Technical University of Munich

School of Management

# GDPR's Role in Regulating Large Language Models

## Analyzing noyb's Complaint Against OpenAI on Internal Data Handling

*Lionel Merz*

Bachelor's Thesis for the attainment of the degree
Bachelor of Science


| | |
|---|---|
| Examiner: | Prof. Dr. jur. Philipp Maume, S.J.D. (La Trobe), Associate Professorship Corporate Governance and Capital Markets Law |
| Supervisor: | Markus Haufellner |
| Study Program: | Bachelor in Management & Technology |
| Submitted by: | Lionel Merz |
| Address: | Freihamer Straße 2B, 82166 Gräfelfing |
| Matriculation Number: | 03734127 |
| Submission Date: | April 22nd, 2025 |

# Abstract

Within the last two years, the usage of Large Language Models (LLMs) has substantially expanded the scope of AI-based solutions, finding applications in many important contexts. As questions around the vast handling of citizens' personal data become more important than ever, European regulators have stepped up to debate the legal obligations for LLM controllers arising from the European Union's General Data Protection Regulation (GDPR). The subsequent Bachelor's Thesis investigates the interplay between LLM controllers, like OpenAI, and compliance with the rights of data subjects, focusing on GDPR's right of access under Article 15 GDPR, the "right to be forgotten" under Art. 17 as well as the obligation to rectify inaccurate personal data as stipulated under Art. 5(1)(d) GDPR.

Using the recent complaint filed by the NGO *noyb* against OpenAI as a pivotal case study, the research demonstrates how personal information embedded within a model's parameters poses a significant challenge for selective data deletion or correction. Making use of a theoretical framework, the legal case study, as well as interviews with leading legal and technical experts, the thesis analyzes OpenAI's argument of a "technical impossibility" as a basis for exempting LLM companies from their legal data obligations.

The following thesis postulates that this argument does not suffice to exempt LLM companies from their GDPR obligations. Rather the opposite, it identifies a clear need for enhanced privacy-by-design measures and regulatory guidance to ensure meaningful protection of data subject rights within the context of the European Union.

# Contents

**Abstract** i

**List of Figures** iv

**Acronyms and Abbreviations** v

**1 Introduction** 1
- 1.1 Background Information . . . 2
- 1.2 Purpose of the Thesis . . . 3
- 1.3 Methodology and Structure of the Thesis . . . 4

**2 Theoretical Framework** 6
- 2.1 Large Language Model (LLM) . . . 6
- 2.2 LLM Training Process and Functionality . . . 7
- 2.3 GDPR Overview . . . 9
  - 2.3.1 Scope of the General Data Protection Regulation (GDPR) . . . 10
  - 2.3.2 Personal Data . . . 10
  - 2.3.3 Rights of Data Subjects . . . 11
- 2.4 Applicability of GDPR to AI/LLMs . . . 12

**3 Case Study: The *noyb* Complaint against OpenAI** 14
- 3.1 Descriptive Overview of *noyb*'s Complaint . . . 14
  - 3.1.1 Structure and Representation . . . 15
  - 3.1.2 Factual Background and Operational Practices . . . 16
- 3.2 Legal Violations and Compliance Issues . . . 17
  - 3.2.1 Principal Complaint Reason . . . 17
  - 3.2.2 Competitive Authority and Jurisdictional Considerations . . . 17
  - 3.2.3 Remedies and Requests . . . 18

**4 Methodology** 19
- 4.1 Case Analysis Method . . . 19
- 4.2 Method and Procedure for Conducting Interviews . . . 19
  - 4.2.1 Interview Objective . . . 19
  - 4.2.2 Interview Design . . . 20
  - 4.2.3 Data Collection Method . . . 21
  - 4.2.4 Data Analysis and Limitations . . . 21

**5 Analysis and Interview Results** 23
- 5.1 Case Analysis Procedure . . . 23
- 5.2 Expert Interviews: Comparative Analysis of Excerpts . . . 25

**6 Discussion** **28**
6.1 Contextual Discussion of Case Findings . . . . . 28
6.2 Legal Implications of GDPR in the Case of LLMs . . . . . 32

**7 Conclusion and Key Findings** **34**

**A Appendix** **36**
A.1 Interview Transcripts . . . . . 36
A.1.1 Interview with Boris Paal . . . . . 36
A.1.2 Interview with Markus Hupfauer . . . . . 41

**Bibliography** **50**

# List of Figures

2.1 Classification of chatbots . . . . . 7
2.2 Structure of an AI chatbot, highlighting its underlying components . . . . . 8
2.3 Preserving user privacy within an LLM . . . . . 13

# Acronyms and Abbreviations

**AI** Artificial Intelligence.

**Art.** Article.

**EU** European Union.

**GDPR** General Data Protection Regulation.

**LLMs** Large Language Models.

**ML** Machine Learning.

**NGO** Non-Governmental Organization.

**NLG** Natural Language Generation.

**NLP** Natural Language Processing.

**NLU** Natural Language Understanding.

***noyb*** NGO *noyb – European Center for Digital Rights*.

**RLHF** Reinforcement Learning with Human Feedback.

# 1 Introduction

> "Large Language Models represent a significant advancement in artificial intelligence, finding applications across various domains. However, their reliance on massive internet-sourced datasets for training brings notable privacy issues[...]."[1]

Within the last years, Large Language Models (LLMs) such as OpenAI's ChatGPT have garnered significant attention due to their transformative impact across diverse sectors including academia, industry, healthcare, law, and education. These models function based on the principle of predicting and generating sequences of text that are coherent and contextually relevant by processing vast quantities of textual data from, among other sources, the internet.[2] Consequently, their performance and functionality are largely reliant on the size and diversity of their training datasets. This dependency, however, introduces significant privacy and compliance challenges, particularly when the processed datasets contain personal or sensitive information.[3]

A particularly noteworthy recent case is the complaint filed with the Austrian Data Protection Authority (DPA) in April 2024 by the NGO *noyb – European Center for Digital Rights* (*noyb*) against the US company OpenAI. The use of LLMs seems to be increasingly becoming the focus of commercial and public applications. This complaint criticizes the fact that ChatGPT repeatedly provides incorrect personal data and that there are no adequate mechanisms for correcting or deleting personal data.[4] The European Union's General Data Protection Regulation (GDPR) requires all data processing stakeholders to process personal data correctly, transparently and in a timely manner. These are all requirements that often seem to reach their limits with the current technology. OpenAI's argument that "technical impossibilities" would prevent selective corrections or deletions seems particularly critical.[5] This position raises not only legal, but also a number of ethical questions, as the rights of the data subjects must still be safeguarded under current European law regardless.

[1] Miranda et al. 2025, p. 1.
[2] Yan et al. 2024, p. 2.
[3] Singh and Namin 2025, p. 3.
[4] *noyb* 2024, p. 2.
[5] Mittal et al. 2024, p. 738.

## 1.1 Background Information

In parallel with technological advancements, legislative measures to protect personal data and ensure privacy compliance have been significantly strengthened, particularly in the European Union with the General Data Protection Regulation (GDPR)'s entry into force in May 2018. The GDPR outlines explicit requirements regarding the handling of personal data, emphasizing principles such as data minimization, transparency, accountability, and accuracy.[6] The regulation also endows data subjects with enhanced rights, such as the right to access their data (Article (Art.) 15 GDPR), rectify incorrect information (Art. 16 GDPR), and demand the erasure of their personal data under specific conditions (Art. 17 GDPR).[7] However, these stringent GDPR requirements clash with the inherent operational mechanisms of LLMs, posing substantial challenges in ensuring compliance. Specifically, the opaque nature of data processing within LLM architectures makes fulfilling these rights challenging.[8]

A notable complexity arises from the fact that the data utilized in training LLMs is extensively sourced from publicly accessible online materials, encompassing an assortment of information that may not have been intended for such usage. Consequently, there is a risk that personal and potentially sensitive data of individuals, not intended for a global audience, might inadvertently be incorporated into the model parameters.[9] Once such data has been embedded in the high-dimensional matrices of LLM weights, its extraction or modification becomes technically challenging without performing retraining or severely impacting model performance. This dilemma illustrates a fundamental tension between the performance optimization of LLMs and adherence to GDPR principles, notably the data minimization and the right to erasure.[10] Recent investigative studies by data protection authorities ("ChatGPT Taskforce") and research groups indicate that popular models like ChatGPT have repeatedly demonstrated significant inaccuracies when processing personal information, thereby breaching GDPR obligations related to data accuracy and transparency.[11] This issue underscores the immediate need for advanced research on both technical solutions (e.g., differential privacy, machine unlearning) and legislative frameworks to harmonize technological capabilities with legal and ethical standards for processing of personal data.[12] Despite these challenges, the application of LLMs continues to expand across various domains due to their unparalleled capabilities in areas such as text generation, data analytics, automation of routine processes, and sophisticated communication tasks like customer

[6] European Data Protection Board 2024, p. 4.
[7] European Data Protection Board 2024, p. 5.
[8] Feretzakis and Verykios 2024, p. 12.
[9] Yan et al. 2024, p. 2.
[10] Feretzakis et al. 2025, p. 4.
[11] European Data Protection Board 2024, p. 6.
[12] Mittal et al. 2024, p. 2.

support. The rapid evolution and broad deployment of these Machine Learning (ML) models have catalyzed critical discourse on their responsible governance and compliance with regulatory standards.[13] As companies and public bodies increasingly leverage LLM-based applications, the urgency of resolving existing tensions between AI functionality and GDPR compliance has become particularly prominent, motivating interdisciplinary research and policy-making initiatives at international levels.[14]

Moving through the current complex dynamics between LLM technological capabilities and regulatory frameworks, notably GDPR, represents a vital challenge for stakeholders, including AI developers, regulatory bodies, data protection authorities, and civil society organizations. The ongoing debates and legal actions—such as the notable complaint filed by the Non-Governmental Organization (NGO) *noyb* against OpenAI for GDPR infringements due to inaccuracies in data processing—highlight the immediate practical and theoretical significance of this intersection between law, technology, and ethics.[15]

## 1.2 Purpose of the Thesis

Considering the increasing prevalence of LLMs and their broad integration into everyday applications, the central purpose of this thesis is to systematically analyze and evaluate the complex interplay between LLMs' technological capabilities and the regulatory demands set forth by the General Data Protection Regulation (GDPR). The recent rise in popularity of LLM services has resulted in many offerings available in global consumer and commercial markets, including within the European Union (EU), even before specific regulations addressing the technology like the EU's AI Act had been established. As such, complaints have been brought forward to various European Data Protection Authorities (DPA) against aspects in the usage and commercial offering of this technology.

This thesis will cover one such complaint, filed by NGO *noyb* in April of 2024 before the Austrian DPA. It has thereby highlighted critical legal and ethical issues surrounding the use of personal data in artificial intelligence systems.[16] The subsequent legal analysis thus focuses on identifying and critically discussing how LLMs conflict with or potentially violate Art. 5(1)(d) and Art. 15 GDPR principles, including transparency, accuracy, and data subjects' rights to rectification and erasure.[17] The aim is to explore the technical arguments and justifications provided by OpenAI, especially concerning its claims of "technical impossibility" related to selective deletion or correction of personal data

[13] Rodriguez et al. 2024, p. 5.
[14] Miranda et al. 2025, p. 3.
[15] *noyb* 2024, p. 4.
[16] *noyb* 2024, p. 2.
[17] Feretzakis and Verykios 2024, p. 12.

embedded within trained model parameters.[18] In this effort, the objective is to explore whether GDPR shall be applicable to the LLM technology, and, if that is the case, what consequences this entails with regard to potential violations of the further EU regulation. By dissecting its arguments through a multidisciplinary lens, drawing insights from technology, law, and its surrounding ethics, this subsequent legal research intends to analyze the validity of these claims, determine their alignment with current legal standards, and outline potential pathways toward a greater regulatory compliance.

## 1.3 Methodology and Structure of the Thesis

The thesis is set to be structured into two key parts to ensure a systematic research approach. This structure is designed to address the central research question of this analysis:

> Analyzing the case C-078 *noyb* against Controller OpenAI, to what extent do LLMs comply with GDPR requirements regarding personal data, and what are the technical and regulatory implications arising from their current limitations?

To answer this, the study employs a systematic theory review, a case analysis, and incorporates qualitative expert interviews as research methodology. The thesis is divided into two main sections, with the aim to reflect both on the case analysis as well as its broader interpretation within the European data economy.

**Part 1: Theoretical Framework and Case Study**

> How do Large Language Models LLMs operate in relation to GDPR principles, particularly concerning data accuracy, transparency, and rights of data subjects?

The first part of the thesis will provide a theoretical exploration of Large Language Models, including their functionality, data processing practices, and specific challenges within the context of GDPR compliance. Through the subsequent review of the case study involving the complaint lodged by *noyb* against OpenAI, this section sets to examine the legal claims related to breaches of GDPR Art. 5(1)(d) and Art. 15. It thereby aims to identify the fundamental technical processes and limitations of LLMs that conflict with European data protection standards. This section sets the foundational knowledge, distinguishing itself clearly from the empirical approach in the second part by remaining predominantly theoretical, analytical, and based on existing literature, legal texts, and official documentation.

[18] Mittal et al. 2024, p. 22.

**Part 2: Case Analysis and Empirical Interpretation**

> How do legal and technical experts interpret the implications of GDPR compliance and the feasibility of enforcing data subject rights in LLM-based systems?

The second part of the thesis shifts to an empirical approach, employing standardized expert interviews which are set to integrate a nuanced insight into the practical implications and potential resolutions of GDPR-related challenges currently posed by LLMs. This analysis involves semi-structured qualitative interviews with legal scholars and data privacy professionals in order to investigate their perspectives on current compliance gaps, feasible solutions, and the broader regulatory impact. By systematically analyzing expert perspectives, this part is aimed to provide a concrete insight into the practical enforceability of GDPR principles and identify emerging regulatory needs and technological innovations required for future compliance.

# 2 Theoretical Framework

The following section provides an overview of the theoretical foundations of modern Large Language Models (LLMs), along with their data protection implications within the context of the GDPR. Understanding these concepts is crucial for the subsequent analysis presented in this thesis.

## 2.1 Large Language Model (LLM)

LLMs are generative Artificial Intelligence (AI) systems which are developed through training on massive, unstructured text datasets. This makes them capable of understanding, processing, and generating language.[1] One very frequent application of LLMs is in chatbots. These have evolved significantly over the past years, enabling powerful capabilities in areas such as marketing, customer retention, and sales development.[2] By utilizing attention mechanisms, LLMs can quickly identify relevant contextual information and incorporate it into their text predictions.[3] They have emerged as a groundbreaking innovation within Natural Language Processing (NLP), significantly impacting various fields such as translation, text summarization, sentiment analysis, content generation, and automated customer support.[4]

Historically, the concept of language modeling has its roots in statistical methods dating back to the mid-twentieth century. However, it was the advent of deep neural networks and specifically transformer-based architectures in 2017, with models such as the seminal Transformer by Vaswani et al. 2023, that revolutionized the capability of these models. Transformers introduced mechanisms such as self-attention, allowing models to efficiently capture contextual dependencies within text, paving the way for increasingly sophisticated and scalable language models.[5] In 2018, the introduction of OpenAI's GPT-2 further highlighted the potential of LLMs, with its unprecedented ability to produce coherent and contextually relevant text across diverse topics. Subsequent iterations, notably

[1] Yan et al. 2024, p. 4.
[2] Singh and Namin 2025, p. 22.
[3] Vaswani et al. 2023, p. 3.
[4] Rodriguez et al. 2024, p. 4.
[5] Yan et al. 2024, p. 3.

GPT-3 released in 2020, demonstrated even greater performance and scalability, drawing considerable attention from both the research community and industry stakeholders. This period marked a shift toward models with billions of parameters trained on vast internet-sourced datasets, raising important questions concerning privacy, ethics, and data governance.[6]

## 2.2 LLM Training Process and Functionality

LLMs function fundamentally on the principle of predicting the next word or sequence of words based on previously provided text inputs, utilizing complex probabilistic models trained on extensive textual corpora. This training involves massive datasets derived predominantly from publicly available internet texts, books, articles, and other textual sources, enabling the models to internalize linguistic structures, contexts, and its semantics.[7]

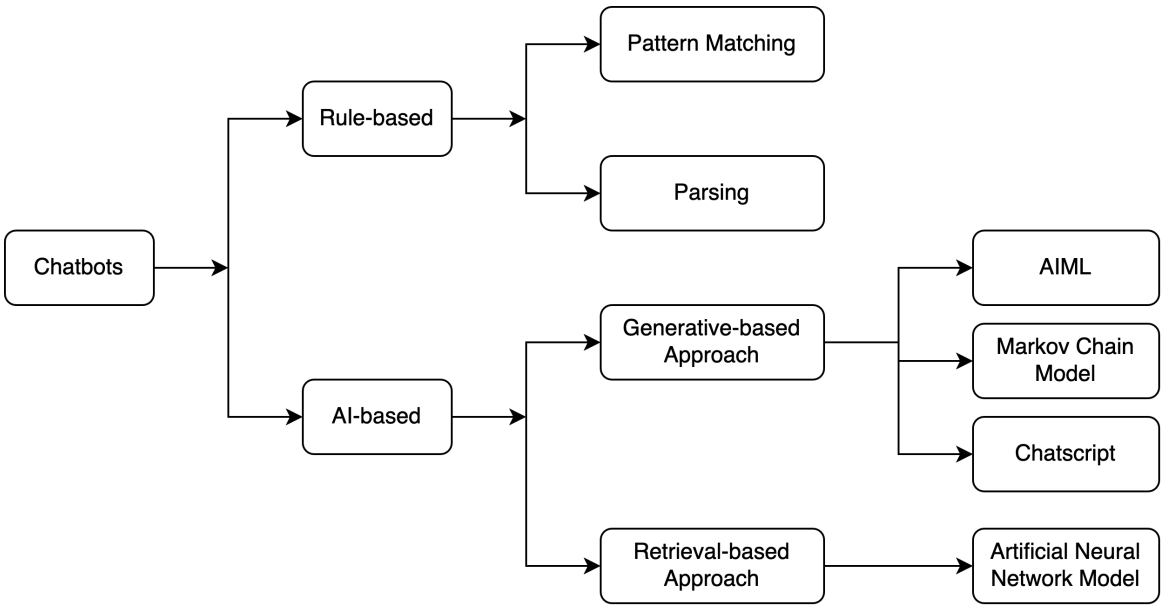


Figure 2.1: Classification of chatbots

In general, the implementation of chatbots can be divided into two main approaches, as shown in Figure 2.1.[8] On one side, there are rule-based systems that depend on well-defined rules and parsing strategies, following a clear pattern-matching logic.[9] On the other side are AI-based chatbots, either following a generative approach, meaning the autonomous construction of sentences, or a retrieval-based approach, i.e. fetching appropriate responses from a knowledge base.[10] Generative approaches

[6] Singh and Namin 2025, p. 5.
[7] Muhammad et al. 2024, p. 144.
[8] Singh and Namin 2025, p. 3.
[9] Muhammad et al. 2024, p. 141.
[10] Singh and Namin 2025, p. 2.

include methods such as AIML utilizing a text-based format to define sets of rules[11], Markov chain models, which determine the probability of the next words, and modern neural networks like the previously mentioned transformer models, capable of generating context-aware and "natural-sounding" responses from extensive training data.[12] In the retrieval-based approach, the chatbot searches for pre-existing responses in a database, using semantic similarity and vector space search methods, with probabilities calculated through neural networks.[13]

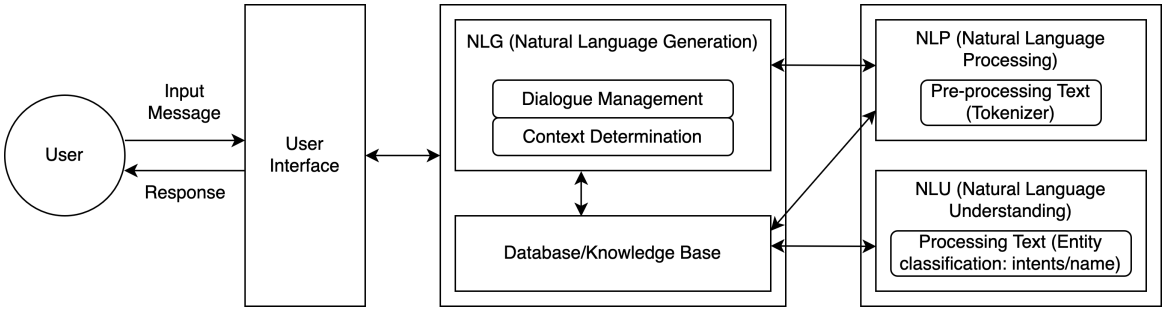


Figure 2.2: Structure of an AI chatbot, highlighting its underlying components

The basic architecture and functioning of an AI chatbot is often the same, as illustrated in Figure 2.2.[14] First, the user enters their input via the user interface, which is then preprocessed using Natural Language Processing (NLP). Next, the system uses Natural Language Understanding (NLU) to identify the user's intentions and relevant entities. Based on this, the dialogue management system controls the flow of the conversation, taking the context into account and, if necessary, retrieving information from a database or other knowledge base. Finally, the Natural Language Generation (NLG) module produces a fluently-worded response, which is then returned to the user.

However, sometimes the input, generated based on statistical probability calculations seems to be distorted due to its incomplete or inaccurate data. A characteristic phenomenon of modern LLMs in this context is what is known as "hallucinations". In such cases, incorrect but still plausible-sounding answers are generated.[15] During training, the model learns statistical relationships between words and contexts. However, if certain information is missing or if the data is contradictory, the system extrapolates likely individual words or entire sentences that do not always reflect reality or factual knowledge. In order to mitigate hallucinations and improve their interpretability and accuracy, recent developments in LLMs have increasingly focused on techniques such as Reinforcement Learning with Human Feedback (RLHF), differential privacy, and federated learning, which are being explored

[11] Cory et al. 2025, p. 82.
[12] Liu et al. 2025, p. 22.
[13] Rodriguez et al. 2024, p. 3.
[14] Singh and Namin 2025, p. 6.
[15] El Naqa and Murphy 2015, p. 5.

to improve model robustness and reduce privacy concerns.[16] That said, especially when processing sensitive data, misunderstandings can still easily arise if incorrect information is to be presented in a way that seems plausible for the LLM. As such, compliance with privacy regulations such as the European Union's General Data Protection Regulation (GDPR) appears to remain a highly contested issue. As the capabilities and deployments of LLMs further expand, resolving the inherent tensions between technological functionality, data privacy, and regulatory frameworks becomes increasingly critical for sustainable and ethical utilization of these powerful AI systems.[17]

## 2.3 GDPR Overview

The European GDPR, known officially as Regulation (EU) 2016/679, entered into force on May 24, 2016 and applies to all member states since May 25, 2018. Developed by the European Union, it is designed to unify and fortify data protection for everyone within the EU and the European Economic Area (EEA).[18] In addition to safeguarding data within its borders, the GDPR regulates the transfer of personal information to regions outside the EU and EEA, giving it global influence due to its extraterritorial provisions.[19]

Art. 3 of the regulation specifically delineates its territorial scope, thereby extending EU data protection obligations to international activities involving personal data. This global reach carries significant implications, especially for technological developments such as LLMs, which routinely process massive volumes of personal information across different jurisdictions. Under the GDPR, data protection obligations are triggered when a data controller or processor maintains an operational presence within the EU, known as the "establishment criterion," or when entities outside the EU target EU residents through the provision of goods, services, or monitoring of behavior.[20] Importantly, even organizations without a physical presence in the EU can fall within the GDPR's jurisdiction if their activities involve processing data related to individuals in the EU, significantly broadening its reach under international law. This extensive regulatory framework positions the GDPR as an important cornerstone in shaping global governance frameworks for artificial intelligence and data-centric technologies.[21]

Given the rapid rise and widespread adoption of technologies such as LLMs, compliance with GDPR presents unique challenges. The regulation explicitly mandates that entities processing per-

[16] Miranda et al. 2025, p. 9.
[17] Feretzakis and Verykios 2024, p. 14.
[18] Voigt and von dem Bussche 2024, p. 33.
[19] Kamarinou et al. 2016, p. 19.
[20] European Data Protection Board 2024, p. 4.
[21] European Data Protection Board 2024, p. 4.

sonal data undertake stringent measures to uphold data privacy, thereby clearly defining roles and responsibilities for both data controllers and processors.[22]

### 2.3.1 Scope of the GDPR

The fundamental goal of the GDPR is to provide individuals greater control over their personal data and to establish a harmonized legal framework across the European Union, thus facilitating cross-border business activities and ensuring robust protection of fundamental rights. The GDPR requires personal data to be processed lawfully, transparently, and fairly. Moreover, personal data must only be collected for explicit, specific, and legitimate purposes; must adhere strictly to data minimization principles; and must remain accurate and up-to-date, with inaccuracies promptly rectified or deleted.[23]

From a temporal perspective, GDPR is applicable to all data processing activities that commenced after the regulation's implementation on May 25, 2018. Additionally, ongoing processing activities that started prior to this date but continued afterward are also subject to the GDPR's provisions, requiring the involved entities to ensure compliance retrospectively and continuously.[24] In terms of material scope, GDPR applies broadly to any activity involving the processing of personal data, whether fully or partially automated, as well as data that form part of structured filing systems.[25] The term "processing" under GDPR covers a wide array of operations including but not limited to collecting, recording, organizing, structuring, storing, adapting, retrieving, disseminating, disclosing, and erasing personal data.

### 2.3.2 Personal Data

Under the GDPR, "personal data" refers explicitly to "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly" (Art. 4 GDPR). Examples of personal data under the GDPR are thereby extensive, ranging from traditional identifiers such as names, identification numbers, and location data to modern identifiers like IP addresses, cookies, and digital footprints. Notably, GDPR recognizes a specific category of sensitive personal data, which includes genetic, biometric, health, racial, ethnic, political, sexual orientation, and religious data, among others. These data types demand higher protection standards due to their potential for significant privacy implications and risks of harm in case of misuse or unauthorized disclosure.[26] Processing such special categories of data is generally prohib-

[22] Feretzakis and Verykios 2024, p. 12.
[23] Albrecht 2016, p. 287.
[24] Mittal et al. 2024, p. 737.
[25] Feretzakis et al. 2025, p. 4.
[26] Albrecht 2016, p. 288.

ited unless explicit consent is given by the data subject or under certain limited exceptions specified explicitly by GDPR, such as substantial public interest or for preventive medicine purposes.[27] This aspect is particularly relevant to the LLM industry, where inadvertent inclusion of sensitive data into training sets poses serious regulatory and ethical challenges.

### 2.3.3 Rights of Data Subjects

The GDPR grants individuals, as data subjects, a set of rights. These include the right of access (Art. 15), which enables individuals to confirm whether their data is being processed and to request access along with information regarding such processing; the right to rectification (Art. 16) to amend any incorrect personal details; and the right to erasure, commonly known as the "right to be forgotten" (Art. 17), which permits data deletion under certain conditions. Other rights include the right to restrict processing (Art. 18), allowing individuals to impose limits on the processing of their data; the right to data portability (Art. 20), which makes it possible for individuals to retrieve and transfer their data in a structured, commonly used, and machine-readable format; the right to object (Art. 21), which permits opposition to data processing in certain contexts such as direct marketing; and protections under the rights related to automated decision-making (Art. 22), which shield individuals from decisions based solely on automated processing, including profiling, that significantly affect their rights or interests.

Organizations that act as either data controller or processor must adopt suitable technical and organizational measures to adhere to the GDPR guidelines. This compliance includes, but is not limited to, performing Data Protection Impact Assessments (DPIAs) for high-risk processing (Art. 35), designating Data Protection Officers (DPOs) in specified circumstances (Art. 37–39), and incorporating data protection principles into the design and default settings of their systems (Art. 25). Failure to comply with the GDPR can lead to substantial fines, potentially reaching up to 4 per cent of a company's annual global turnover or €20 million. Since its implementation, the regulation's robust enforcement has been highlighted by several notable cases[28] Its influence extends globally, serving as a model for privacy legislations such as the California Consumer Privacy Act (CCPA) in the United States, Brazil's General Data Protection Law (LGPD), and Japan's Act on the Protection of Personal Information (APPI).

[27] Voigt and von dem Bussche 2024, p. 34.
[28] For example, against Google. (European Data Protection Board 2019)

## 2.4 Applicability of GDPR to AI/LLMs

Applying the GDPR to LLMs, one of the most commonly invoked legal bases is Art. 6(1)(f) GDPR. This article permits the lawful processing of personal data where it is necessary to pursue the legitimate interests of the controller or a third party, provided these interests are not overridden by the fundamental rights and freedoms of data subjects.[29] This provision illustrates a core tension in LLMs: balancing business or research interests against the often complex and evolving framework of data protection and privacy standards.

Beyond this, it is important to note that through Art. 25, the GDPR further obligates controllers to integrate data protection measures already during the conception and design stages of data processing systems, including AI language models. Consequently, it must be ensured that, by default, only those personal data necessary for the specific purpose of processing are processed.[30] This principle requires the proactive embedding of data protection principles throughout the entire development lifecycle. However, this requirement presents a significant challenge, especially in the context of developing LLMs.[31] The complex and highly nested architectures of these models considerably complicate the seamless integration of privacy-enhancing measures. Furthermore, many LLMs necessitate continuous updates with new data, which further complicates the sustained adherence to data protection standards. An additional tension arises between data protection and model performance, as the more robust the implemented data protection mechanisms is, the higher the potential impact on system functionality gets.[32]

To address these challenges, developers increasingly rely on technical measures such as differential privacy, federated learning, or homomorphic encryption. These methods enable the protection of personal data both during the training and inference phases. Other approaches include modular model architectures, wherein individual components can be specifically equipped with data protection mechanisms. Additionally, integrating automated compliance tools into the development process promises effective implementation of "Privacy by Design".[33]

Potential solutions for implementing data protection through technical design in LLMs include the aforementioned privacy-preserving techniques to safeguard personal data across all phases of model usage.[34] Moreover, modular architectures can be developed to facilitate the targeted integration

[29] Feretzakis and Verykios 2024, p. 8.
[30] Feretzakis et al. 2025, p. 9.
[31] Laurelli 2024, p. 15.
[32] Feretzakis and Verykios 2024, p. 27.
[33] Mittal et al. 2024, p. 737.
[34] Feretzakis et al. 2025, p. 12.

of privacy-relevant components.[35] Through the adoption of such measures, data protection can be structurally embedded from the outset rather than ensured merely as an afterthought, thus consistently aligning with the concept of "Privacy by Design" as stipulated by GDPR.

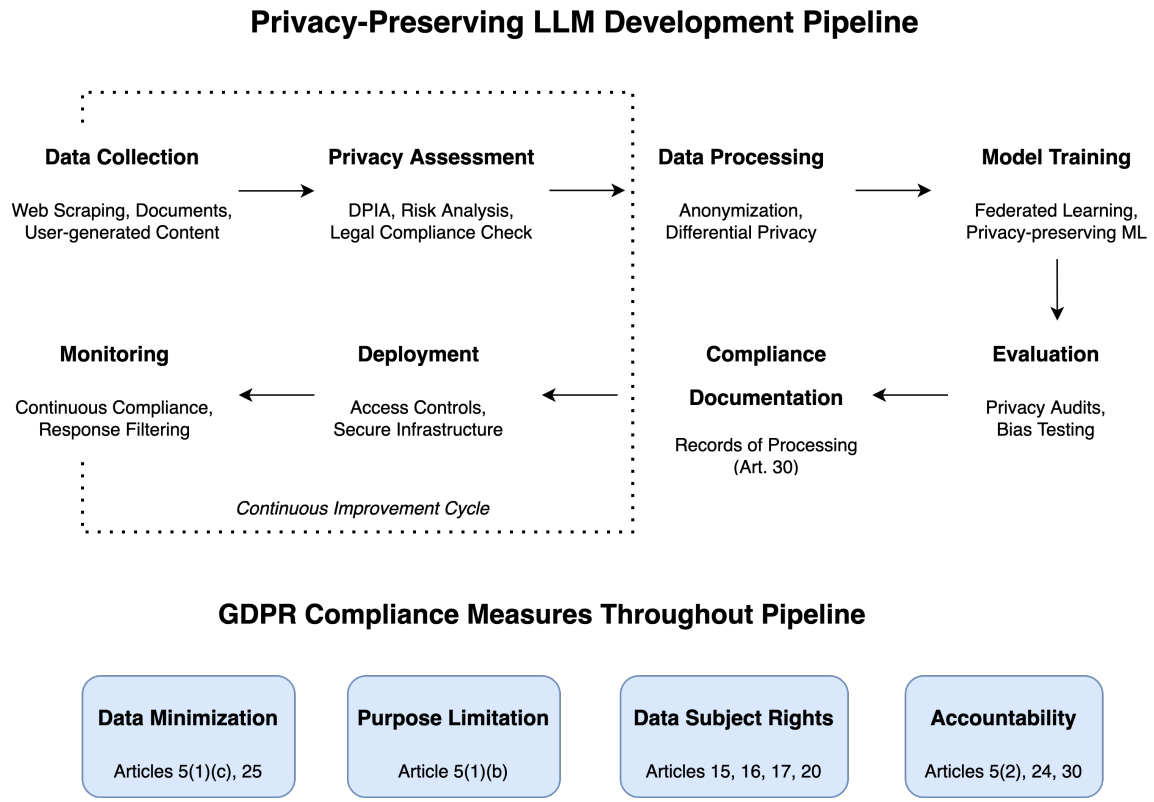


Figure 2.3: Preserving user privacy within an LLM

The model of privacy-compliant processing depicted in Figure 2.3[36] illustrates this integrative approach. From data collection, risk assessment, and data processing to model training, evaluation, documentation, and continuous monitoring, technical and organizational data protection measures are considered at each phase. At least theoretically, this aims for a continuous improvement process ensuring compliance with GDPR—particularly concerning data minimization (Art. 5(1)(c)), purpose limitation (Art. 5(1)(b)), rights of data subjects (Art. 15–17, 20), and accountability (Art. 5(2), 24, 30).

[35] Cory et al. 2025, p. 5.
[36] Feretzakis et al. 2025, p. 9.

# 3 Case Study: The *noyb* Complaint against OpenAI

This chapter examines the recent complaint filed by the data protection NGO *noyb – European Center for Digital Rights* against OpenAI.[1] The complaint thereby centers on allegations that OpenAI's renowned LLM "ChatGPT" has failed to adhere to critical GDPR obligations, including the principles of data accuracy and the right of access. Challenging the internal data processing practices within LLMs, the case provides an illustrative example of the ongoing tensions between advanced AI technologies and its stringent data protection standards, safeguarding users privacy.

## 3.1 Descriptive Overview of *noyb*'s Complaint

The data protection NGO *noyb* filed a complaint on April 29, 2024, with the *Österreichische Datenschutzbehörde (DSB)*, the federal Data Protection Authority (DPA) for Austria, against the US company OpenAI OpCo, LLC based in California, US. The complaint primarily concerned the principle of accuracy under Art. 5(1)(d) GDPR, as well as an access request under Art. 15 GDPR directed at OpenAI. The *noyb* complaint, distinctively designated internally as "Case C-078," represents a significant challenge regarding the applicability of the GDPR to the internal data handling of AI systems such as ChatGPT. The complaint was submitted on behalf of a data subject, who is represented by *noyb* in accordance with Art. 80 GDPR.[2] At its core, the complaint alleges that OpenAI's AI-driven language model ChatGPT has generated incorrect personal data about the data subject. Specifically, when queried, ChatGPT delivered an inaccurate date of birth for the individual. This example underscores a fundamental issue: the inability of systems like ChatGPT to provide accurate and reliable personal data verifiably. The data subject had already filed an access and erasure request on December 4, 2023. However, in its response dated February 7, 2024, OpenAI acknowledged the existence of user account data, but stated that it is technically infeasible to correct or delete specific personal data

[1] *noyb* 2024, p. 1.
[2] *noyb* 2024, p. 1.

within the model. Instead, OpenAI offered to block the display of these data according to its filtering guidelines, while leaving the underlying inaccuracy intact. *noyb* states this approach to be in conflict with the GDPR, as it mandates the prompt rectification or erasure of incorrect personal data.

### 3.1.1 Structure and Representation

The general complaint starts by outlining the representation. *noyb* acts under Art. 80(1) GDPR to defend the rights of the complainant. It identifies the respondent as OpenAI—the legal entity responsible for operating ChatGPT—and provides detailed contact information, including both its European establishment and its US headquarters. This introductory section sets the procedural framework and reaffirms that complaints may only be directed to one joint controller, even if operational and decision-making functions are distributed internationally.[3]

#### ***noyb*'s advanced allegations**

The allegations by *noyb* can be broadly divided into two main categories. First, there is the violation of the principle of accuracy as stipulated in Art. 5(1)(d) GDPR. This principle obliges data controllers to ensure that personal data is factually correct and kept up-to-date; in the case of ChatGPT, it means that the generated information must be correct. The generation of incorrect personal data clearly constitutes a breach of this obligation. Secondly, the complaint highlights a violation of the right of access under Art. 15 GDPR, which guarantees that data subjects have the right to receive confirmation on whether their personal data are being processed and, if so, to obtain access to such data along with specific details regarding its processing. According to the complaint, OpenAI has failed to provide detailed information about the processing of personal data within the algorithms of ChatGPT, thereby infringing on this right.[4]

#### **OpenAI's preliminary response**

As a response, OpenAI has presented several counter-arguments intended to refute these allegations. The company contends that the complex architecture of ChatGPT makes it technically unfeasible, at present, to selectively correct or delete specific personal data, even with the best of intentions. However, the argument of technical complexity does not exempt a controller from its legal obligations under the GDPR. Data controllers are expected to implement "Privacy by Design" from the early stages of system development, as required by Art. 25(1) GDPR. OpenAI's proposal to block the

[3] *noyb* 2024, p. 4.
[4] *noyb* 2024, p. 2.

display of incorrect information through an existing filtering portal, rather than to delete or rectify the inaccurate data, is criticized in the complaint as merely a superficial attempt to comply with the GDPR's requirements. Moreover, OpenAI argued that ChatGPT is a research project and should be held to different standards. Nonetheless, under current legal interpretations, the processing of personal data, even in a research context, remains bound by the provisions of the GDPR, as emphasized by Art. 89 GDPR. Thus this line of reasoning does not hold regardless of whether ChatGPT, which is provided as a commercial offering, even constitutes a research project.

### 3.1.2 Factual Background and Operational Practices

In its exposition of the case, the complaint describes ChatGPT as an artificial intelligence application built upon LLMs. The complaint delineates how these models function by statistically predicting word sequences based on vast training datasets that include personal data. Next, it highlights a critical failure: when the concerned data subject's date of birth is requested, ChatGPT returns multiple inaccuracies. This erroneous output, verifiable by any user of the ChatGPT system, illustrates a broader systemic issue, namely the model's inability to accurately process or correct personal data despite being trained on comprehensive datasets.

The factual narrative is further being reinforced by a detailed timeline. The data subject first submitted an access and erasure request on December 4, 2023, seeking not only to verify the personal data processed by the system, but also to have any inaccuracies rectified or removed. OpenAI's response, issued on February 7, 2024, acknowledged the existence of user account data but failed to sufficiently address the processing of personal data embedded within the language model itself.
Instead, OpenAI explained that its only available option is to deploy an all-encompassing blocking function. This means that although the erroneous data may be hidden from public view, the underlying inaccuracies are not corrected or deleted, and they persist in the system's internal processing. Further adding to the concerns, the complaint points out that this approach has significant implications for data accuracy as mandated by the GDPR. The persistent presence of incorrect personal data not only undermines the rights of the data subject to accurate information under Art. 16 GDPR but also compromises the overall reliability of the system, particularly if the information is used for decision-making or further data processing. By failing to provide the required level of transparency about the processing mechanisms and by not implementing measures to rectify such inaccuracies, OpenAI's practices appear to circumvent the obligation under Art. 5(1)(d) GDPR to ensure that personal data is accurate and up-to-date.[5] The complaint thereby raises broader questions about the internal opera-

[5] *noyb* 2024, p. 5.

tional practices of AI systems like ChatGPT. It suggests that the integration of extensive personal data into the training process, combined with the technical limitations in correcting such data post-training, creates a situation in which systemic inaccuracies may become inherently entrenched. This scenario not only has immediate consequences for the data subject affected in this particular case but may also signal a more widespread issue across similar AI systems, which could impact the fundamental rights of countless data subjects whose information has been inadvertently or inaccurately processed.

The factual background and operational practices outlined in the complaint seem to underscore the current challenge that arises when LLMs process personal data. The evidence presented indicates that despite the technological advancements of these systems, there remains a significant gap between the intended regulatory requirements, specifically the need for accuracy and transparency under the GDPR, and the current technical and operational capabilities of AI models like ChatGPT. This gap raises critical questions about how controllers can effectively reconcile advanced AI functionalities with the stringent demands of data protection law.

## 3.2 Legal Violations and Compliance Issues

### 3.2.1 Principal Complaint Reason

As mentioned, the complaint has two principal grounds for its challenge. First, there is a breach of the principle of accuracy, as specified under Art. 5(1)(d) GDPR. Under this provision, the controller must promptly rectify or erase inaccurate personal data. However, the respondent claims that the system's technical limitations preclude any selective correction of the complainant's date of birth without affecting other stored information. Further, the violation of the right of access: the respondent's failure to provide comprehensive information regarding the processing of personal data within the internal ChatGPT system contravenes Art. 12(3) and 15 of the GDPR.[6] Although OpenAI did disclose certain details about user account data, it did not explain how personal data is utilized in the underlying language model. The complaint argues that this rationale does not constitute a legally acceptable exemption from the accuracy obligation, especially given that the inaccurate data does not contribute to any public interest debate and violates the data subject's right to privacy.

### 3.2.2 Competitive Authority and Jurisdictional Considerations

A further aspect of the complaint focuses on the issue of jurisdiction and the delineation of roles among joint controllers. While OpenAI maintains a presence in Ireland through OpenAI Ireland

[6] *noyb* 2024, p. 4.

Limited, the substantive decisions including those involving data processing for the design and implementation of their LLMs remain centralized in the US-based OpenAI Limited Liability Company. This distribution of control is critical because the data subject is resident in Austria, and the Austrian *Datenschutzbehörde* (Data Protection Authority) is thus accorded competence over the complaint under Art. 55 and 77 GDPR. By filing the complaint against OpenAI OpCo, LLC, the complainant not only accentuates the legal responsibilities of the primary decision-making entity but also preserves the right to pursue additional measures against any other joint controller if necessary[7][3,4]noybComplaintPDF

### 3.2.3 Remedies and Requests

The complaint concludes with a series of concrete requests designed to enforce compliance and ensure that the rights of the data subject are protected. In essence, the complainant here asks that the competent supervisory authority undertake a comprehensive investigation into OpenAI's internal processing practices, with a particular focus on how personal data are managed within the ChatGPT system, thereby addressing the accuracy issues under Art. 5(1)(d) GDPR and the right of access under Art. 15 GDPR.

The complainant further demands a declaratory decision that formally recognizes the infringement of these key GDPR provisions. Moreover, it is requested that the authority impose corrective measures requiring OpenAI to fully comply with data subject requests, not merely blocking the display of inaccurate information, but by actually rectifying or erasing the erroneous data as mandated by Art. 16 and 17 GDPR. To ensure that these remedial steps result in sustained compliance, the complaint also calls for the imposition of an administrative fine that is proportionate to the severity of the infringement, serving both as a punitive and deterrent measure.[8] This comprehensive set of requests emphasizes that the technical limitations claimed by OpenAI do not justify non-compliance with the GDPR, and stresses the need for data protection obligations to be integrated into the AI system's operational framework.

[7] ..

[8] *noyb* 2024, p. 6.

# 4 Methodology

## 4.1 Case Analysis Method

The thesis primarily uses the IRAC case analysis method. Within the framework Issue, Rule, Application, Conclusion are used as a structured scaffolding for legal analysis. It facilitates a systematic approach to dissecting legal problems, ensuring clarity and coherence in reasoning.[1] The process begins with identifying the Issue, which involves pinpointing the central legal question arising from the case's facts, followed by the relevant legal principles or statutes applicable to the issue. Both have already been mentioned within chapter 3. Application and Analysis are covered in chapters 5 and 6.

## 4.2 Method and Procedure for Conducting Interviews

In order to deepen the analysis of the legal and technical challenges identified in the complaint against OpenAI, this study incorporates an empirical component based on qualitative expert interviews. The purpose of the interview process is to validate and refine the research findings, providing insights into the interaction between data protection law and the technological realities of large language models (LLMs). This section details the methodology used for the interviews, including the objectives, design, data collection methods, and a critical reflection on the data analysis process and its inherent limitations.

### 4.2.1 Interview Objective

The primary objective of the interviews was to substantiate my understanding of the complex technical and legal issues raised by *noyb*'s complaint. The interviewees were asked to comment on key challenges identified in the complaint, such as the difficulties of ensuring data accuracy and transparency in AI-based processing systems. Their perspectives serve to confirm whether the technical limitations

[1] Bittner 1990, p. 228.

and the associated legal risks, such as those concerning the right to erasure and the obligations under Art. 5 and 15 of the GDPR, are pervasive in current operational practices.

Given the interdisciplinary nature of this thesis, the interviews were also designed to capture insights from both technical experts and legal scholars. This dual perspective enables a more robust discussion regarding how privacy by design can be integrated into LLM development and what legal precedents might inform future regulatory compliance. The discussions sought to explore practical examples of how data inaccuracies in AI systems manifest and the implications for data subject rights. Furthermore, the interviews aimed to identify current trends, challenges, and potential solutions in the legal regulation of LLMs, thereby setting the stage for policy recommendations and future research. By engaging with important experts, the research sought to bridge the gap between theoretical legal frameworks and the operational realities of AI systems. These interviews provide an essential supplement to the documentary analysis of legal texts and case studies presented in previous chapters.

### 4.2.2 Interview Design

The design of the interview process was carefully tailored to ensure that both the legal and technical aspects of the complaint were thoroughly examined. I opted for a targeted selection of experts representing the two principal domains relevant to the study: data protection law and advanced AI technology. Two primary profiles were identified as ideal candidates for the interview process, being a technical expert and a legal scholar.

The leading legal expert providing insights in the first interview is Professor Boris Paal, who was selected due to his extensive background in the regulatory aspects of digital transformation. As the founder of the Chair for Law and Regulation of the Digital Transformation within the Technical University of Munich and the author of several influential GDPR-related articles, his insights were particularly relevant to understanding how the GDPR's technology-neutral design might affect the regulation of LLMs. The second interviewee is Markus Hupfauer, who was selected given his prominent role as a well-known specialist and writer in the field of information technology, specifically the application of artificial intelligence in cybersecurity. As a manager at KPMG, he brings practical expertise in advising about complex IT systems and understanding the implications of data processing in large-scale technological environments.

The interview questions were developed to cover a broad range of topics, including the operational mechanics of AI models, the challenges of ensuring GDPR compliance, the specific technical limitations that may hinder accurate data processing, and the practical implications of current regulatory practices. The questions were structured in a semi-structured format to allow flexibility and deep

exploration of the topics while ensuring that critical points were addressed consistently across both interviews. Examples of the key questions include:

- *How do the current technical limitations of AI models impact the ability to ensure data accuracy and facilitate data subject rights?*
- *What practical measures can be employed to integrate the principles of Data Protection by Design in LLM development?*
- *In your view, how effective are the current regulatory frameworks in addressing the privacy challenges posed by AI technologies?*

By structuring the interviews around these thematic areas, the research ensured a comprehensive exploration of the intersection between advanced AI systems and data protection law.

### 4.2.3 Data Collection Method

For both experts, video interviews were scheduled and conducted using secure conferencing tools. These sessions allowed for dynamic discussions, where relevant follow-up questions could be posed immediately in response to the experts' comments. Each video interview was recorded with the consent of the interviewee and subsequently transcribed to ensure that all pertinent details were accurately captured. The transcription process, as described in the Appendix under section A.1 (Interview Transcripts), involved both automated speech-to-text technology and subsequent manual editing to correct any inaccuracies and improve clarity. The data collection methods ensured a rich and diverse set of qualitative data, capturing both spontaneous and reflective insights from the experts.

### 4.2.4 Data Analysis and Limitations

The qualitative data collected from the interviews were analyzed using a thematic approach to identify recurring themes and insights across the various expert responses. This method was chosen because it allows for systematic categorization of the data into key topics that directly relate to the research questions of this thesis. For example, common themes such as the challenges of implementing "Privacy by Design" in complex AI systems, the limitations of current technical solutions like differential privacy and federated learning, and the legal justifications concerning claims of "technical impossibility" were all identified and discussed in depth throughout the analysis.

To begin with, all interviews were transcribed and compiled into a single data set. Following the transcription, the data were systematically organized into categories that aligned with the predetermined research topics. These categories were then grouped into broader themes that captured both the

technical and legal aspects of the complaint, revealing patterns and highlighting areas of consensus as well as divergence among the experts. Once the themes were established, they were interpreted in the context of the overall research objectives, integrating the insights derived from the interviews with the documentary and case analysis data discussed in previous chapters. This interpretative stage was crucial for developing a comprehensive understanding of the challenges inherent in aligning AI operations with GDPR requirements and provided empirical evidence that supported the findings of the case analysis.

However, this method was not without limitations. The scope of the interview process was limited by the selection of only two experts, chosen for their high topical relevance to the study and experience. While these individuals provided valuable insights, the absence of a spokesperson from OpenAI, for instance, meant that the study could not capture direct explanations or counter-arguments from the respondent's perspective. Additionally, temporal constraints also play a role, as the interviews reflect insights available at a specific moment in time, and subsequent technological or regulatory developments, or decisions made about the complaint itself might impact the relevance of these findings. Despite these challenges, the thematic analysis of the expert interviews constitutes a critical component of this thesis. It not only reinforces the conclusions drawn from the case analysis but also offers a well-rounded perspective on the multifaceted difficulties of integrating robust data protection measures within advanced AI systems.

# 5 Analysis and Interview Results

## 5.1 Case Analysis Procedure

In the following section, the approach which is to be applied to the case at hand is described. The actual procedure is then performed in Chapter 6 Discussion. The subsequent application phase within the IRAC scheme entails analyzing how the identified rules apply to the specific circumstances of the case, considering all pertinent facts and potential counter-arguments.[1] Afterwards, the conclusion sets to summarize the findings, providing a reasoned answer to the legal question posed. This method is widely recognized in legal education and practice for its effectiveness in organizing legal arguments and has been endorsed by various legal scholars and institutions.

### Direct Applicability of the Regulation

Within the context of the analysis the first step, to examine whether the GDPR (Regulation (EC) No. 679/2016) is directly applicable to the case, was formulated. As a regulation, the GDPR is binding throughout the entire European Union without requiring national transposition, meaning that all its provisions apply uniformly. This stage establishes that any processing activities affecting EU data subjects are subject to the GDPR, irrespective of the geographic location of the data controller, as long as the conditions set out in Art. 3 are met.

### Applicability of the GDPR

Next, the broader applicability of the GDPR is to be assessed by considering its temporal, territorial, and material scopes. Temporal scope is thereby to be determined by verifying that the processing activities under scrutiny occurred after the GDPR came into force in May 2018, ensuring the regulation's relevance. Territorial scope is examined using Art. 3 of the GDPR, which involves determining whether the data processing is related to the personal data of EU citizens or residents and whether

[1] PLT 2024.

the respondent offers services to the EU—even if the respondent is not established in the EU. Material scope involves confirming that the processing in question falls within the definition of personal data as provided in Art. 4(1) GDPR, and that the activities performed, such as collection, recording, organization, and use of personal data (for example, in powering ChatGPT), fully comply with the regulatory definitions without any applicable exemptions under Art. 2(2).

## Identification of Infringement(s)

Within this context the analysis sets out to identify the specific behaviors or practices by the data controller/processor that potentially breach key GDPR provisions. This includes evaluating whether the processing activities violate the overarching principles established in Art. 5 GDPR, particularly the requirements for fairness and accuracy, as well as the transparency and data subject rights outlined in Art. 15, 16, and 17 GDPR. Key issues examined include the refusal to delete personal data (breaching the right to erasure under Art. 17 GDPR), the failure to rectify inaccurate personal data (violating the obligation under Art. 5(1)(d) GDPR and the right to rectification as specified in Art. 16 GDPR), and the lack of clear, comprehensive information regarding data processing operations, which contravenes the transparency obligations under Art. 12 and 15 GDPR.

## Justification and Counter-Arguments

This step requires a critical examination of any justifications or counter-arguments put forward by OpenAI in response to the allegations. The analysis tests claims such as technical limitations—specifically the assertion that it is not technically feasible to selectively correct or delete certain pieces of personal data without impacting other stored information. In this context, it assesses these claims against the requirements of data accuracy and data minimization stipulated in the GDPR, as well as relevant case law and legal precedents that address similar issues. Furthermore, the analysis documents any discrepancies or failures in the legal rationale provided by OpenAI, noting that technical complexity cannot serve as a valid legal excuse under Art. 5(1)(d) and 25(1) GDPR.

## Expected Outcome and Legal Consequences

The final step forecasts the expected outcomes and potential legal consequences based on the earlier findings. This involves summarizing which GDPR provisions are likely to have been breached—for example, the right to erasure under Art. 17 and the principle of accuracy under Art. 5(1)(d). It further considers the legal implications for a data controller that persists in displaying inaccurate data

or fails to rectify or delete erroneous information, including possible enforcement actions by supervisory authorities. Anticipated remedial measures may include orders to mandate the full deletion or correction of personal data, comprehensive transparency regarding data processing practices, and the imposition of administrative fines proportionate to the severity of the infringement. This stage sets the foundation for understanding the broader legal and operational impacts of non-compliance in the context of modern AI systems such as ChatGPT.

Together, these steps provide a robust methodology that ensures a precise and legally anchored analysis of the case, thereby facilitating an in-depth examination of both the technical and regulatory challenges inherent to processing personal data within AI systems.

## 5.2 Expert Interviews: Comparative Analysis of Excerpts

The interviews conducted for this thesis provide complementary insights into both the legal and technical challenges arising in *noyb*'s complaint against OpenAI. On one side, the conversation with the legal expert seemed to illustrate the ongoing tension between the GDPR's technology-neutral framework and the its pragmatic difficulties that surface when it is applied to AI. On the other side, the interview with the tech expert highlighted the deeply embedded nature of personal data in large language models (LLMs) and the resulting impediments to meeting certain GDPR obligations, such as data erasure or rectification.

From a juridical standpoint, the legal expert seemed to emphasize the fundamental rights enshrined in the GDPR, e.g. that one can insist that any personal data relating oneself can be shown and processed correctly, whether on the internet or elsewhere as it is a fundamental right.[2] This statement underscores how personal data, once identified as inaccurate, should be corrected in compliance with Art. 16 and 17 GDPR. According to him, the notion of "technical impossibility" does not itself absolve a controller from fulfilling legal obligations: "I do not believe that such a justification is likely to stand. The GDPR is quite clear that technical complexity does not negate the law's requirements".[3] He also addressed strategic considerations often encountered in complaints, suggesting that filing a request for access first, rather than invoking multiple provisions at once, can be a deliberate, tactical choice: "I would indeed suspect that, as you suggested, there are strategic considerations, initially starting with the right of access, and based on the outcome, pursuing deletion and rectification claims".[4]

While he recognized that the GDPR was drafted to be technology-neutral, the legal expert further

[2]Paal 2025.
[3]Paal 2025.
[4]Paal 2025.

showed skepticism about the likelihood of a swift legislative reform tailored specifically to AI. According to his reasoning, "It is unrealistic to expect a quick update to the GDPR, as the legislative process at the European level is cumbersome and requires many compromises".[5] Instead, he pointed to guidance from data protection authorities and the European Data Protection Board as more feasible interim solutions. In his view, such interpretative instruments could offer clearer instructions on how to reconcile large-scale data-driven operations with core GDPR principles, such as data minimization and purpose limitation.

By contrast, the interview with the technology expert draws attention to the concrete technological barriers that emerge when individuals demand rectification or deletion of personal data within AI systems. "When I train a model using all sorts of text data from the internet, the model does not store that data one-to-one—it compresses it into a different format, much like a ZIP[6] file".[7] This analogy captures how personal data becomes interwoven with the model's parameters in ways that are not easily reversed. He explained that "the model's weights are set so that to change one data point, you would likely have to retrain the whole thing. That is currently neither cheap nor straightforward,"[8] illustrating why standard approaches to data deletion or correction cannot simply be mapped onto LLM architectures. Crucially, the technology expert believes that "no technical method exists to isolate a single piece of personal data and remove it without undermining the model's overall functionality".[9] He also noted that while advanced fine-tuning methods might appear to solve the problem superficially, they do not actually remove the underlying data from the system's learned representations, and can even introduce new vulnerabilities. In his view, "the only genuinely sure way to comply with a request for deletion is to retrain the entire system without [including] that data, which is immensely costly".[10]

Taken together, both interviews highlight a clear disjunction between the GDPR's insistence on data subject rights—whether in the form of rectification or erasure—and the fundamental design of large language models. As the legal expert's observations show, legal instruments are designed to ensure that controllers cannot claim a "technical impossibility" defense, whereas the tech expert' comments reflect the reality that advanced AI systems currently lack robust, cost-effective mechanisms for selective data correction or deletion. This tension underscores the need for ongoing dialogue among legislators, regulators, and AI developers, as well as potential new technologies that could offer more granular control over model parameters. While legal experts do not foresee an imminent revision

[5] Paal 2025.
[6] ZIP refers to a method of data compression that minimizes redundancy by encoding information more efficiently.
[7] Hupfauer 2025.
[8] Hupfauer 2025.
[9] Hupfauer 2025.
[10] Hupfauer 2025.

of the GDPR, there is wide recognition that practical strategies—such as pre-processing techniques, pseudonymization, or more nuanced interpretative guidance—might be necessary to align AI's capabilities with the existing data protection framework.

# 6 Discussion

The intersection between LLM technology and data protection law, as highlighted by the *noyb* complaint against OpenAI, underscores the broader challenges emerging from the rapidly expanding use of AI-driven solutions in various sectors. While these models enable innovative use cases and broad automation possibilities, they simultaneously introduce tensions around data accuracy, transparency, and the rights of individuals under the GDPR.

## 6.1 Contextual Discussion of Case Findings

Building on the arguments presented in Sections 3.1 and 3.2, it becomes clear that the growing integration of LLMs into digital services poses new and significant challenges for European data protection law. A tension emerges between the technical constraints of current AI systems and the extensive legal requirements of the General Data Protection Regulation (GDPR). At the core of this divide lies the question of whether it is even feasible to selectively rectify or erase personal data—pursuant to Art. 16 and 17 GDPR—once those data are embedded in LLM architectures, and what "adequate" implementation of these rights might look like in the AI context. In light of the present case involving *noyb*'s complaint against OpenAI, the following observations can be made.

### Direct Applicability

A central contention in the complaint from *noyb* is that the GDPR is directly applicable to OpenAI's activities. As a regulation rather than a directive, the GDPR is binding in its entirety across all EU Member States without requiring national implementation measures. Under Art. 3 GDPR, the regulation covers the processing of personal data even by entities established outside the EU, so long as the processing targets data subjects within the Union. In line with this principle, the complaint argues that ChatGPT's availability to EU users places OpenAI's data-processing activities firmly under the GDPR's jurisdiction.

Notably, during one expert interview, a similar point emerged. One interviewee observed that

"there is no reason to treat a US-based AI provider differently when its services are clearly offered to individuals in the EU".[1] Such views on the regulatory framwork reinforces the idea that offering any data-driven service to EU residents automatically triggers GDPR obligations, thereby reflecting the regulation's core aim of preventing non-EU organizations from circumventing European privacy standards through offshore operations. This uniformity of legal protection—regardless of a company's physical location—lies at the heart of the GDPR's extraterritorial reach and directly underpins the complaint's foundational claim that GDPR rules bind OpenAI's data-processing practices.

## Applicability of the GDPR

When establishing the applicability of the GDPR, three core dimensions, temporal, territorial, and material scope, come into focus again. First, the temporal scope is satisfied in this case as the data processing in question occurred between 2023 and 2024, well after the GDPR's entry into force on May 25, 2018. During one of the expert interviews, it was noted that "the regulation has been in effect for several years now, so there is no ambiguity about GDPR obligations applying to any processing carried out post-2018,"[2] reinforcing that activities within this timeframe are indisputably governed by the regulation.

Second, the territorial scope extends to OpenAI even though it is not headquartered within the EU. Art. 3(2) GDPR stipulates that the regulation applies to organizations offering goods or services to data subjects in the Union, or monitoring their behavior, irrespective of the organization's location. This principle directly counters any notion that OpenAI can claim exemption based on being a non-EU entity. Indeed, one interviewee also remarked by saying that "the same GDPR standards apply if a service is accessed by EU residents, even if the service provider is physically located elsewhere,"[3] underscoring the idea that transnational AI services like ChatGPT are not beyond the reach of European privacy laws.

Finally, the material scope is to also be fulfilled by virtue of the processing of personal data. The complaint alleges that ChatGPT handles identifiable information—such as names or birth dates—which Art. 4(1) GDPR defines as personal data. Operations like collection, organization, or usage of these data also thereby qualify as "processing" under Art. 4(2) GDPR. In the interviews, it was consistently highlighted that "any time an AI model absorbs user-specific details, it becomes subject to data protection rules, given that these details constitute personal data".[4] No exception under Art. 2(2), which might exempt purely domestic or household processing, shall apply here, since ChatGPT is

[1] Hupfauer 2025.
[2] Paal 2025.
[3] Hupfauer 2025.
[4] Paal 2025.

publicly accessible and intended for wide-scale interactions, thereby bringing the contested activities fully within the GDPR's scope.

## Infringement

As already mentioned within Section 3.2, the complaint identifies several potential GDPR violations arising from OpenAI's refusal or inability to remove or correct personal data within ChatGPT. First, it alleges that OpenAI's practice of "blocking" erroneous data rather than fully erasing it constitutes a breach of Art. 17 GDPR (right to erasure). According to evidence presented by *noyb*, the controller may limit the visibility of inaccurate or sensitive information but still retains it in the underlying model's parameters, thereby infringing upon the obligation to erase personal data without undue delay. One interviewee noted that "simply hiding personal data is not the same as deleting it—if the information still resides in the system's internal workings, the rights of the data subject remain unmet".[5] This approach runs counter to the spirit of the GDPR, which explicitly requires complete removal of the data in question to safeguard the data subject's privacy.

Secondly, there is an alleged failure to fulfill rectification and erasure rights under Art. 16 and 17 GDPR. In particular, the complaint highlights that OpenAI has made only limited disclosures about user account data and has not offered transparent information regarding personal data embedded in the training of the large language model. By restricting access to partial or incomplete records, the controller undermines the data subject's ability to rectify inaccuracies, such as an incorrect date of birth. As stressed during one of the expert interviews, "the principle of accuracy under Art. 5(1)(d) GDPR and the rights to rectification and erasure stand or fall together; if the system preserves false data, then it effectively denies the data subject's right to have it corrected or removed".[6] Merely blocking output does not equate to delivering the "right to be forgotten," because the erroneous information continues to reside within the AI model's learned parameters.

Finally, the complaint also questions whether personal data is being improperly processed or transmitted to third parties—potentially including other ChatGPT users—without meeting GDPR requirements such as informed consent or adherence to data minimization. This concern rests on the premise that insufficient transparency about which data is shared, how it is shared, and for what purpose breaches Art. 12 and 15 GDPR. Specifically, a lack of clarity surrounding the data used to train and run ChatGPT suggests that OpenAI's approach may not fulfill the transparency requirements of Art. 12(3) and 15(1)(3) GDPR. The overall picture painted by the complaint underscores a consistent shortfall in compliance with key GDPR provisions, primarily due to the manner in which personal

[5] Hupfauer 2025.
[6] Paal 2025.

data, once integrated into the model, cannot be selectively updated or purged.

## Justification

An already mentioned key argument advanced by OpenAI, as gathered from the complaint there, is that it is impossible on a technical level to remove or correct specific data, such as a user's date of birth or personal details, without corrupting the entire model. The principle of accuracy set forth in Art. 5(1)(d) GDPR, however, requires that personal data be kept correct and up to date, and the mere fact that a large language model is trained on vast datasets does not excuse the controller from fulfilling its obligations. One technical expert interviewed commented that "from a purely engineering standpoint, modifying a single data point in an AI model is extremely cumbersome and potentially necessitates retraining, yet this does not override the fundamental legal requirement to rectify inaccurate data".[7] The complaint thus disputes OpenAI's assumption that a "technical limitation" can serve as a valid rationale for non-compliance, stressing that controllers must devise compliance strategies aligned with the GDPR's demands—even if these involve technical or financial burdens.

Additionally, the selective provision of only certain data during access requests, while withholding information processed in the core model, appears to breach the transparency obligations found in Art. 12 and 15. One interviewee noted, "The controller cannot selectively disclose account-level data but conceal details processed in the language model itself, since the data subject has the right to know exactly how personal data is handled, especially in a system that potentially disseminates incorrect information".[8] The complaint itself emphasizes that no lawful basis has been identified to justify such omissions; a combination of partial disclosure and technical complexity does not meet the GDPR's rigorous standards, underscoring the overall lack of sufficient legal or factual justification for OpenAI's current practices.

## Possible Outcome

In light of the various alleged GDPR breaches, the complaint concludes that OpenAI's retention of inaccurate personal data and its failure to disclose relevant processing details amount to a persistent infringement of Art. 12, 15, and 17(1) GDPR. By continuing to block data rather than deleting it, and by omitting crucial information on how personal data is integrated into the language model and from which sources, OpenAI appears to deny data subjects their rights to rectification and erasure, as well as to full transparency about how their information is used. One expert interviewed commented that

[7] Hupfauer 2025.
[8] Paal 2025.

"the principle of accuracy under Art. 5(1)(d) [GDPR] remains violated as long as the erroneous data is embedded in the system, irrespective of whether the data is simply hidden from display".[9] Against this backdrop, the complaint anticipates that Austria's supervisory authority will find OpenAI in contravention of key GDPR provisions. Among the measures the complainant expects are (i) an order mandating the complete deletion or rectification of inaccurate personal data, (ii) directives to provide detailed transparency regarding data processing procedures, and (iii) potentially an administrative fine commensurate with the gravity of the violations. As further underscored in the interviews, "technical impossibility" is unlikely to stand as a persuasive legal defense; despite any genuine engineering obstacles, the GDPR does not exempt controllers from their obligation to ensure data subjects' rights remain enforceable.

## 6.2 Legal Implications of GDPR in the Case of LLMs

As previously indicated in Chapter 2, Large Language Models (LLMs) such as OpenAI's ChatGPT do not store personal data in a conventional database format ("rule-based chatbot") but rather incorporate it into a statistical network of weights during the training phase. This design makes selective access, erasure, or correction of specific personal data decidedly non-trivial.[10] The opaque boundary between trained model weights and original training data often complicates compliance with accountability (Art. 5(2) GDPR) and documentation (Art. 30 GDPR) requirements.[11] Furthermore, the GDPR's rights to rectification (Art. 16) and erasure (Art. 17) appear difficult to enforce once data is deeply embedded in the model's structure.

In the context of *noyb*'s complaint, OpenAI concedes that selectively removing inaccuracies—such as a wrong birth date—cannot be done without blocking all references to the individual in question.[12] This approach, which one interviewee described as "merely hiding the data while still storing it within the model's architecture,"[13] clashes with the principle of data accuracy (Art. 5(1)(d) GDPR) and the data subject's right to an effective remedy (Art. 77 and 79 GDPR). One might also set caution that a distinction must be made between truly "impossible" corrections and an unwillingness to develop innovative solutions.[14] As one interviewee pointed out, "technical complexity alone does not negate the GDPR's requirement to ensure accurate personal data".[15] Simply "blocking" rather than

[9] Hupfauer 2025.
[10] Yan et al. 2024, p. 9.
[11] Feretzakis et al. 2025, p. 9.
[12] *noyb* 2024.
[13] Hupfauer 2025.
[14] Feretzakis and Verykios 2024, p. 17.
[15] Paal 2025.

definitively removing erroneous data fails to meet GDPR demands, since the incorrect information remains embedded in the model.[16]

One possible strategy lies in privacy-preserving techniques such as "differential privacy" as mentioned in Section 2.2, which generate statistical outputs without pinpointing any individual.[17] Additional methods like homomorphic encryption, federated learning, and zero-knowledge proofs could help protect personal data during both training and inference.[18] Implementing these technologies may enable a functional compromise between the model's performance and GDPR compliance. On the architectural side, a modular LLM design—as proposed by some[19]—could allow more targeted data protection measures in specific components, but is currently not at the bleeding-edge of the technology. This is especially relevant given one interviewee's assertion that "once data is intertwined with the overall network, selectively purging it becomes prohibitively complex".[20] Concurrently, automated compliance tools can track and document obligations like data minimization (Art. 5(1)(c)), purpose limitation (Art. 5(1)(b)), and data subject rights (Art. 15–17) within the development workflow.[21] Meta and OpenAI, among others, are currently experimenting with semantic filters that could at least prevent the inadvertent release of sensitive information[22], techniques which however appear to have a negative impact on LLM performance in terms of both output quality and potentially execution speed.[23] Fullfilling the principle of data minimization in that regard also requires preprocessing training data to remove, pseudonymize, or fully anonymize personal details. However, as one expert noted, "simply anonymizing data is not foolproof, because cross-referencing with other sources can undermine anonymity".[24] This risk significantly increases the technical and organizational measures needed to uphold privacy obligations.[25] Overall, while the GDPR lays out robust rights for data subjects, current AI architectures complicate the real-world enforceability of those rights, underscoring the urgency of developing advanced privacy-preserving methodologies and governance strategies.

[16] Laurelli 2024, p. 14.
[17] Feretzakis et al. 2025, p. 12.
[18] Mittal et al. 2024, p. 739.
[19] Cory et al. 2025, p. 82.
[20] Hupfauer 2025.
[21] Mittal et al. 2024, p. 740.
[22] Miranda et al. 2025, p. 11.
[23] Hupfauer 2025.
[24] Paal 2025.
[25] Rodriguez et al. 2024, p. 3.

# 7 Conclusion and Key Findings

It seems that in the last years, Large Language Models (LLMs) have significantly expanded the horizons of AI-driven applications, yet they also bring unique regulatory and ethical challenges with them. The *noyb* complaint against OpenAI thereby illustrates the real-world complexity of enforcing GDPR provisions within large-scale, data-intensive systems. By examining both the legal requirements and the technical realities of LLMs, this thesis has highlighted the urgent need for innovative solutions that reconcile model performance with data protection rights.

One of the core challenges identified revolved around enforcing the GDPR's data subject rights—rectification, erasure, and transparency—when personal data is embedded in the deep architecture of an LLM. OpenAI's rationale seems to be that selective data removal is infeasible, emphasizing the difficulty of balancing technological feasibility with the GDPR's stringent obligations. Expert insights have consequently confirmed that while certain privacy-preserving methods exist, they are not yet widely adopted at scale. It seems necessary that developers, regulators, and legal experts must further coordinate to find new frameworks and technical approaches (e.g., differential privacy, modular LLM design) that enable compliance without undermining AI functionality. Within the European Data Protection Board (edpb) the "ChatGPT Taskforce" also arrived at a similar conclusion.

One distinguishing feature of this specific complaint, compared to other cases involving major gatekeeper technology companies lies in the highly integrated nature of the data within the actual language model. Prior EU-driven complaints against large tech platforms typically concerned structured data or user-provided information e.g. on social media apps. By contrast, LLMs merge vast amounts of user-generated and open-source data into a single probabilistic model, thereby complicating selective deletion or rectification. This difference calls for deeper interdisciplinary research into privacy-enhancing technologies, model-specific data governance, and standardized auditing frameworks. Further studies may also need to investigate whether novel methods, such as "machine unlearning" of already compressed LLM data can actually be scaled up to address GDPR obligations in an effective manner. When it comes to potential consequences, if the practices around LLMs are not adequately aligned with GDPR requirements, significant legal, operational, and ethical consequences

could ensue. Regulatory authorities might initiate enforcement actions, including substantial fines or sanctions, setting precedents that influence future AI-related compliance cases. On an operational level, AI providers may be compelled to reassess their data processing and training methodologies, potentially necessitating the adoption of entirely new privacy-focused architectures and practices. In this regard a sustained failure to comply with data protection laws also risks undermining public trust in artificial intelligence technologies, particularly if inaccurate or sensitive personal information remains accessible or uncorrected within widely used AI systems. In the research area directly involved with LLMs, practitioners may have to come up with novel ways to source and process training data either with consent, or without personal data at all.

As LLM-based services continue to gain prominence, European Data Protection Authorities (DPAs) will increasingly need to adapt their oversight strategies to address AI-specific scenarios. This could involve issuing targeted guidelines for AI developers regarding model design, data minimization practices, and transparency obligations. Additionally, DPAs may also play a critical role in promoting technological innovation by fostering cooperation among regulators, academia, and industry stakeholders to accelerate the adoption of privacy-by-design methodologies. Furthermore, given the inherently international scope of these AI services, strengthening cross-border enforcement through enhanced collaboration among DPAs seems to be essential for ensuring a sustainable, consistent and effective application of GDPR principles across its jurisdictions.

# A Appendix

## A.1 Interview Transcripts

Interviews conducted using a videoconferencing tool were recorded, had their conversation transcribed after the meeting concluded using WhisperKit with the Whisper model `openai_whisper-large-v3-v20240930` for transcription, then manually edited for clarity, grammar and portions which were still incorrect from automatic transcription alone.

### A.1.1 Interview with Boris Paal

The following is a transcript of the interview conducted on February 26, 2025.
The interview participants were:

- **Lionel Merz**, author of this thesis.
- **Prof. Dr. Boris P. Paal, M.Jur. (Oxford)**, introduced under section 4.2.2 in Interview Design.

**LM:** Guten Tag. Ja, freut mich, dass es geklappt hat. Und ich würde jetzt einfach anfangen und eine kleine Einführung geben, vielleicht für den Fall, und dann zu meinen Fragen kommen.
**LM:** Also, es ist ja so, dass *noyb* von Schrems öfters Leute vertritt und in dem Fall ging es darum, dass OpenAI falsche Daten über eine Person des öffentlichen Lebens veröffentlicht hat im Rahmen von ChatGPT, also konkret DS-GVO Artikel 12, da wurden nach einer Anfrage die Daten nicht vollständig herausgegeben und darüber geht dann diese Beschwerde. Und OpenAI sagt jetzt, dass es technisch nicht möglich wäre, dass sie sich quasi an diese Regeln halten. Also inwiefern wäre das vertretbar als Argument, dass diese Vorschriften einfach nicht eingehalten werden können, weil GPT nun das so in sich hereingebacken hat, dass die Daten nicht quasi gesäubert werden können nach einer Anfrage oder sogar korrigiert werden können oder eben gelöscht werden können.
***Note***: At this point, the interviewee's audio track in the recording file got corrupted until the point further along in the interview marked below. For transparency's sake, a brief listing of the topics discussed during this timeframe is provided below from my memory, but not referenced in great extent within the thesis itself.

- Zu urteilen ist: Sind personenbezogene Daten enthalten; wenn ja, ist DS-GVO anwendbar; wenn anwendbar, welche Implikationen beinhaltet dies?
- "Technische Unmöglichkeit" ist grundsätzlich kein valides Argument
- Verbot- und Erlaubnisvorbehalt Art. 6 DS-GVO
- Interessenabwägung: Schutz eines personenbezogenen Datums gegenüber Interesse der Allgemeinheit, wobei es sicherlich eine Rolle spielen wird, dass das Datum öffentlich zugänglich war
- Berichtigungsanspruch: korrekt, existiert grundsätzlich für Betroffene

**LM:** Das habe ich mir tatsächlich auch schon gedacht, also wenig überraschend in dem Fall. OpenAI sagt jetzt weiterhin, es gebe ja, also in der Beschwerde wurde das von *noyb* aufgefasst, Freedom of Expression oder Freedom to Inform the Public. Und inwiefern lässt sich das abwägen, wenn jetzt hier auf der Input-Ebene quasi ein an sich frei zugängliches Datum kam, aber es wurde wahrscheinlich nicht selbst veröffentlicht, sondern war halt zum Beispiel Bestandteil eines Wikipedia-Artikels.
(***Note****: The recording file is still broken until partway into this next answer*)
**BP:** ... personenbezogener Anspruch. Also ich kann verlangen, jeder kann verlangen, dass personenbezogene Daten, die einen betreffen, korrekt im Internet angezeigt oder wo auch immer angezeigt und verarbeitet werden. Also das sind zwei unterschiedliche Perspektiven und dieses Freedom of

Expression, also Meinungsfreiheit, Meinungsäußerungsfreiheit Argument ist sicherlich eines, das im Rahmen der Interessenabwägung, also Zulässigkeit der Datenverarbeitung eine Rolle spielt.

**LM:** Okay. Daran möchte ich noch anknüpfen. Da habe ich vielleicht falsch formuliert. Es geht nicht um Schrems selber, sondern eine andere Person des öffentlichen Lebens, die aus der Beschwerde rausgeschwärzt wurde. Aber das nur, damit ich es nicht falsch rüberbringe. Okay. Jetzt wäre noch meine Frage, warum *noyb* vielleicht sich nicht bezogen hat auf weitere Artikel. Also konkret waren ja in der Beschwerde nur gelistet quasi 12(3) und 15 [Art. DS-GVO]. Und jetzt könnte man noch fragen, warum sie nicht gleich noch 14, 16, 17 [Art. DS-GVO] alle mit reinnehmen in die Anfrage. Gibt es da vielleicht taktische Gründe, nach einem Ersterfolg weiterzumachen oder sowas in die Richtung?

**BP:** Ich kann es nicht sagen, ich würde aber auch mit Sicherheit sagen, ich würde in der Tat vermuten, wie auch Sie angedeutet haben, strategische Überlegungen, dass man zunächst mal mit dem Auskunftsanspruch startet und auf diesen Auskunftsanspruch, also Ergebnis des Auskunftsanspruchs wäre ja, personenbezogene Daten und dann zweite Information, welche, darauf gestützt dann, Löschungsansprüche, Berichtigungsansprüche geltend zu machen. Das wäre meine Vermutung, das wäre ein übliches prozesstaktisches Vorgehen. Also ohne jede Bewertung, ist ganz normal, da ist der Streitwert erst mal geringer und dann schaut man, das ist kostengünstiger und dann schaut man, wie sich das entwickelt.

**LM:** Und vielleicht noch ein bisschen Ausblick auf die mögliche Änderung auf diese Technologie als Reaktion der EU. Also im Prinzip ist es ja ziemlich offensichtlich, dass Grundsätze der Datenminimierung und Transparenz und Zweckbindung alles schwer eingehalten werden kann, wenn es um ein generelles KI-Modell gehen soll, was wirklich alles wissen soll und auch ausspucken soll. Ist dann vielleicht eine Aktualisierung der DS-GVO notwendig, wenn der Gesetzgeber vermutet, es wäre jetzt angemessen, die anzuwenden auf diese Technologie?

**BP:** Also vielleicht vorweggeschickt, die DS-GVO ist, man sagt technologieneutral formuliert, offen formuliert. Aber sie ist eben auch nicht mit einem Fokus auf KI-Technologien konstruiert. Und deswegen gibt es ganz offensichtliche Brennpunkte im Zusammenspiel von KI auf der einen Seite und DS-GVO auf der anderen Seite. Sie haben es bereits benannt, Transparenz ist ein Problem, also Blackbox-Prinzip. Das Prinzip der Datenminimierung gegenüber den Big Data Needs auf der anderen Seite. Und auch die Frage der Zweckänderung, also ich habe die Daten für den einen Zweck erhoben und jetzt will ich sie für die Zwecke nutzen. Ich glaube, dass hier vieles möglich ist im Wege der Auslegung, der Anwendung der DS-GVO. Datenminimierung bedeutet jetzt nicht zwingend, ich muss immer alle Daten löschen, sondern ich muss gute Gründe haben, wenn ich die Daten speichere. Aber trotzdem haben wir dann natürlich Rechtsunsicherheiten und potenzielle Hemmnisse für KI-Entwicklung. Und jetzt war Ihre Frage ja: Besteht da Hoffnung, Erwartung, Aussicht

darauf, dass die DS-GVO modifiziert wird, also angepasst wird mit Blick auf diese KI-Bedürfnisse? Und da wäre meine Antwort: Das glaube ich nicht, dass das passiert. Das ist deswegen unrealistisch, weil der Gesetzgebungsprozess auf europäischer Ebene, wie Sie wissen, ein sehr mühsamer ist. Ein sehr zeitaufwendiger, auch vieler Kompromisse bedarf und auch natürlich unterschiedliche Interessen – Datenschutz auf der einen Seite, Big Data, Big Tech-Unternehmen auf der anderen Seite – und deswegen glaube ich, dass der Weg, den wir gehen können, realistischerweise derjenige ist, der Auslegung auf Vorgaben des Europäischen Datenschutzausschusses, also von Behörden, die Guidelines, Richtlinien, Leitlinien herausgeben. Das ist, glaube ich, der realistischere Weg, als auf eine Novellierung der DS-GVO zu hoffen.

**LM:** Ja, soweit nachvollziehbar. Jetzt wäre noch die Frage, wenn sich aber nun nichts verändert, Technologieneutralität beibehalten wird, was passiert denn jetzt konkret in diesem Fall? Also natürlich schwer vorherzusehen, aber was wären denn mögliche Strafmaße? Also von Korrektur dieses einen Datums bis hin zu Neutraining nach jeder einzelnen Beschwerde ist ja alles vorstellbar – oder gleich ganz Einstellen des Angebots am europäischen Markt.

**BP:** Also das ist sicherlich die Königsfrage in dem Zusammenhang, die sich in dem Fall, in dem Punkt sich alles kristallisiert. Meine Vermutung, meine Annahme wäre, dass man dazu kommen wird, dass so ein LLM tatsächlich personenbezogene Daten enthält und damit sind grundsätzlich DS-GVO-Vorschriften anwendbar. Aber jetzt ist auf der Rechtsfolge-Ebene durchaus zu überlegen: Wie kann ich Berichtigung umsetzen? Ich glaube nicht, dass es dazu führt und dazu führen sollte, dass das LLM als solches gelöscht wird. Das kostet ja Ressourcen, das ist ein finanzielles, auch ein Nachhaltigkeitsthema. Deswegen glaube ich, es wird verlangt werden, dass die LLM-Anbieter entsprechende Vorkehrungen treffen, also dass sie ihrerseits Mögliche tun, sowohl im Vorfeld als auch in Reaktionen daraus. Und dass aber auch berücksichtigt wird, dass so ein LLM eben keine Wahrheitsmaschine ist. Also das ist eine Wahrscheinlichkeitsmaschine und keine Wahrheitsmaschine. Und da muss man ja auch fragen: Erwarten die Nutzer denn tatsächlich, dass das ein richtiges Ergebnis immer ist? Oder wird mit eingepreist, dass das womöglich gar nicht zutreffend ist. Also wenn es bei Wikipedia steht, muss es ja auch nicht zwingend richtig sein. Und ich glaube, da ist dann der Blick auf die Empfängerseite, auf die Rezipienten zu richten, was eine berechtigte Nutzererwartung ist. Und im Zusammenspiel dieser verschiedenen Aspekte hoffe ich, dass die Gerichte eine abgewogene Entscheidung treffen, die zum einen den Persönlichkeitsschutz von Herrn Schrems und anderen angemessen reflektiert, aber auf der anderen Seite auch reflektiert, was ist ein solches LLM, wie sollte es genutzt werden? Und da sind wir bei der grundsätzlichen Frage KI-Kompetenz zu verstehen. Und das wollen wir natürlich auch in der Ausbildung auf allen Ebenen, Stichwort Lifelong Learning, dass die Nutzenden verstehen, was kann ein solches LLM und was kann es eben nicht, was ist überhaupt

dessen Aufgabe. Ich glaube, das ist ein ganz wichtiger Punkt und häufig ist ja so DS-GVO-Bashing, also der Datenschutz ist an allem schuld. Ich glaube, eine klügere Anwendung des Datenschutzes kann hier viele Probleme überwinden, aber wir haben ganz offensichtliche Spannungsverhältnisse hier.

**LM:** Ja, sehe ich ähnlich. Dann vielleicht noch abrundend, wie das im Verhältnis steht zu einfach den Nutzern, aktiven Nutzern, die ja selbst die Plattform potenziell einfach als Wahrheit ansehen. Das sind ja dann potenziell einfach komplett unbetroffene Leute sozusagen, die das Produkt nutzen mögen oder auch nicht. Also, wie unterscheidet sich das vielleicht von einem typischen Fall, wo man ganz klar sagen kann, ich habe als Firma eine Privacy Policy veröffentlicht und daran müssen sich Nutzer halten. Hier haben wir ja keine Nutzer mehr in dem Sinne.

**BP:** Ja, also das kommt darauf an. Man könnte sich unterschiedliche Sachverhalte vorstellen. Aber wir müssen uns ja vorstellen, dass jeder, der in einem LLM oder in die Schnittstelle, ein KI-System, personengezogene Daten eingibt, auch betroffene Person natürlich ist. Weil die gehen da rein, die werden vielleicht genutzt für das Feintuning, für das weitere Training der KI. Ich glaube, ein ganz wichtiger Punkt, den wir uns hier anschauen müssen ist: Was sind die Rollen der verschiedenen Akteure in einer solchen Konstellation? Sie haben schon zu Recht angesprochen, Privacy Policy der Unternehmen ist ein ganz wichtiger Punkt, wie ein Unternehmen das für sich definiert, aber auch Nutzende sich klar machen, was passiert da eigentlich und was passiert mit meinen betroffenen, mit meinen personenbezogenen Daten?

**LM:** Okay. Ja, ich sehe, jetzt haben wir schon die Zeit erreicht. Dann danke ich Ihnen vielmals.

### A.1.2 Interview with Markus Hupfauer

The following is a transcript of the interview conducted on April 1, 2025.

The interview participants were:

- **Lionel Merz**, author of this thesis.
- **Markus Hupfauer** of KPMG, introduced under section 4.2.2 in Interview Design.

**LM:** Genau. Vorneweg, ich würde uns aufnehmen, dann kann ich es akkurater wiedergeben. Ich denke, das wird passen.

**MH:** Gar kein Problem.

**LM:** Also, ich würde es so machen, ich würde einen Überblick geben über diesen Fall und dann ein paar Fragen stellen, damit ich eine Quelle habe für mein Verständnis. Und zwar hat letztes Jahr im April schon die NGO *noyb*, also von Max Schrems, eine Beschwerde gestellt in Österreich der Datenschutzbehörde gegenüber mit einer Anfrage für Artikel 12, DS-GVO, also die Datenanfrage und Artikel 15 mit dem Right of Access und Artikel 5 mit der generellen Verarbeitung der personenbezogenen Daten bei OpenAI. Und jetzt ist es so, dass OpenAI dann gesagt hat, das ist Gegenstand der aktuellen Forschung und sie können sich quasi nicht dran halten. Und das Interessante sozusagen, es ging um ein Geburtsdatum von der Person, die vertreten wird von dieser NGO und das kann halt nicht technisch gesehen wieder rausgelöscht werden aus dem Sprachmodell. Sehe ich das richtig?

**MH:** Ja, also wie trainiert so ein Modell ist da, glaube ich, die unterliegende Frage. Und ich verwende ja mehr oder weniger in den ersten Phasen von der Erstellung eines Large Language Models einfach jeden Text, den ich finden kann. Und das Ziel des Modells ist, das nächste Wort vorauszusagen. Das heißt, in der ersten Phase sind diese Modelle nicht wie wir die kennen, dass ich sagen kann, ChatGPT schreibe mir das und das, sondern eher einen Satz vervollständigen. Ich gebe Ihnen ein bisschen Text und das Modell kann den dann logisch vervollständigen. Und hier ist es eben so, ich habe natürlich im Internet personenbezogene Daten, Google-Suchergebnisse etc. Und wenn ich das Modell eben auf diesen Daten trainiere, die vorher sagen zu lassen, dann nimmt das Modell ja diese Daten nicht eins zu eins und speichert die in irgendeiner Art und Weise, sondern in diesem Modell, man kann sich das —erzähle ich immer—so ein bisschen wie eine ZIP-Datei vorstellen. Sie komprimieren die Informationen in anderes Format einfach, das wesentlich effizienter ist. Und dabei geht Ihnen natürlich eine gewisse Nuance verloren. Sie können zum Beispiel aus einer ZIP-Datei, jetzt gibt mittlerweile technische Möglichkeiten, aber schwierig eine einzelne Datei aus dem Inhalt rauslöschen, ohne die neu zu komprimieren. Wenn du das machst, ist jetzt im Hintergrund und sieht man das nicht mehr. Aber normalerweise ist es eben so, dass das, was in das Modell reinkommt, dann Teil von diesem Modell wird. Und ich kann nicht mehr genau sagen, ich zum Beispiel, ich habe am 17. April Geburtstag, ändere jetzt exakt diesen Datensatz, denn der ist in dieses große ganze Modell eingeflossen. Und das kann man dann auch nicht wieder rausholen.

**LM:** Also sprich, die Gewichte sind so festgesetzt, dass man dann neu trainieren müsste, vermutlich, also für jedes einzelne.

**MH:** Ganz genau. Also gerade eben in diesen Themen ist es eben so, ich habe die einzelnen Model-Weights und die Activations zwischen diesen Weights und in diesen Texten oder in diesen

numerischen Repräsentationen der Daten, der statistischen Häufigkeit, wie diese Daten zueinander auftreten, enkodiere ich natürlich auch personenbezogene Daten. Es ist einfach nur eine andere Darstellung formt, die jetzt eine Text-Datei. Und hier kann ich einzelne Sachen nicht rausnehmen. Jetzt kommt häufig der Gedanke, okay, ich gehe einfach in einem späteren Schritt hin und sage, Trainingsziel ist es jetzt nicht, 17. April zu sagen. Ich kann das Modell ja sozusagen immer wieder mit einer zusätzlichen Schicht weiter trainieren. Das Problem ist, es gibt sehr gute Möglichkeiten dann die Bad Words sozusagen zu extrahieren. Das heißt, trainiere ich dieses Modell gezielt personenbezogene Daten nicht zu sagen von jetzt Leuten, die die gelöscht haben wollten, ann ich als Angreifer sehr wohl das Modell dazu bringen, mir die Schranken, die es kennt, zu nennen, was wiederum die Personen bezogenen Daten sind. Das heißt, ein nachträgliches Fine-Tuning, wie man das so gern dazu sagt, der um die personenbezogenen Daten heraus zu tunen, ist eigentlich auch technisch nicht möglich.

**LM:** Ja, und also was mir als Gedanke aufkam, das ist ja prinzipiell immer noch nicht selbst so, wenn Sie das dann schaffen, dass man mit keiner Attacke quasi an die versteckten Informationen kommt, ist das ja eigentlich immer noch keine Einhaltung der DS-GVO, weil die ja eine Löschung auch innerhalb der Firma verlangt. Das würde ja das mit einbeziehen, oder?

**MH:** Absolut. Denn der Layer, der die Daten oder der diesen Finetuner enthält, der das Modell unterdrücken soll, muss selber wieder die Daten halten, um es unterdrücken zu können. Und dann möchte ich OpenAI rechtgeben. Das ist Stand aktiver Forschung. Das ist, glaube ich, keine Fehlaussage. Und ich kenne auch niemand anderen, der hier eine Lösung kennt, um personenbezogene Daten aus Modellen herauszulöschen. Das muss nicht allinklusiv sein, aber soweit ich das beurteilen kann, haben wir da wenig. Und jetzt ist es eben so, wenn wir, wenn ich mit Mandanten spreche, der Punkt ist eigentlich der, sobald ich personenbezogene Daten in dieses Modell reintrainiere, enthält dieses Gesamtmodell diese personenbezogene Daten. Und damit bin ich natürlich, wenn ich eine Anfrage bekomme, muss ich das Gesamtmodell löschen und ohne die Daten neu trainieren. Das ist unseres Erachtens, meines persönlichen Erachtens, im Moment glaube ich der einzige Weg, diese DS-GVO-Themen abzudecken.

**LM:** Von Open AI kam noch ein Gegenargument in Richtung Freedom of Expression oder Freedom to Inform the Public. Also es handelt sich ja so gesehen um frei zugängliche Daten, aber in der Regel auch um ein selbst veröffentliches Datum. Also natürlich gibt es so gesehen keine Zustimmung, aber da haben sie eben versucht, so ein Argument heraufzuspinnen. Was halten Sie davon?

**MH:** Ohne da jetzt direkt die Argumentationslinie von Open AI anzugreifen, aber vielleicht im Allgemeinen es ist so, dass der DS-GVO das relativ egal ist. Personenbezogene Daten sind personenbezogene Daten. Explizit in der DS-GVO sind auch die Punkte des Data-Minings erwähnt, dass ich

aus unzusammenhängenden Daten Zusammenhänge extrapoliere, ist als spezieller Use Case in der DS-GVO genannt, der Zustimmung bedarf und besonderer Vorsichtsmaßnahmen. Und ich würde stark vermuten, dass das in ein Large Language Model zu trainieren, das über alle Daten, die ihm gegeben worden sind, dann Reasoning betreiben kann. Das für mich nicht unwahrscheinlich klingt, dass es da drunter fallen könnte.

**LM:** Ja, Genau. Die DS-GVO ist erstmal technologieoffen oder technologieneutral formuliert. Wäre denn denkbar, dass eine Anpassung als Reaktion auf Sprachmodelle kommt, weil an sich ist es ja, wie wir jetzt festgestellt haben, nicht wirklich kompatibel. Also man müsste ja dann entweder davon ausgehen, dass wirklich tagtäglich Strafen reinprasseln für alles und jeden, der sich quasi mit der Technologie und irgendwelchen Datensätzen, die öffentlich zugänglich sind, befasst oder dass sie die Benutzung innerhalb der EU zum Beispiel dann einstampfen. Ist es denn in irgendeiner Form realistisch, dass da sowas kommt?

**MH:** Also mein Gefühl sagt nein. Die EU hat ja den AI Act veröffentlicht, in dem es sehr viele harmonisierende Vorschriften gibt, die nicht Vorschriften der DS-GVO harmonisiert haben. Das wäre an dieser Stelle möglich gewesen und man tat das nicht. Ob das Absicht war oder ob das ein Oversight war, das kann ich natürlich nicht beurteilen. Vielmehr glaube ich aber persönlich daran, dass die Provider in der Pflicht wären, das ordentlich zu tun. Denn ich kann ja auch einfach nicht mit personenbezogenen Daten trainieren. Das Wort der Pseudonymisierung kennt die DS-GVO ja als Sicherheitsmaßnahme. Es ist sehr wohl denkbar in einem Vorverarbeitungsschritt der Daten, bevor ich die ins Training gebe, aktiv mit Filtern – das muss ja auch keine 100%-Lösung sein, aber wenn es eine 90%-Lösung ist, ist es deutlich besser, wie was es jetzt ist. Wenn ich mit Vorfiltern meine Daten nach klassischen Pattern von Personen bezogenen Daten, also Namen, Adressen, Postleitzahlen, Geburtsdaten, das sind alles einfach standardisiert zur erkennende Sachen. Und ich speichere Dinge aus diesen Daten heraus und erzeuge ich sage mal pseudonyme Datenpunkte, die so ähnlich sind, die logisch weiterhin miteinander funktionieren, dass immer wenn der Punkt Markus Hupfauer aufkommt, dann das Geburtsdatum, was dazu, ich sage mal gefaked oder pseudonymisiert wird, halt ein anderes ist als der 17. April. Und ich halte mir als Provider dann eben diesen Mapping-Datensatz vor. Und wenn jemand seine personenbezogenenen Daten gelöscht haben will, dann lösche ich einfach in dieser Referenz-Tabelle die personenbezogenenen Daten. Und in meinem Modell verbleiben lediglich die pseudonymisierten Daten, die ich nicht ent-pseudonymisieren kann. Das Modell funktioniert genauso gut, es ist irrelevant für ChatGPT, ob ich, wann ich Geburtstag habe. Da wird es natürlich Ausnahmen geben von historisch wichtigen Daten, wo ich eben die realen Daten benötige. Aber die werden ja auch in dieser Pseudonymisierungs-Tabelle weiterhin vorhanden sein, die lässt ja niemand löschen. Das ist jetzt technisch nicht unmöglich, würde ich behaupten. Natürlich, ich bin kein Anbieter eines

Großsprachsmodells, aber von dem, was wir sehen, glaube ich, dass das technisch eine Möglichkeit wäre. Fügt halt zusätzliche Kosten, zusätzlichen Aufwand hinzu, ist träger als einfach die Daten direkt zu trainieren. Ich könnte mir auch vorstellen, dass die großen Anbieter nicht im Moment sämtliche Trainingsdaten speichern, denn die würden quasi dafür eine Kopie des Internets vorhalten müssen, das sie crawlen. Und ich glaube auch, dass deshalb die Probleme von Open AI da sind, auch so eine Anfrage zu beantworten. Denn ja, Speicherplatz ist günstiger geworden die letzten Jahrzehnte, aber halt immer noch nicht so billig, dass ich einfach mal eine Hard Copy vom Internet rumliegen lasse, "just in case".

**LM:** Mh-hm. Wobei jetzt sicherlich ein Teil auch sein wird, bei dem, sagen wir mal, Sträuben der Anbieter bisher. Also jetzt in dem Fall von *noyb*, war es so, die vertreten eine Person des öffentlichen Lebens, die nicht weiter genannt wird. Und da ist ja bestimmt auch im Geschäftsinteresse der Anbieter, dass sie quasi doch akkurate Daten hätten. Also das hilft ja bestimmt nicht in dem Fall, dass die Sprachmodelle für Nutzer so wirken sollen wie ein schnelles Nachschlageverzeichnis im Prinzip.

**MH:** Absolut. Aber genau das gleiche Problem hat Google auch. Wenn ich nicht möchte, dass Google meine Website und meinen Namen indiziert, dann kann ich durch die EU meine Rechte geltend machen und Google hält sich da auch dran. Und Google gefällt das bestimmt auch nicht. Und da muss ich sagen gleiches Recht für alle, denn ChatGPT ist für viele mittlerweile eine glorifizierte Suchmaschine. Und dann wüsste ich nicht, wieso sich der eine US-Konzern den Regeln unterwirft, wenn es der andere nicht tut, weil die Suchtechnologie technologisch deutlich ineffizienter ist wie die vom anderen. Also da bin ich, das ist eine klare Sache eigentlich meiner Meinung nach, dass sich hier einfach ChatGPT nicht dranhält. Und das ist auch kein ChatGPT-exklusives Problem, das ist bei Google, bei AWS, bei Mistral von den Franzosen bei Anthropic, bei allen.

**LM:** Ja, ein Punkt vielleicht noch. Es gibt teilweise dieses Argument, dass die Nutzer ja den Nutzungsbedingungen zugestimmt hätten. Das bezieht sich aber auf, was OpenAI tatsächlich in der Anfrage zuerst falsch verstanden hat, bezieht sich auf die Auskünfte über "welche Themen habe ich angesprochen in einem Sprachmodell während ich registrierter Nutzer war". Aber das trifft ja ziemlich sicher nicht zu auf irgendwelche Daten, wenn die Nutzer zufälligerweise auch ein Benutzerkonto haben sollten bei diesen Firmen.

**MH:** Also das ist so eine ganz interessante Thematik. Da gab es doch auch den Case, ich weiß gar nicht mit irgendeinem Roboter-Taxi, wo mit einem Vertrag irgendwelche Liability ausgeschlossen wurde. Ich bin mir nicht sicher, aber es gab auch bei irgendeinem Fall, ich bin mir gar nicht mehr sicher, wo es war, dass auch jemand mit einem US-Konzern und mit irgendeiner Tochter einen Vertrag hatte, der quasi Liability freistellen sollte.

**LM:** Ich glaube, das war Disney, kann das sein?

**MH:** Ah, das stimmt, das war Disney+ glaube ich.

**LM:** Disney+ und das Restaurant mit der Allergie, ja.

**MH:** Ja, und das ist ja auch so eine ganz ähnliche Sache, weil ich hier irgendwo etwas zugestimmt habe und jetzt bin ich kein Rechtsanwalt, ich bin ein Informatiker, also mit dem Hintergrund bitte nehmen. Aber die Zustimmung bezieht sich ja auf mein Rechtsgeschäft, das ich hier gerade abschließe und das geht um den Dienstvertrag, dass ich einen über die Website dargebotenen Dienst konsumiere. Inwieweit der meine Zustimmung der DS-GVO ermöglicht oder gleichstellt, das bezweifle ich ganz stark, denn ich habe ja Double Opt-in und sonstiges in der DS-GVO als Anforderung mit einem dauerhaften Widerspruchsrecht und ich kann mir nicht vorstellen, dass man andere wichtige Gründe vorschieben kann, warum sie die Daten denn behalten müssten. Also es ist ja keine Rechnungsaufbewahrungspflicht oder sowas, das OpenAI daran hindern könnte. Also da habe ich das Gefühl, dass man hier sich versucht an jeden Strohhalm zu klammern, aber die Wahrheit glaube ich ist eine ganz nüchterne, dass die Regeln ganz klar auch für OpenAI gelten, und ich kann mir nicht vorstellen, dass die EU hier die DS-GVO stark verbessern würde, weil wenn ich dann als Unternehmen mich nicht mehr die DS-GVO halten will, dann enkodiere ich einfach alle meine personenbezogenen Daten von Kunden in so ein Modell und speichere die da und dann bin ich raus mit der DS-GVO. Also wenn ich diesen gedanklichen Schritt jetzt tue, egalisiert sich ja das ganze Thema sofort, weil dadurch würden wir ja die ganze DS-GVO aufweichen, wenn die Speichermechanik der Daten ausschlaggebend wäre über die Anwendbarkeit der DS-GVO.

**LM:** Weswegen sie ja auch vermutlich dann so technologieoffen und -neutral formuliert ist.

**MH:** Ja, das ist glaube ich nicht unabsichtlich, dass es irrelevant ist, die Technologie, die da drunter liegt. Und ich kann mir auch vorstellen, dass es bei den Open Source-Modellen auch noch interessant werden wird, denn auch ein Llama-Modell [Anm.: LLM-KI der Firma Meta] ist ja von einem Konzern trainiert worden und dann veröffentlicht auch noch unter Apache 2-Lizenz. Und da sage ich mal, dass ist auch eine ganz interessante Frage, ob ich diese ganzen Daten, die mir nicht gehören, die definitiv selbst nicht unter Apache 2-Lizenz lizenziert waren—ob ich die jetzt nur in anderer Darreichungsform unter Apache 2 im Internet veröffentlichen kann?

**LM:** Ja, da gab es erst diese LibGen, also diesen wissenschaftlichen Datensatz, der wirklich riesig ist, der vermutlich auch in dem Modell gelandet ist.

**MH:** Ja, stimmt. Die Thematik zum Beispiel, was vielleicht ganz interessant ist, es gibt ja diesen GitHub Co-Pilot, der mir beim Software-Entwickeln hilft. Microsoft stellt Nutzer dieses Modells in unbegrenzter Höhe von Rechtsschäden frei durch IP, also Intellectual Property, Violation, wenn der Co-Pilot quasi Source Code generiert, der IP von anderen Leuten darstellt, übernimmt Microsoft hierfür die Haftung, weil Microsoft dieses Thema ganz genauso sieht.

**LM:** Ja, interessant. Vielleicht noch gegen Ende, das ist natürlich sehr schwer vorherzusehen immer, aber ist es denn absehbar, dass vielleicht eine Technik kommt, mit der man dann, sagen wir, Impräzision aus den Modellen wegkriegt? Also das war jetzt in dem konkreten Fall Teil des Problems, glaube ich, nicht unbedingt. Also eigentlich schon, weil das Datum, um das es ging, bei der Person des öffentlichen Lebens war einfach falsch. Es gibt ja nicht nur das Recht auf Löschung, sondern auch auf Korrektur, dass man da quasi im Nachhinein so eine Lookup-Tabelle noch anlegt und dann entweder Korrekturen einpflegt oder gleich von Anfang an quasi das richtige Datum erwischt.

**MH:** Ja, ich glaube auch, das Modell generiert ja durch statistische Wahrscheinlichkeiten, das tut sich mit Zahlen ja ganz schwer. Der klassische Case ist immer, "wie viele Rs sind in Strawberry" – das kann das Modell auch nicht wirklich. Und es wäre möglich, das Modell quasi nachträglich zu finetunen, dass ich dem Modell sage, immer wenn jemand nach Person X Geburtsdatum fragt, dann bitte produziere dieses Geburtsdatum. Das wäre extrem aufwendig und kostspielig für OpenAI, für alle Anbieter, das zu tun. Und ich glaube auch, dass das die Qualität des Gesamtmodells deutlich senken würde, wenn ich hier 8000 Pflaster kleben muss. Das geht aber. Oder eine Lookup-Tabelle. Also das ist technisch möglich, das ist halt nur teuer und ineffizient, zumindest im Moment. Das Thema des Löschens, da bin ich vorsichtig. Es gibt im Moment technische Möglichkeiten, mit denen wir KI, also es kommt so ein bisschen auch aus der, ich sag mal, der Performance-Steigerung, das heißt Activation Aware Quantization. Wenn ich so ein Modell komprimieren möchte, sind manche Weights wichtiger als andere für die Funktionalität, für meinen Usecase. Das heißt, die wichtigen lasse ich in ihrer richtigen Präzision und die anderen verkleinere ich. Ich könnte mir vorstellen, dass ich durch Analyse des Modells während der Laufzeit herausfinden kann, wo oder welche Neuronen konkret aktiv sind, um dieses Geburtsdatum zu produzieren oder diesen personenbezogenen Datensatz zu produzieren. Und ich könnte mir vorstellen, diese aktiv zu löschen oder zu verändern, so dass das nicht mehr funktioniert, dieses Datum oder diesen Personennamen herzustellen. Das Problem ist, an der Stelle, das zerstört mir das ganze Modell quasi, weil das ja sukzessive aufeinander aufbaut. Das heißt, ja, es geht dann nicht mehr. Ich habe den Datensatz gelöscht, aber das Modell funktioniert auch nicht mehr. Das wäre im Moment meine einzige Idee, wie man das angehen könnte. Und das ist natürlich keine gute Idee. Und ich stelle mir das extrem schwer vor, da eine schnelle Lösung zu finden.

**LM:** Vor allem, weil es ja dann auch noch unvorhersehbare Auswirkungen hat auf ganz andere Queries vielleicht.

**MH:** Genau. Also, da glaube ich einfach, da haben wir im Moment viel zu wenig Explainability der Modelle, dass wir konkret hingehen könnten und etwas Spezifisches an einem fertigen Modell verändern können, an dieser Blackbox. Und so lang sich das, glaube ich, auch nicht ändert, während

wir da in dem Thema keinen Fortschritt sehen, dass wir gezielt etwas ändern können, wenn wir im Moment noch gar nicht wirklich verstehen und nachvollziehen können, wie es zu dieser Ausgabe kommt.

**LM:** Ja, super. Also, das wäre es soweit mit meinen Fragen.

**MH:** Wunderbar.

**LM:** Dann bedanke ich mich vielmals.

**MH:** Sehr gerne. Falls es irgendwas gibt, mal was lesen, mal einen Kommentar oder zu schreiben, immer gerne. Wir sind bei uns im Team da sehr research-offen. Also, falls es da mal eine zweite Meinung oder sowas braucht, können wir da dann mal nochmal drüber sprechen oder was gucken.

# Bibliography


Albrecht, J. P. (2016). "How the GDPR Will Change the World". In: *European Data Protection Law Review* 2.3, pp. 287–289. DOI: 10.21552/EDPL/2016/3/4.

Bittner, M. (1990). "The IRAC Method of Case Study Analysis: A Legal Model for the Social Studies". In: *The Social Studies* 81.5, pp. 227–230. DOI: 10.1080/00377996.1990.9957530.

Cory, T. et al. (2025). "Word-level Annotation of GDPR Transparency Compliance in Privacy Policies using Large Language Models (Version 1)". In: *arXiv*. DOI: 10.48550/ARXIV.2503.10727.

El Naqa, I. and Murphy, M. J. (2015). "What Is Machine Learning?" In: *Machine Learning in Radiation Oncology: Theory and Applications*. Springer International Publishing, pp. 3–11. DOI: 10.1007/978-3-319-18305-3_1.

European Data Protection Board (2019). *The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC*. Accessed: 2025-03-29. URL: https://www.edpb.europa.eu/news/national-news/2019/cnils-restricted-committee-imposes-financial-penalty-50-million-euros_en.

— (2024). *Report of the work undertaken by the ChatGPT Taskforce (No. 23 May 2024)*. Accessed: 2025-02-19. URL: https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf.

Feretzakis, G. and Verykios, V. (2024). "Trustworthy AI: Securing Sensitive Data in Large Language Models". In: *AI* 5.4, pp. 2773–2800. DOI: 10.3390/ai5040134.

Feretzakis, G. et al. (2025). "GDPR and Large Language Models: Technical and Legal Obstacles". In: *Future Internet* 17.4, p. 151. DOI: 10.3390/fi17040151.

Hupfauer, M. (Apr. 2025). *Interview with Markus Hupfauer on GDPR, LLMs, and the* noyb *complaint against OpenAI in Austria*. Personal interview. See Appendix for transcript.

Kamarinou, D., Millard, C., and Singh, J. (2016). *Machine Learning with Personal Data*. URL: https://papers.ssrn.com/abstract=2865811.

Laurelli, M. (2024). "Brain Surgery: Ensuring GDPR Compliance in Large Language Models via Concept Erasure (Version 1)". In: *arXiv*. DOI: 10.48550/ARXIV.2409.14603.

Liu, Y. et al. (2025). “Understanding LLMs: A comprehensive overview from training to inference”. In: *Neurocomputing* 620, p. 129190. DOI: 10.1016/j.neucom.2024.129190.

Miranda, M. et al. (2025). “Preserving Privacy in Large Language Models: A Survey on Current Threats and Solutions”. In: DOI: 10.48550/ARXIV.2408.05212.

Mittal, S. et al. (2024). “Democratizing GDPR Compliance: AI-Driven Privacy Policy Interpretation”. In: *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, pp. 735–743. DOI: 10.1145/3675888.3676142.

Muhammad, S. et al. (2024). “Safeguarding Data Privacy: Enhancing Cybersecurity Measures for Protecting Personal Data in the United States”. In: *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence* 15.1.

*noyb* (2024). *Complaint against OpenAI (Redacted)*. Accessed: 2025-04-17. URL: https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf.

Paal, B. (Feb. 2025). *Interview with Prof. Boris Paal on GDPR, LLMs, and the* noyb *complaint against OpenAI in Austria*. Personal interview. See Appendix for transcript.

PLT, iracmethod (2024). *What is The IRAC Method—Everything You Need to Know*. Accessed: 2025-03-02. URL: https://www.iracmethod.com/irac-methodology.

Rodriguez, D. et al. (2024). “Large language models: A new approach for privacy policy analysis at scale”. In: *Computing* 106.12, p. 7993. DOI: 10.1007/s00607-024-01331-9.

Singh, S. U. and Namin, A. S. (2025). “A survey on chatbots and large language models: Testing and evaluation techniques”. In: *Natural Language Processing Journal* 10, p. 100128. DOI: 10.1016/j.nlp.2025.100128.

Vaswani, A. et al. (2023). “Attention Is All You Need (No. arXiv:1706.03762)”. In: *arXiv*. DOI: 10.48550/arXiv.1706.03762.

Voigt, P. and von dem Bussche, A. (2024). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Nature Switzerland. DOI: 10.1007/978-3-031-62328-8.

Yan, B. et al. (2024). “On Protecting the Data Privacy of Large Language Models (LLMs): A Survey (Version 2)”. In: *arXiv*. DOI: 10.48550/ARXIV.2403.05156.

# Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For comparing my work with existing sources, I agree that it shall be entered in a database where it shall also remain after examination to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This work was not previously presented to another examination board and has yet to be published.

Gräfelfing, April 22, 2025

Place, Date

Signature