ”forgetting” creates a misconception, as people believe data can be erased entirely under the RTBF, a notion at odds with the practical impossibility of completely erasing online information.[245] This discrepancy between expectation and reality makes the RTBF concept misleading and impractical to implement in real-life scenarios.[246]

Considering the difficulties in removing data from text producers, it seems appropriate to adjust the RTBF concept to better reflect the current technology-driven world. Subdividing it into sections incorporating available remedies, such as delisting, machine unlearning, and differential privacy could be a step forward. Adjusting the title of the RTBF to represent a more accurate representation could help align with the realities of modern technology. Shifting the focus from complete erasure to controlling or restricting access to personal data better acknowledges the challenges of removing information entirely from the digital world. this revised title would reflect a more realistic approach to data management within the GDPR framework, emphasising the

[241] Zhao, ‘The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws’ (n 30) 107; Christopher Kuner and others, ‘Machine Learning with Personal Data: Is Data Protection Law Smart Enough To Meet the Challenge?’ (2017) 7 International Data Privacy Law 1; EDPS, ‘Opinion of the European Data Protection Supervisor on the Data Protection Reform Package’ (n 195) 23.

[242] Ausloos, *The Right To Erasure in EU Data Protection Law* (n 151) 114.

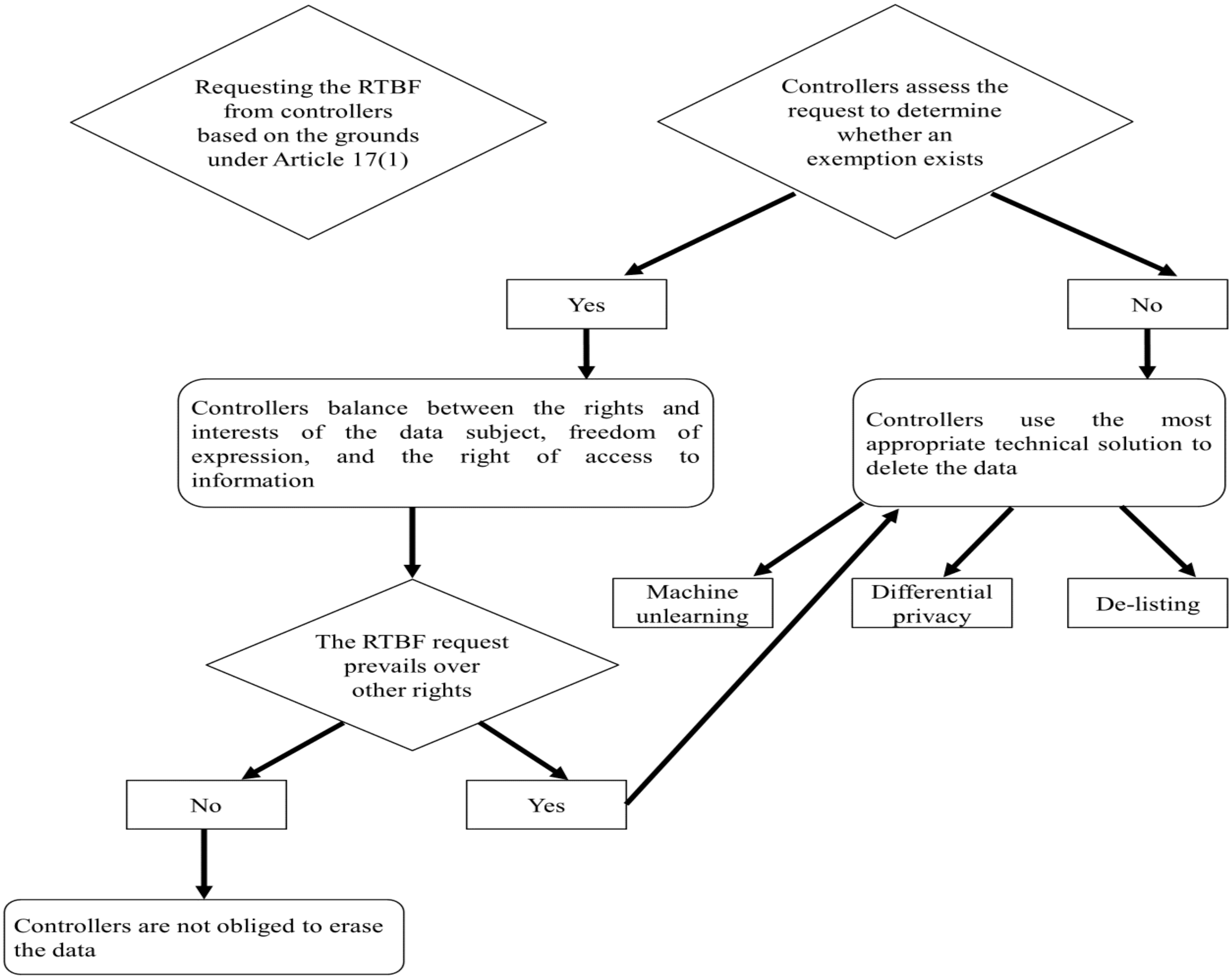
[243] Hawkins, ‘A Decision-Making Process to Implement the ‘Right to be Forgotten’ in Machine Learning’ (n 30) 3; EDPS, ‘Opinion of the European Data Protection Supervisor on the Data Protection Reform Package’ (n 195) 23.

[244] Lindsay, ‘The ‘Right To Be Forgotten’ in European Data Protection Law’ (n 194) 298; ENISA, ‘The Right To Be Forgotten – Between Expectations and Practice’ (n 195) 8-9.

[245] Bert-Jaap Koops, ‘The Trouble with European Data Protection Law’ (Tilburg Law School Legal Studies Research Paper Series No 04/2015) 11; Paulan Korenhof, ‘Forgetting Bits and Pieces: An Exploration of the ”Right To Be Forgotten” As Implementation of ”Forgetting” in Online Memory Processes’ (IFIP International Summer School, The Netherlands, June 2013) 118; ICO, ‘Proposed New EU General Data Protection Regulation: Article-by-article Analysis Paper’ (12 February 2013) 23.

[246] ibid.

regulation of access rather than the misleading notion of complete removal. Crucially, clear guidelines from the EDPB and national data protection authorities are needed to effectively implement these emerging solutions. This approach will dispel illusions about data removal and provide individuals with a more realistic perspective.


Requesting the RTBF from controllers based on the grounds under Article 17(1)
Controllers assess the request to determine whether an exemption exists
Yes
No
Controllers balance between the rights and interests of the data subject, freedom of expression, and the right of access to information
Controllers use the most appropriate technical solution to delete the data
Machine unlearning
Differential privacy
De-listing
The RTBF request prevails over other rights
No
Yes
Controllers are not obliged to erase the data

The RTBF Implementation Flow

4.4. INTERIM CONCLUSION

To effectively enforce the RTBF, two promising technical solutions, namely machine unlearning and differential privacy, offer potential remedies. Machine unlearning, though still in its infancy, lacks adequate evidence to gauge its effectiveness in RTBF implementation, whereas differential privacy has demonstrated success despite its shortcomings. Although technical remedies play a crucial role, embracing a promising legal framework becomes imperative, especially considering the practical and technological constraints of the RTBF.

5. CONCLUSION

This Thesis aims to delve deeper into the possible implementation of the RTBF in text producers. It seeks to address a fundamental question: *How to exercise the right to be forgotten in text producers?*

The research first examines: *In what ways do text producers impact the legal aspects of the GDPR in exercising the right to be forgotten?* As the relationship between big data, AI, and GDPR compliance in text-producing AI models is complex and multifaceted. Big data fuels AI's capabilities, but its incorporation into text producers raises GDPR concerns due to the inclusion of personal data. The broad definition of personal data under the GDPR poses challenges in ensuring compliance, especially when dealing with vast and varied datasets. Text producers gather data through web scraping, user interactions, and generated outputs, potentially containing personal information. Establishing lawful bases for data processing, especially with training data, becomes critical, and transparency in processing methods remains essential, considering the accountability principle. Furthermore, security vulnerabilities pose a significant risk, complicating compliance with data breach notification requirements. ultimately, aligning text producers with the GDPR entails navigating legal grounds, ensuring transparency, and addressing security concerns.

Subsequently, it endeavours to provide an answer to the following: *To what extent are text producers obliged to erase personal data under the GDPR, considering their technological infrastructure and capabilities?* The examination is expanded to weighing legal obligations against

AI system practicalities concerning erasing personal data for text producers. The RTBF traces its origins from the *Google Spain* case, however, the criticisms emphasise the challenge of actual data erasure rather than just delisting. The application of the RTBF might be triggered by withdrawal of consent which provides a basis for data processing, the lack of legal basis, the right to object to processing, and inaccurate data. Considering the non-absolute nature of this right, balancing it with freedom of expression poses complexities. Moreover, implementing the RTBF in text producers faces technological hurdles, making exact data erasure challenging. Overall, the necessity of the RTBF should be acknowledged while highlighting complexities and limitations in its practical application for text producers due to technological and legal intricacies.

Finally the research analyses: *What design principles and practices should be implemented by text producers to promote the effective use of the right to be forgotten?* It explores technical solutions and policy adjustments for text producers implementing the RTBF within the GDPR framework. Technical remedies like machine unlearning and privacy by design offer solutions but grapple with complexities and drawbacks. Policy suggestion proposes subdividing the RTBF into sections that encompass available remedies and regulating it broadly under the concept of restricting access to personal data, rather than removing or erasing them. This approach acknowledges the limitations of complete data erasure in the digital landscape and aims to provide clearer expectations for data subjects while maintaining GDPR compliance. Broadly, it highlights the need for a more realistic approach to data erasure and suggests regulatory guidance to navigate evolving technological landscapes effectively.

BIBLIOGRAPHY

Primary Sources

TABLE OF CASES

European Court of Justice

Case 507/17 Google LLC v the Commission Nationale ne l'informatique et des Libertés (CNIL) [2019]

Case 507/17 Google LLC v the Commission Nationale de l'informatique et des Libertés (CNIL) [2019], Opinion of AG Szpunar

Case C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González [2014]

Case C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González [2014], Opinion of AG Jääskinen

National Data Protection Authorities

CNIL, 'Restricted Committee Deliberation No. SAN-2022-019 of 17 October 2022 concerning Clearview AI

CNIL, SAN-2022-019, 17 October 2022 <https://www.cnil.fr/en/facial-recognition-20-million-euros-penalty-against-clearview-ai> accessed 25 November 2023

Garante Per La Protezione Dei Dati Personali (Italian Data Protection Authority), '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> accessed 11 November 2023

TABLE OF LEGISLATION

European Union Law

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 (General Data Protection Regulation) [2016] OJ L 119/1

Legislative Proposals

Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (AI Act)' COM (2021) 206 final

Secondary Sources

OFFICIAL MATERIAL

European Commission

Commission, 'A Comprehensive Approach on Personal Data Protection in the European Union' COM (2010) 609 final 4

Commission, 'AI Watch Defining Artificial Intelligence' 2020 (EUR 30117)

Commission, 'Historical Evolution of Artificial Intelligence' 2020 (EUR 30221)

Commission, 'Towards a Thriving Data-driven Economy' COM (2014)

European Parliament

European Parliament, 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' June 2020 (PE 641.530)

European Data Protection Board

EDPB, 'Decision by the Austrian SA against Clearview AI Infringements of Article 5, 6, 9, 27' (*EDPB*, 12 May 2023) <https://edpb.europa.eu/news/national-news/2023/decision-austrian-sa-against-clearview-ai-infringements-articles-5-6-9-27_en> accessed 25 November 2023

EDPB, 'Facial Recognition: Italian SA Fines Clearview AI EUR 20 Million' (*EDPB*, 10 May 2023) <https://edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en> accessed 25 November 2023

EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (4 May 2020)

EDPB, 'Guidelines 5/2019 on the Criteria of the Right To Be Forgotten in the Search Engines Cases under the GDPR' (2 December 2019)

EDPB Resolves Dispute on Transfers by Meta and Creates Task Force on ChatGPT' (EDPB, 13 April 2023) <https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en> accessed 5 December 2023

EDPB, 'Guidelines 3/2018 on the Territorial Scope of the GDPR' (12 November 2019)

Others

Agencia Espanola Proteccion Datos and EDPS, '10 Misunderstanding Related To Anonymisation' 2 <https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf> accessed 14 December 2023

EDPS, 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (2012)

ENISA, 'Privacy and Data Protection by Design – from Policy to Engineering' (December 2014)

ENISA, 'The Right To Be Forgotten – Between Expectations and Practice' (18 October 2011)

Confederation of European Data Protection Organizations, 'Generative AI: The Data Protection Implications' (2023) CEDPO AI Working Group

Congressional Research Service, 'Generative Artificial Intelligence and Data Privacy: A Primer' (23 May 2023)

Council of European Union, 'ChatGPT in the Public Sector – Overhyped or Overlooked?' (24 April 2023)

Hayes G, 'Have You Read Your Privacy Policies' (Privacy Commissioner, 16 August 2019) <https://www.privacy.org.nz/blog/have-you-read-your-privacy-policies/> accessed 26 November 2023

ICO, 'Anonymisation: Managing Data Protection Risk Code of Practice' (November 2012)

ICO, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (2017)

ICO, 'Data Processing by Design and Default' (19 May 2023)

ICO, 'Guide to the General Data Protection Regulation (GDPR)' 14 October 2022

ICO, ‘Joint Statement on Data Scraping and the Protection of Privacy’ (24 August 2023) 1 <https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf> accessed 15 December 2023

ICO, ‘Lawful Basis for Processing’ (19 May 2023)
ICO, ‘Lawful Basis for Processing: Legitimate Interests’ (22 March 2018)

ICO, ‘Lawful Basis for Processing: Special Category Data’ 17 October 2022

ICO, ‘Proposed New EU General Data Protection Regulation: Article-by-article Analysis Paper’ (12 February 2013)

Norwegian Consumer Council, ‘Ghost in the Machine’ (June 2023)

OECD, ‘AI Language Models: Technological, Socio-Economic and Policy Considerations’ (April 2023, OECD Digital Economy Papers No 352)

OECD, ‘Harnessing the Power of AI and Emerging Technologies’ 15 November 2022 (DSTI/CDEP(2022)14/FINAL)

OECD, ‘Initial Policy Considerations for Generative Artificial Intelligence’ September 2023 (No 1)

OECD, ‘Recommendation of the Council on Artificial Intelligence’ (2023, OECD/LEGAL/0449)

UNESCO, ‘Global Toolkit on AI and the Rule of Law for the Judiciary’ 2023 (CI/DIT/2023)

WP 29, ‘Advice Paper on Special Categories of Data’ (20 April 2011)

WP 29, ‘Guidelines 4/2019 on Article 25 Data Protection by Design and by Default’ (20 October 2020, Version 2.0)

WP 29, ‘Guidelines on the Implementation of the Court of the Justice of the European Union Judgment on ”Google Spain and Inc V. Agencia Espanola De Proteccion De Datos (AEPD) and Mario Costeja Gonzalez” C-131/12’ (26 November 2014, 14/EN WP 225)

WP 29, ‘Guidelines on Personal Data Breach Notification under Regulation 2016/679’ (6 February 2018, 18/EN WP250rev.01)

WP 29, ‘Guidelines on Transparency under Regulation 2016/679’ (11 April 2018, 17/EN WP260 rev.01)

WP 29, ‘Opinion 05/2014 on Anonymisation Techniques’ (10 April 2014, 0829/14/EN WP 216)

WP 29, ‘Opinion 4/2007 on the Concept of Personal Data’ (20 June 2007, 01248/07/EN WP 136)
WP 29, ‘Opinion 1/2008 on Data Protection Issues Related to Search Engines’ (4 April 2008, WP 148 00737/EN)

WP 29, ‘Opinion 15/2011 on the Definition of Consent’ (13 July 2011, 01197/11/EN WP187)

WP 29, ‘Opinion 3/2012 on Developments in Biometric Technologies’ (27 April 2012, WP 193, 00720/12/EN)

WP WP29, ‘Opinion 06/2014 on the Notion of Legitimate Interest of the Data Controller under Article 7 of Directive 95/46/EC’ (WP 217, 844/14/EN, 9 April 2014)

29, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse’ (5 June 2013, 1021/00/EN WP 207)

WP 29, ‘Opinion 3/2010 on the Principle of Accountability’ (13 July 2010, 00062/10/EN WP 173)

WP 29, ‘Opinion 03/2013 on Purpose Limitation’ (2 April 2013, 00569/13/EN WP 203)

BOOKS

Altobelli C and others, ‘To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR’ in Jean Herveg (eds), *Deep Driving into Data Protection* (Larcier 2021)

Ausloos J, *The Right To Erasure in EU Data Protection Law* (OUP 2020)

Bermudez J.L, *Machine Learning: From Expert Systems to Deep Learning* (3rd edn, CUP 2020)

Burton C, ‘Article 34. Communication of A Personal Data Breach to the Data Subject’ in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Bygrave L.A, ‘Article 25. Data Protection by Design and by Default’ in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Bygrave L.A and Tosoni L, 'Article 14(4). Biometric Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Bygrave L.A and Tosoni L, 'Article 4(1). Personal Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Bygrave L.A and Tosoni L, 'Article 4(2). Processing' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

De Terwangne C, 'Article 5. Principles relating to Processing of Personal Data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Hacker P, 'AI Regulation in Europe: From the AI Act to Future Regulatory Challenges' in Ifeoma Ajunwa and Jeremias Adams-Prassl (eds), *Oxford Handbook of Algorithmic Governance and the Law* (OUP 2023)

Hutchinson T, 'Doctrinal Research: Researching the Jury' in Dawn Watkins and Mandy Burton (eds), *Research Methods in Law* (Routledge 2017)

Hutchinson T, *Research and Writing in Law* (4th edn, Lawbook Co 2018)

Kaplan J, *Artificial Intelligence: What Everyone Needs to Know* (OUP 2016)

Kelleher J.D and Tierney B, *Data Science* (MIT Press 2018)

Kotschy W, 'Article 6. Lawfulness of Processing' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Kranenborg H, 'Article 17. Right to Erasure ('Right to Be Forgotten')' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Lambert P, *The Right To Be Forgotten* (2nd edn, Bloomsbury Professional 2022)

Lindsay D, 'The 'Right To Be Forgotten' in European Data Protection Law' in Normann Witzleb and others (eds), *Emerging Challenges in Privacy Law* (CUP 2014)

Lindsay D, 'The 'Right To Be Forgotten' by Search Engines under Data Privacy Law: A Legal and Policy Analysis of the Costeja Decision' in Andrew T. Keyton (ed), *Comparative Defamation and Privacy Law* (CUP 2016)

Mayer-Schönberger V, *Delete: The Virtue of Forgetting in the Digital Age* (Princeton University Press 2009)

Ortega-Fernandez I and others, 'Large Scale Data Anonymisation for GDPR Compliance' in Dimosthenis Kyriazis and John Soldatos (eds), *Big Data and Artificial Intelligence in Digital Finance* (Springer 2022)

Paal B.P, 'Artificial Intelligence as A Challenge for Data Protection Law and Vice Versa' in Silja Voeneky and others (eds), *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (CUP 2022)

Polcak R, 'Article 12. Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Scopino G, *Algo Bots and the Law* (CUP 2020)

Sharma S, *Data Privacy and GDPR Handbook* (Wiley 2020)

Svantesson D.J.B, 'Article 3. Territorial Scope' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Vrabec H.U, *Data Subjects Rights under the GDPR* (Oxford 2021)

Zanfir-Fortuna G, 'Article 14. Information To Be Provided Where Personal Data Have Not Been Obtained from the Data Subject' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

Zanfir-Fortuna G, 'Article 21. Right To Object' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020)

ARTICLES

Al-Rubaie M and Chang J.M, 'Privacy-Preserving Machine Learning: Threats and Solutions' (2018) 17 IEEE Security and Privacy 49

Ausloos J and Kuczerawy A, 'From Notice-and-Takedown to Notice-and Delist: Implementing Google Spain' (2016) 14 Colorado Technology Law Journal 291

Burri M and Schar R, ‘The Reform of the EU Data Protection Framework: Outlining Key Changes and Assessing Their Fitness for A Data-Driven Economy’ Journal of Information Policy (2016) 6 479

Cao Y and Yang J, ‘Towards Making Systems Forget with Machine Unlearning’ (IEEE Symposium of Security and Privacy, 2015) 463

Celeste E and Fabbrini F, ‘The Right To Be Forgotten In the Digital Age: The Challenges of Data Protection Beyond Borders’ (2020) 21 German Law Journal 55

Chahal A and Gulia P, ‘Machine Learning and Deep Learning’ [2019] International Journal of Innovative Technology and Exploring Engineering 4910

Cooper A.F and others, ‘Report of the 1st Workshop on Generative AI and Law’ (*Arxiv*, 2023) 11 <https://arxiv.org/pdf/2311.06477.pdf> accessed 30 November 2023

Cowls J and Floridi L, ‘A Unified Framework of Five Principles for AI in Society’ [2019] Harvard Data Science Review

Cummings R and Desai D, ‘The Role of Differential Privacy in GDPR Compliance’ [2018] Proceedings of the Conference on Fairness Accountability, and Transparency 20

Custers B and Ursula H, ‘Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection’ (2016) 6 International Data Privacy Law 4

Custers B and Ursic H, ‘Legal Barriers and Enablers to Big Data Reuse’ (2016) 2 European Data Protection Law Review 208

De Hert P and Kamara I, ‘Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach’ (2018) 4 Brussels Privacy Hub 321

Dohmann I.S, ‘A New Framework for Information Markets: Google Spain’ (2015) 52 Common Market Law Review 1033

Duncan N and Hutchinson T, ‘Defining and Describing What We Do: Doctrinal Legal Research’ (2012) 17 Deakin Law Review 101

Dwork C and others, ‘Differential Privacy in Practice: Expose Your Epsilons!’ (2019) 9 Journal of Privacy and Confidentiality

Elena Esposito, ‘Algorithmic Memory and the Right to Be Forgotten’ (2017) 4 Big Data & Society 1

Felzmann H and others, ‘Transparency You Can Trust: Transparency Requirement for Artificial Intelligence Between Legal Norms and Contextual Concerns’ (2019) 6 Big Data & Society

Friesen J, ‘The Impossible Right To Be Forgotten’ (2021) 47 Rutgers Computer & Technology Law Journal 173

Floridi L, ‘Machine Unlearning: Its Nature, Scope, and Importance for A ”Delete Culture”’ (2023) 36 Philosophy and Technology

Floridi L, ‘Supporting Trustworthy AI Through Machine Unlearning’ (2023) 3 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4643518> accessed 13 December 2023

Flyverbom M and others, ‘The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business’ (2019) 58 Business and Society 3

Gandhi R and Jayanti A, ‘Differential Privacy’ (2021, Harvard Kennedy School)

Gaurav Pathak, ‘Manifestly Made Public: Clearview and GDPR’ (2022) 8 European Data Protection Law Review 419

Hacker P, ‘A Legal Framework for AI Training Data-from First Principles to the Artificial Intelligence Act’ (2021) 13 Law, Innovation and Technology 257

Hacker P, ‘Sustainable AI Regulation’ (*Arxiv*, 2023) <https://arxiv.org/abs/2306.00292> accessed 13 December 2023

Hintze M, ‘Viewing the GDPR through A De-identification Lens: A Tool for Compliance, Clarification, and Consistency’ (2018) 8 International Data Privacy Law 86

Hosain T and others, ‘Path to Gain Functional Transparency in Artificial Intelligence with Meaningful Explainability’ (2023) 3 Journal of Metaverse 166

Hotzel J, ‘Differential Privacy and the GDPR’ (2019) 5 European Data Protection Law Review 184

Jones K, ‘AI Governance and Human Rights: Resetting the Relationship’ (*Chatnam House*, January 2023) 24 <https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights> accessed 15 December 2023

Kiriiak O.V, ‘The Right To Be Forgotten: Emerging Legal Issues’ (2021) 3 Review of European and Comparative Law 27

Kokott J and Sobotta C, ‘The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR’ (2013) 3 International Data Privacy Law 222

Koops B, ‘The Trouble with European Data Protection Law’ (Tilburg Law School Legal Studies Research Paper Series No 04/2015)

Kranenborg H, ‘Google and the Right To Be Forgotten’ (2015) 1 European Data Protection Law Review 70

Kuner C and others, ‘Machine Learning with Personal Data: Is Data Protection Law Smart Enough To Meet the Challenge?’ (2017) 7 International Data Privacy Law

Lagone L, ‘The Right To Be Forgotten: A Comparative Analysis’ [2012] <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2229361> accessed 10 December 2023

Lee K and others, ‘AI and Law: The Next Generation’ 2023 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4580739> accessed 24 November 2023

Lewis C, ‘The Need for a Legal Framework to Regulate the Use of Artificial Intelligence’ (2022) 47 University of Dayton Law Review 285

Luciano Floridi, ‘Supporting Turstworthy AI Through Machine Unlearning’ [2023] <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4643518> accessed 13 December 2023

McDonagh M and Paterson M, ‘Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data’ (2018) 44 Monash University Law Review 1

Meszaros J and others, ‘ChatGPT: How Many Data Protection Principles Do You Comply with?’ (2023) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4647569> accessed 4 December 2023

Mpame M.E and Niedermeier R, 'Processing Personal Data under Article 6(f) of the GDPR: The Concept of Legitimate Interest' (2019) 3 International Journal for the Data Protection Officer, Privacy Officer and Privacy Counsel 18

Nagy P and Neff G, 'Talking to Bots: Symbiotic Agency and the Case of Tay' (2016) 10 International Journal of Communication 4915

Padova Y, 'Is the Right To Be Forgotten A Universal Regional, or 'Glocal' Right?' (2019) 9 International Data Privacy Law 15

Parks A.M, 'Unfair Collection: Reclaiming Control of Public Available Personal Information from Data Scrapers' (2022) 120 Michigan Law Review 913

Politou E and others, 'Forgetting Personal Data and Revoking Consent under the GDPR: Challenges and Proposed Solutions' (2018) 4 Journal of Cybersecurity

Post R.C , 'Data Privacy and Dignitary Privacy: Google Spain, the Right To Be Forgotten, and the Construction of the Public Sphere' (2018) 67 Duke Law Journal 981

Propp K, 'Introductory Note To Google LLC V. Commission Nationale de l'informatique et des Libertés (CNIL) and Eva Glawischnig-Piesczek V. Facebook Ireland LTD. (C.J.E.U.)' (2020) 59 The Maerican Society of International Law 428

Sebastian G, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (2023) 15 International Journal of Security and Privacy in Pervasive Computing

Shaik T and others, 'Exploring the Landscape of Machine Unlearning: A Comprehensive Survey and Taxonomy' (*Arxiv*, 2023) <https://arxiv.org/abs/2305.06360> accessed 14 December 2023

Solaiman I and others, 'Evaluating the Social Impact of Generative AI Systems in Systems and Society' (*Arxiv*, 2023) <https://arxiv.org/pdf/2306.05949.pdf> accessed 1 December 2023

Trigo P, 'Can Legitimate Interest Be An Appropriate Lawful Basis for Processing Artificial Intelligence Training Datasets' (2023) 48 Computer Law & Security Review

Turing A.M, 'Computing Machinery and Intelligence' (1950) 236 Mind 433

Turing A.M, 'Intelligent Machinery, A Heretical Theory' (1996) 4 Philosophia Mathematica 256

Veale M and others, ‘Algorithms That Remember: Model Inversion Attacks and Data Protection Law’ [2018] 376 Philosophical Transactions of the Royal Society

Villaronga E.F and others, ‘Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten’ (2018) 34 Computer Law & Security Review 304

Wach K and others, ‘The Dark Side of Generative Artificial Intelligence: A Critical Analysis of Controversies and Risks of ChatGPT’ (2023) 11 Entrepreneurial Business and Economics Review 7

Wachter S, ‘Data Protection in the Age of Big Data’ (2019) 2 Nature Electronics

Weitzenboeck E.M and others, ‘The GDPR and Unstructured Data: Is Anonymization Is Possible’ (2022) 12 International Data Privacy Law 184

Welborn A, ‘ChatGPT and Fake Citations’ (*Duke University*, 9 March 2023) <https://blogs.library.duke.edu/blog/2023/03/09/chatgpt-and-fake-citations/> accessed 7 December 2023

Wodi A.A, ‘The EU General Data Protection Regulation (GDPR): Five Years After and the Future of Data Privacy Protection in Review’ [2023] <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4601142> accessed 10 December 2023

Wolf C, ‘Impact of the CJEU’s Right To Be Forgotten’ (2014) 21 Maastricht Journal of European and Comparative Law 547

Wu X and others, ‘Unveiling Security, Privacy, and Ethical Concerns of ChatGPT’ [2023] Journal of Information and Intelligence

Zhang D and others, ‘Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions’ (*Arxiv*, 2023) <https://arxiv.org/pdf/2307.03941.pdf> accessed 10 November 2023

Zhao R and others, ‘Can ChatGPT-like Generative Models Guarantee Factual Accuracy? On the Mistakes of New Generation Search Engines’ (*Arxiv*, 2023) <https://arxiv.org/pdf/2304.11076.pdf> accessed 1 December 2023

Zhao Z, ‘The Application of the Right to be Forgotten in the Machine Learning Context: From the Perspective of European Laws’ (2022) 31 Catholic University Journal of Law and Technology 73

Conference Papers

Bourtoule L and others, ‘Machine Unlearning’ (IEEE Symposium of Security and Privacy, December 2020)

Ganguli D and others, ‘Predictability and Suprise in Large Generative Models’ (Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, 2022)

Goldsteen A and others, ‘Anonymizing Machine Learning Models’ (ESORICS 2021 International Workshops, Germany, October 2021)

Hacker P and others, ‘Regulating ChatGPT and other Large Generative AI Models’ (ACM Conference on Fairness, Accountability, and Transparency, June 2023)

Hawkins K and others, ‘A Decision-Making Process to Implement the ‘Right to be Forgotten’ in Machine Learning’ (Annual Privacy Forum, France, June 2023)

Korenhof P, ‘Forgetting Bits and Pieces: An Exploration of the ”Right To Be Forgotten” As Implementation of ”Forgetting” in Online Memory Processes’ (IFIP International Summer School, The Netherlands, June 2013)

Seinen W and others, ‘Compatibility as a Mechanism for Responsible Further Processing of Personal Data’ (6th Annual Privacy Forum, Barcelona, June 2018)

Sekhari A and others, ‘Remember What You Want To Forget: Algorithm for Machine Unlearning’ (35th Conference on Neural Information Processing Systems, December 2021)

Lobo J.L and others, ‘The Right To Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges’ (IEEE Conference on Artificial Intelligence, Santa Clara, 2023)

WEBSITES AND BLOGS

‘All about Google Bard’ (*Medium*, 10 July 2023) <https://ip-specialist.medium.com/all-about-google-bard-aae73b5534f3> accessed 24 November 2023

Bard FAQ’ <https://bard.google.com/faq?hl=en> accessed 24 November 2023

Bhasker S and others, ‘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> accessed 8 November 2023

‘ChatGPT Can Now See, Hear, and Speak’ (*OpenAI*, 25 September 2023) <https://openai.com/blog/chatgpt-can-now-see-hear-and-speak> accessed 3 December 2023

Chilton J, ‘The New Risks ChatGPT Poses to Cybersecurity’ Data’ (*Harvard Business Review*, 21 April 2023) < https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity> accessed 3 December 2023

Chui M and others, ‘The Economic Potential of Generative AI: The Next Productivity Frontier’ (*McKinsey & Company*, 14 June 2023) 18 and 24 <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/> accessed 15 December 2023

‘Dall-E 3 Is Now Available in ChatGPT Plus and Eterprise’ (*OpenAI*, 19 October 2023) <https://openai.com/blog/dall-e-3-is-now-available-in-chatgpt-plus-and-enterprise> accessed 3 December 2023

‘Differential Privacy’ (*Harvard University*) <https://privacytools.seas.harvard.edu/differential-privacy> accessed 14 December 2023

Dowe D and Graham O, ‘The Turing Test’ Winter 2021 The Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/archives/win2021/entries/turing-test/> accessed 1 September 2023

Escott J, ‘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> accessed 12 December 2023

‘Europe Privacy Policy’ (*OpenAI*, 22 June 2023) <https://openai.com/policies/eu-privacy-policy> accessed 26 November 2023

Europe Terms of Use’ (*OpenAI*, 14 November 2023) <https://openai.com/policies/eu-terms-of-use> accessed 24 November 2023

Facing Lawsuit and Investigation, Netflix Cancels Contest’ (*Lexology*, 7 April 2010) <https://www.lexology.com/library/detail.aspx?g=21ae0d4e-30d7-40f5-b753-efa7a9cb6d5b> accessed 14 December 2023

‘Factsheet 4: Types of Misinformation and Disinformation’ (*UNHCR*) <https://www.unhcr.org/innovation/wp-content/uploads/2022/02/Factsheet-4.pdf> accessed 1 December 2023

Gal U, 'ChatGPT Is A Data Privacy Nightmare. If You've Ever Posted Online, You Ought To Be Concerned' (*The University of Sydney*, 8 February 2023) <https://www.sydney.edu.au/news-opinion/news/2023/02/08/chatgpt-is-a-data-privacy-nightmare.html> accessed 3 December 2023

'Germany Launches Data Protection Inquiry over ChatGPT' (*The Local de*, 24 April 2023) <https://www.thelocal.de/20230425/germany-launches-data-protection-inquiry-over-chatgpt> accessed 5 December 2023

Greene T, 'ChatGPT Will Not Replace Google Search' (*The London School of Economics and Political Science*, 27 January 2023) <https://blogs.lse.ac.uk/impactofsocialsciences/2023/01/27/chatgpt-will-not-replace-google-search/> accessed 12 December 2023

Hao K, 'Training A Single AI Model Can Emit As Much Carbon As Five Cars in Their Life Time' (*MIT Technology Review*, 6 June 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4643518> accessed 13 December 2023

'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/> accessed 24 November 2023

Kern M.E, 'ChatGPT Generates More Than Data Outputs; Data Security and Privacy Concerns Grow as Artificial Intelligence Technology Rapidly Advances' (*Lexology*, 17 May 2023) <https://www.lexology.com/library/detail.aspx?g=1523494d-4f33-4602-b399-65be00d96d64> accessed 3 December 2023

Kovacs E, 'Simple Attack Allowed Extraction of ChatGPT Training Data' (*Security Week*, 1 December 2023) <https://www.securityweek.com/simple-attack-allowed-extraction-of-chatgpt-training-data/> accessed 3 December 2023

Krawczyk J, 'Bard's Latest Update: More Features, Languages and Countries' (*Google*, 13 July 2023) <https://blog.google/products/bard/google-bard-new-features-update-july-2023/> accessed 3 December 2023

Lomas N, 'Spanish Privacy Watchdog Says It's Probing ChatGPT Too' (*TechCrunch*, 13 April 2023) <https://techcrunch.com/2023/04/13/chatgpt-spain-gdpr/> accessed 5 December 2023

Lomas N, 'How to Ask OpenAI for Your Personal Data to Be Deleted or Not Used To Train Its AIs' (*TechCrunch*, 2 May 2023) <https://techcrunch.com/2023/05/02/chatgpt-delete-data/> accessed 11 December 2023

‘March 20 ChatGPT Outage: Here’s What Happened’ (*OpenAI*, 24 March 2023) <https://openai.com/blog/march-20-chatgpt-outage> accessed 7 December 2023
‘Misinformation vs Disinformation’ (*Taylor & Francis*) <https://insights.taylorandfrancis.com/social-justice/misinformation-vs-disinformation/> accessed 1 December 2023

Mustac T, ‘An Elephant Never Forgets and Neither Does ChatGPT’ (JD Supra, 13 April 2023) <https://www.jdsupra.com/legalnews/an-elephant-never-forgets-and-neither-7307484/> accessed 13 December 2023

Niles R, ‘GPT-3.5 Turbo Updates’ (*OpenAI*, November 2023) <https://help.openai.com/en/articles/8555514-gpt-3-5-turbo-updates> accessed 26 November 2023

‘OpenAI Personal Data Removal Request’ (*OpenAI*) <https://share.hsforms.com/1UPy6xqxZSEqTrGDh4ywo_g4sk30> accessed 11 December 2023

Pedregosa F and Triantafillou E, ‘Announcing the First Machine Unlearning Challenge’ (*Google Research*, 29 June 2023) <https://blog.research.google/2023/06/announcing-first-machine-unlearning.html> accessed 13 December 2023

Requests to Delist Content Under European Privacy Law’ (*Google*, 19 November 2023) <https://transparencyreport.google.com/eu-privacy/overview> accessed 19 November 2023

Schade M, ‘How ChatGPT and Our Language Models Are Developed (*OpenAI*, November 2023) <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed> accessed 5 December 2023

Schade M, ‘How Your Data Is Used to Improve Model Performance’ (*OpenAI*, November 2023) <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance> accessed 5 December 2023

York A, ’12 Best ChatGPT Alternatives & Competitors in 2023’ (*ClickUp*, 18 October 2023) <https://clickup.com/blog/chatgpt-alternatives/ - 57-12-claude> accessed 24 November 2023

NEWSPAPERS

‘Byron Kaye, ‘Australian Mayor Readies World’s First Defamation Lawsuit over ChatGPT Content’ *Reuters* (5 April 2023) <https://www.reuters.com/technology/australian-mayor-readies-worlds-first-defamation-lawsuit-over-chatgpt-content-2023-04-05/> accessed 7 December 2023

‘ChatGPT Can Now Browse the Internet for Updated Information’ *Aljazeera* (28 September 2023) <https://www.aljazeera.com/news/2023/9/28/chatgpt-can-now-browse-the-internet-for-updated-information> accessed 24 November 2023

James Clayton and Ben Derico, ‘Clearview AI Used Nearly 1m Times by US Police, It Tells the BBC’ *BBC* (San Francisco, 27 March 2023) <https://www.bbc.com/news/technology-65057011> accessed 25 November 2023

OpenAI May Leave the EU If Regulations Bite’ *Reuters* (25 May 2023) <https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/> accessed 13 November 2023

Zoe Kleinman and Antoinette Radford, ‘ChatGPT Can Now Access Up To Date Information’ *BBC* (27 September 2023) <https://www.bbc.com/news/technology-66940771> accessed 19 November 2023