ChatGPT From a Data Protection Perspective

Kanan Naghiyev[*]

Abstract

This paper explores the complex legal issues related to the use of ChatGPT, a cutting-edge language model developed by OpenAI. The main legal concerns are related to data privacy, intellectual property rights, and the possible misuse of generated content. It emphasises the significance of these concerns in the present digital era, where data is frequently referred to as the "new oil". The paper highlights the reality of these issues, considering the extensive use of AI models such as ChatGPT in various sectors, ranging from customer service to content creation. With a particular focus on compliance with the General Data Protection Regulation (GDPR), the study delves deeply into the intricacies of how ChatGPT manages user data. This includes data retention policies, user consent mechanisms and data subject rights. The article explores the impact of these practices on users and offers insights into the wider data protection landscape in the field of artificial intelligence. This research is a significant resource for researchers, policymakers, and users wishing to understand the intersection of data protection and AI.

Annotasiya

Bu məqalə "OpenAI" tərəfindən hazırlanmış qabaqcıl dil (süni intellekt) modeli olan "ChatGPT"-nin istifadəsi ilə bağlı mürəkkəb hüquqi problemləri araşdırır. Əsas hüquqi problemlər məlumatların məxfiliyi, əqli mülkiyyət hüquqları və yaradılan məzmunun potensial olaraq sui-istifadəsi ilə bağlıdır. Məqalə məlumatların tez-tez "yeni neft" adlandırıldığı rəqəmsal dövrdə bu problemlərin əhəmiyyətini vurğulayır. Müştəri xidmətlərindən məzmunun yaradılmasına qədər müxtəlif sektorlarda "ChatGPT" kimi süni intellekt modellərinin geniş istifadə edildiyi bu dövrdə sözügedən problemlər təhlil edilir. Avropa İttifaqının Ümumi Məlumatların Qorunması Qaydalarına (General Data Protection Regulation) uyğunluğa xüsusi diqqət yetirməklə tədqiqat "ChatGPT"-nin istifadəçi məlumatlarını necə idarə etməsinin incəliklərini dərindən araşdırır. Buraya məlumatların saxlanması siyasətləri, istifadəçi razılığı mexanizmləri və məlumat subyektinin hüquqları daxildir. Məqalə bu təcrübələrin istifadəçilərə təsirini araşdırır və süni intellekt sahəsində daha geniş məlumatların mühafizəsi sahəsinə dair fikirlər təklif edir. Bu tədqiqat mövcud fərdi məlumatların qorunması qanunvericiliyi ilə süni intellektin kəsişməsini anlamaq istəyən istifadəçilər üçün əhəmiyyətli bir mənbədir.

CONTENTS

[*] LL.M. Candidate in Intellectual Property Law, Technical University of Dresden.

Introduction

There is no single consensus definition for Artificial Intelligence (AI), but its capabilities are undeniable and continue to expand with advancements in technology. It is a set of technologies that combine data, algorithms, and computing capacity to create computer systems capable of performing tasks that traditionally required human intelligence. This field of computer science has been greatly influenced by the growing availability of data and the advancement of computational processing capacity in recent years.[1]

AI has a rich history dating back to the mid-20th century, with the term gaining prominence in the 2010s. The concept was first introduced in Alan Turing's seminal 1950 paper.[2] In this paper, Turing proposed a heuristic test, known as the Turing Test, to determine whether something is intelligent or not. The purpose of this test is to answer the question, "Can machines think?" The test involves a computer, a human, and a human judge. Both the computer and the human attempt to convince the judge that they are humans by providing typewritten answers to his questions. The computer is said to have passed the test if the judge cannot consistently distinguish between the two.[3] After over fifty years of technological development and the resulting advances in algorithmic and computational performance, along with the availability of large datasets, AI now plays a significant role in nearly all aspects of life.

This study specifically explores ChatGPT, a variant of the Generative Pre-trained Transformer (GPT) models developed by OpenAI. The generative nature of ChatGPT enables it to create new content, while its transformer architecture allows the model to focus on different parts of the input, enabling it to generate detailed responses to prompts and follow-up questions.[4]

[1] Carlos Ragazzo, Morgana Tolentino and Bruna Cataldo, Inteligência artificial: o que é e como se aplica às finanças? (Artificial Intelligence: What is It and How Does Finance Use It?), 4 (2023). Available at: http://dx.doi.org/10.2139/ssrn.4579348 (last visited Dec. 18, 2023).

[2] Alan M. Turing, *Computing Machinery and Intelligence*, 236 Mind 433, 433-460 (1950).

[3] *Id.*, 434.

[4] Gero Strobel, Leonardo Banh, Frederik Möller and Thorsten Schoormann, Exploring Generative Artificial Intelligence: A Taxonomy and Types, in *Proceedings of the 57th Hawaii International Conference on System Sciences* 4546, 4546-50 (2024).

Moreover, the model undergoes a rigorous two-step training process: pre-training and fine-tuning. During the pre-training phase, the model acquires a comprehensive understanding of the statistical patterns of the language by training on a vast corpus of text data. The subsequent fine-tuning phase involves additional training on a more specific dataset with human feedback to adapt to particular tasks.[5] For example, a model pre-trained on 14 million images of flowers can be fine-tuned to recognize only five types: rose, tulip, daisy, sunflower, and dandelion.[6]

Given ChatGPT's prevalence and widespread usage, it is important to consider the legal concerns over how it handles the personal data input by users. For example, a recent survey among chatbot users found that 87.8% of respondents believe that chatbots can be used to collect personal information or manipulate users.[7] The fact that a large percentage of users believe ChatGPT can collect personal information suggests a need for clear communication about data handling and privacy policies. The concern about manipulation could stem from the potential misuse of persuasive techniques by AI. This underlines the importance of ethical guidelines in AI development and usage.

ChatGPT must comply with various data protection laws which safeguard individuals' rights regarding their personal data. The protection of personal data is essential for the right to privacy. Article 8 of the EU Charter of Fundamental Rights states that everyone has the right to the protection of their personal data, which must be processed fairly for specific purposes and based on the consent of the person concerned or some other legitimate basis laid down by law. Additionally, everyone has the right to access data collected about them and to have it rectified.[8] Article 16 (1) of the Treaty on the Functioning of the European Union (TFEU) provides that everyone has the right to the protection of their personal data.

The General Data Protection Regulation (hereinafter GDPR) is based on the aforementioned legal acts and plays a crucial role in regulating the protection of personal data within the EU.[9] Consequently, this paper will examine ChatGPT's compliance with the GDPR.

[5] Xiaoming Zhai, ChatGPT User Experience: Implications for Education, 2 (2022). Available at: http://dx.doi.org/10.2139/ssrn.4312418 (last visited Jan. 4, 2024).

[6] Kenneth Ward Church, Zeyu Chen and Yanjun Ma, *Emerging trends: A Gentle Introduction to Fine-tuning,* 27 Natural Language Engineering 763, 769 (2021).

[7] Glorin Sebastian, *Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk? - An Exploratory Study*, 15 International Journal of Security and Privacy in Pervasive Computing 1, 4 (2023). Available at: http://dx.doi.org/10.2139/ssrn.4363843 (last visited Feb. 9, 2024).

[8] European Parliament, Charter of Fundamental Rights of the European Union, art. 8 (1)–8 (2) (2000).

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on

Firstly, it should be established whether the information collected by OpenAI qualifies as personal data[10] under the GDPR as the regulation only applies to data processing operations concerning personal data. According to Article 4 (1) of the GDPR, personal data is any information relating to an identified or identifiable natural person. OpenAI expressly states that it processes various personal data for ChatGPT, including account details, communication information (names, contacts, message content), and users' contact information from social media. The processed information can be used to easily identify the user. Thus, it falls under the scope of the GDPR.

One of the main challenges ChatGPT faces in complying with the GDPR is the right to be forgotten. ChatGPT's compliance with the right to be forgotten may be problematic if removing personal data compromises its performance or is technically impractical, and legal or contractual obligations may necessitate retaining some personal data. Additionally, GDPR's transparency and accountability requirements are difficult for ChatGPT due to its complex data processing and inability to fully explain its decision-making process.

This paper provides an overview of ChatGPT's handling of personal data and the challenges faced in complying with the current data protection laws, especially within European Union legislation. The first chapter of the paper examines OpenAI's privacy policy, with a specific focus on how it applies to ChatGPT. The article's second chapter discusses how ChatGPT adheres to GDPR's key principles. It covers two main perspectives: the rights of users and the rights of third parties. The third chapter discusses potential measures that OpenAI could implement to enhance GDPR compliance.

I. ChatGPT's Privacy Policy and Data Processing

As the developer of ChatGPT, OpenAI determines the purposes of the processing of personal data and is therefore the controller. To better understand OpenAI's role, it is essential to note that a controller defines the processing purposes and means of personal data, while a processor processes this data on the controller's behalf.[11] The controller and the processor could be the same subject or vice versa.

OpenAI's Privacy Policy (hereinafter Privacy Policy) covers the collection of personal information, how personal information is processed, how personal information is disclosed, and the rights provided by law. In the privacy policy of OpenAI, the security of the privacy of the users is emphasised, rather than the privacy of third parties: "*We at OpenAI OpCo, LLC respect your privacy and are strongly committed to keeping secure any information*

the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), recital 1 (2016).

[10] Personal data is any information relating to an identified or ascertainable individual, for example name, email address, location and so on.

[11] *Supra* note 9, art. 4 (7)–4 (8).

we obtain from you or about you. (...) This Privacy Policy does not apply to content that we process on behalf of customers of our business offerings, such as our API".[12] OpenAI's data processing activities with ChatGPT fall within the scope of the GDPR. According to the GDPR, online identifiers, including user inputs to ChatGPT, are considered personal data. Even if these inputs do not contain personally identifiable information, they may still be classified as personal data if they can be linked to the user during interaction.[13] Additionally, the processing of this data, which includes a wide range of activities regarding personal data, is governed by the GDPR as well.[14] For instance, when a user asks ChatGPT about the weather, the system processes this input, understands the user's request and generates a response about the current weather conditions. This constitutes a type of data processing that falls under the GDPR.

However, views on ChatGPT's GDPR compliance vary. According to Kesa and Kerikmäe, it might be impossible to ensure compliance with GDPR during data processing when AI is used. They argue that complete compliance with GDPR's requirements to ensure certain rights is not possible for data processors using complex machine-learning systems that process vast amounts of data through multistep instructions.[15]

In a similar manner, the Scientific Foresight Unit of the European Parliamentary Research Service conducted an analysis which concluded that although the GDPR may be compatible with the development of AI, it "*does not provide sufficient guidance for controllers*". Therefore, it may need to be "*expanded and concretised*" to offer clearer explanations of how its provisions apply to AI systems.[16]

Moreover, the broad characteristics of general artificial intelligence, or the "*foundation models*" as described in the proposed artificial intelligence legislation,[17] present specific challenges in aligning with GDPR requirements. Some EU regulators have proposed changes to the EU AI Act to classify AI

[12] Privacy Policy (2023), https://openai.com/policies/privacy-policy (last visited Dec. 13, 2023); Here, the term "API" refers to a tool released by OpenAI for the development of conversational AI using natural language processing (NLP). With the API, users can create an application (such as GPT) or a plugin for the current ChatGPT model.

[13] *Supra* note 9, recital 30.

[14] *Id.*, art. 4 (2).

[15] Aleksandr Kesa & Tanel Kerikmäe, *Artificial Intelligence and the GDPR: Inevitable Nemeses?*, 10 TalTech Journal of European Studies 68, 70–71 (2020).

[16] European Parliamentary Research Service, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, 3 (2020). Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf (last visited Feb. 4, 2024).

[17] European Parliament, General-purpose artificial intelligence, 1-2 (2023). Available at: https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA%282023%29745708_EN.pdf (last visited Feb. 4, 2024).

systems, such as ChatGPT, that generate complex texts without any human oversight, as "*high-risk*" AI systems that would fall under the requirements of the Act. This classification has been controversial with certain regulators arguing that technologies such as ChatGPT, which only generate text, are not risky at all.[18]

The provisions of the GDPR have the potential to be either a barrier to or a facilitator of the development of AI systems such as GPTs. The GDPR encourages the use of privacy-enhancing technologies, such as encryption, which can help protect personal data while still allowing it to be used in AI systems.[19]

This chapter explores how the ongoing development of Generative AI technologies affects cybersecurity tools and threats, which in turn impacts the effectiveness of the GDPR and privacy policies. The focus is on investigating how ChatGPT may hinder the enforcement and compliance of the GDPR. For the purpose of analyzing the Privacy Policy of OpenAI in the framework of the GDPR, it is better to divide it into three main categories: data collection and storage, use of data, and data disclosure.

A. Data Collection and Storage

This section aims to describe the types of data collected and stored by ChatGPT. Moreover, it evaluates ChatGPT's security protocols and practices for handling sensitive data, with a specific focus on scenarios involving confidential business information.

In the context of the GDPR, data collection refers to collecting or obtaining personal data from individuals or data subjects. OpenAI collects personal information when a user creates an account to use their services or communicate with them. This includes account information,[20] user content,[21] communication information, and social media information.

ChatGPT stores all user input information, including confidential information a user may enter.[22] In a business context, ChatGPT could be used to generate an executive summary of a company's quarterly performance. For instance, employees might input sensitive and confidential actual sales data into ChatGPT, which then processes the data to produce the desired text. However, there are significant concerns about the handling of such sensitive data. ChatGPT can still use the data to improve its AI even if the employee

[18] Josephine Wolff, William Lehra and Christopher S. Yoo, Lessons from GDPR for AI Policymaking, 1 (2023). Available at: http://dx.doi.org/10.2139/ssrn.4528698 (last visited Feb. 4, 2024).

[19] *Supra* note 9, art. 32-33.

[20] The information relating to the account created by the user, including his/her name, contact information, account details, payment card information, transaction history.

[21] The input or prompt, file uploads or feedback information generated during the use of OpenAI services.

[22] *Supra* note 12, section 1.

deletes their conversations. This poses a risk if users enter sensitive personal or company information that would be attractive to malicious parties in the event of a breach.

In addition to the information collected with user consent, ChatGPT automatically gathers technical information each time a user visits, uses or interacts with the service. Technical information includes data logs, usage data, device information, cookies and analytics.

Moreover, the offering of API-based services for businesses emphasizes data elements that play a role in the thorough assessment of ChatGPT's privacy and security measures. OpenAI offers API-based services to businesses using the same AI technologies as those available to individuals. An API, (Application Programming Interface), enables two software programs to communicate with each other by exposing certain functions and data from one application to the other. For instance, with the OpenAI API, developers can integrate ChatGPT's language understanding and generation capabilities into their own software, allowing them to enhance their applications with advanced natural language processing features. These services are covered by the Enterprise Privacy policy.[23]

OpenAI stores user data on its systems and sub-systems, including international systems. Users can opt out of the use of this data to improve OpenAI's services through a form provided. However, OpenAI still collects and stores this data.[24] OpenAI states that this form addresses the main concern of GDPR compliance, which is the collection of personal data for training purposes. However, if ChatGPT is asked about a previous user, there is a possibility that it could unintentionally provide other users with personal data that has been processed for training purposes.[25] This could lead to significant losses in a business context. For example, a user could inadvertently provide their customer's contact details to OpenAI when using ChatGPT to collect feedback and compile it into a report.

Thus, complying with GDPR demands more than simply offering the option to opt-out. It requires OpenAI to ensure compliance with the principles of data minimisation, purpose and limited retention because the data is still being collected and stored. For instance, when using ChatGPT to gather feedback and compile a report, a user's client contact information must be

[23] Enterprise Privacy at OpenAI (2023), https://openai.com/enterprise-privacy (last visited Jan. 26, 2024).

[24] How your data is used to improve model performance (2023), https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance (last visited Jun. 7, 2023).

[25] Glorin Sebastian, *Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information*, 15 International Journal of Security and Privacy in Pervasive Computing 1, 6 (2023). Available at: http://dx.doi.org/10.2139/ssrn.4454761 (last visited Feb. 8, 2024).

handled securely and kept confidential. It should not be shared with other users or used for training purposes.

In conclusion, this analysis reveals insights into data collection, storage, and security measures regarding ChatGPT, especially in relation to confidential business information. It underscores the need for safeguards and compliance with the GDPR to protect data privacy.

B. Use of Data

This section explores the various purposes identified by OpenAI for collecting personal information. Moreover, it touches upon OpenAI's practices aimed at ensuring data privacy and emphasizes the user's responsibility to follow the GDPR regulations. The text also notes OpenAI's challenge of aligning with the purpose limitation principle.

OpenAI has identified several purposes for the use of personal information collected. These purposes include (a) providing, administering, maintaining and analysing the AI services; (b) conducting research to improve the AI; (c) communicating with the user; (d) developing new programmes and services; (e) preventing fraud, criminal activity and misuse of the OpenAI services, ensuring the security of IT systems, architecture and networks; (f) complying with legal obligations and legal processes and protecting the rights, privacy, safety or property of OpenAI's affiliates, users and third parties.[26]

Furthermore, OpenAI emphasises its commitment to data privacy within the specified purposes by outlining practices to safeguard data. For example, in the Privacy Policy, OpenAI highlights the possibility of aggregating or de-identifying[27] personal data to achieve this goal. These practices serve to analyze service effectiveness, improve or add features, and conduct research for other similar purposes.

Considering OpenAI's stated purposes for using personal information and its dedication to data privacy, users who interact with the chat interface are responsible for fulfilling certain obligations. These include providing privacy notices and obtaining consent, as detailed in OpenAI's policies. If a user wants to enter personal data into the chat, they will need to provide the people involved with an appropriate privacy notice. In addition, users will need to obtain consent from those individuals and demonstrate to OpenAI that it is processing that data in a lawful manner.[28] There is no specific supervising mechanism for this procedure. However, if individuals in the EU encounter

[26] *Supra* note 12, section 2.

[27] "Aggregating" or "de-identifying" personal data refers to the process of removing or altering information that could be used to identify individuals, thereby protecting their privacy. *See* California Consumer Privacy Act (2018). Available at: https://www.oag.ca.gov/privacy/ccpa (last visited Dec. 21, 2023).

[28] Terms of Use (2023), https://openai.com/policies/terms-of-use (last visited Jun. 28, 2023).

any privacy issues during the consent procedure, they can file a complaint with national data protection authorities in accordance with the GDPR.[29]

Furthermore, if the user intends to use the information entered, which is defined as personal data under the GDPR, for processing, they will need to contact OpenAI to execute the Data Processing Addendum (DPA). OpenAI's Data Processing Addendum governs the processing of Customer Data. These include data provided by the customer through OpenAI's API or any OpenAI services for businesses (API Services), or data provided pursuant to OpenAI's provision of the ChatGPT Enterprise service for businesses (ChatGPT Enterprise Services). The DPA applies to all Services.[30] The DPA is incorporated into the agreement between the customer and OpenAI that governs the customer's use of the Services. This could be the OpenAI Business Terms, an Enterprise Agreement, or another individual agreement.

To guarantee the smooth integration of OpenAI's privacy policies and obligations with the regulatory framework, it is crucial to comprehend GDPR's purpose limitation principle. This principle limits the use of personal data to the purposes for which it was collected.[31] To be in line with it, OpenAI must be clear about the purposes of its processing from the outset. Personal data should only be used for new uses that are compatible with the original uses of OpenAI, where consent has been obtained, or where there are explicit legal obligations or roles.

Overall, when it comes to the use of personal data, OpenAI specifies certain categories of purposes that are limited by the Privacy Policy. However, it is problematic to define whether ChatGPT adheres to the principle of purpose limitation or not, as the Privacy Policy is subject to constant changes by OpenAI. This is especially the case when OpenAI publicly states that they may use personal data to improve their models.

C. Data Disclosure

This section examines OpenAI's data disclosure practices by outlining justifications for the disclosure. It also focuses on OpenAI's challenge to comply with the GDPR in the context of disclosure.

According to the Privacy Policy, personal information may be transferred or disclosed to third parties without further notice to users, except as required by law. Among these third parties, vendors and service providers such as providers of hosting, cloud and other IT services, email communication software and newsletter services, and web analytics services are at the

[29] *Supra* note 8, art. 8 (3).

[30] Data processing addendum (2023), https://openai.com/policies/data-processing-addendum (last visited Dec. 21, 2023).

[31] Principle (b): Purpose limitation, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/the-principles/purpose-limitation/ (last visited Dec. 25, 2023).

forefront.[32] Apart from the Vendors and Service Providers, in the event of a reorganisation, bankruptcy, receivership or transfer of the Service (in this case ChatGPT) to another provider, personal data may be disclosed or transferred to a successor or affiliated company.

Where required by law, OpenAI may disclose personal data to law enforcement authorities to comply with legal obligations and to protect the personal safety of users or the public.[33] This statement is problematic because it implies the disclosure of personal data of any user worldwide, including EU citizens, to US law enforcement agencies on the basis of US law enforcement legislation.

Moreover, OpenAI may disclose personal information to its affiliates, which means any entity that controls, is controlled by or is under common control with OpenAI. These affiliates may only use the information in ways consistent with the Privacy Policy.

Any disclosure of data constitutes data processing and is therefore subject to the GDPR,[34] requiring adherence to the principles of fairness, lawfulness, and transparency, as outlined in Articles 12, 13 and 14. Therefore, it is necessary to closely examine whether OpenAI's data disclosure practices align with these GDPR requirements.

II. ChatGPT's Compliance with the GDPR

As previously discussed in the Introduction, the personal data processed by OpenAI falls under the scope of "personal data" under the GDPR. As OpenAI's services are accessible to EU citizens, it must comply with the GDPR.[35]

OpenAI's privacy policy regarding ChatGPT details the company's efforts to comply with the California Consumer Privacy Act (hereinafter CCPA). However, it lacks details concerning its adherence to international laws, including the GDPR. The two regulations overlap when it comes to some rights and compliance with the GDPR often means that a company is well on their way to meeting CCPA requirements. Under both the CCPA and the GDPR, businesses (controllers) need to be transparent about what personal data they collect and what they do with that data.[36] Individuals are entitled to access their personal data and can request copies of their personal information verbally or in writing in accordance with both legal acts.[37] Both acts confirm that under certain circumstances, individuals have a right to opt out of having

[32] *Supra* note 12, section 2.

[33] *Id*., section 4.

[34] His Majesty's Revenue and Customs, Information Disclosure Guide, https://www.gov.uk/hmrc-internal-manuals/information-disclosure-guide/idg40160 (last visited Dec. 25, 2023).

[35] *Supra* note 9, art. 3 (2) (a).

[36] California Civil Code, § 1798.100 (2018).

[37] *Ibid*.

their personal data processed by an organization.[38] When it comes to the right to erasure, under CCPA, an organisation is required to delete information that it obtained directly from the consumer.[39] If this data is obtained from other sources, it falls outside the scope of the right to be forgotten within CCPA. On the other hand, the GDPR extends to data collected by the organization from the consumer directly or data regarding a consumer that they acquired indirectly. As seen, the requirements outlined in the GDPR are more extensive and demanding.

Although the CCPA and the GDPR share similarities, OpenAI has faced regulatory challenges in the EU, specifically in Italy, despite its efforts to comply with these regulations. These challenges underscore the difficulties of navigating various data protection laws and emphasize the significance of transparency and age verification in AI services.

According to an order issued by the Italian Data Protection Authority (hereinafter GPDP) on 31st March 2023, OpenAI lacked a legal justification for collecting users' personal data. Specifically, the order requires more transparency from OpenAI, which has been, and probably still is, secretive about how training data and user instructions are processed in the development of ChatGPT. In addition, OpenAI did not have a mechanism to prevent underage users from accessing the services, exposing them to completely inappropriate answers compared to their level of development and self-awareness.[40] This order by the GPDP had a similar legal basis to the order banning Replika AI. There, the GPDP emphasised the lack of an age verification mechanism, in breach of the GDPR in both orders.[41] This is because the processing of personal data cannot be based on an agreement between a controller and a person who lacks the legal capacity to enter into a contract. The GPDP gave OpenAI a deadline to address the issues and make changes to its privacy policy. OpenAI temporarily geo-blocked ChatGPT in Italy to conform with the order.

As a result, OpenAI has created a new form for users from the EU with the aim of removing personal data under the GDPR.[42] The form requires individuals to provide evidence that they are included in responses to

[38] *Id.*, § 1798.120.

[39] *Id.*, § 1798.105.

[40] Artificial intelligence: stop to ChatGPT by the Italian SA personal data is collected unlawfully, no age verification system is in place for children (2023), https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english (last visited Dec. 13, 2023).

[41] Italy Bans AI Chatbot Replika from Using Personal Data (2023), https://www.pcmag.com/news/italy-bans-ai-chatbot-replika-from-processing-user-data#:~:text=San%20Francisco%2Dbased%20AI%20chatbot,minors%20and%20emotionally%20vulnerable%20people (last visited May 24, 2023).

[42] OpenAI Personal Data Removal Request (2023), https://share.hsforms.com/1UPy6xqxZSEqTrGDh4ywo_g4sk30 (last visited Dec. 13, 2023).

prompts that they provide, so screenshots and prompts must be provided. In response to concerns about minors, ChatGPT users were required to provide their age during the registration procedure.[43] Following these changes to the Privacy Policy, the GPDP lifted the ban on ChatGPT on 28th April 2023.

Although the terms of service for ChatGPT have been updated to require users to be at least 13 years old to access the service, the problem still remains. It has been noted that OpenAI does not verify users' ages, which could potentially allow users under 13 to sign up for the service.[44] This has raised concerns about the collection and processing of data from minors, as discussed previously. According to Article 8 of the GDPR, the processing of the personal data of a child under the age of 16 can only be lawful with the consent given or authorised by the holder of parental responsibility over the child. Any Member State may set a lower age for processing, provided that such a lower age does not fall below 13. Consequently, questions about legal problems about ChatGPT's compliance with the GDPR remain unresolved.

Investigations into ChatGPT within EU countries are not limited to Italy. In April, the Spanish Data Protection Agency (hereinafter AEPD) said it had launched its own initiative and asked the EU Data Protection Board (hereinafter EDPB) to assess the privacy concerns surrounding ChatGPT, as global scrutiny of AI systems intensifies.[45] Unlike GPDP, AEPD has not initiated blocking access to or prohibiting the use of ChatGPT yet. EDPB has launched a dedicated task force on ChatGPT in response to a request from the AEPD. The task force aims to promote cooperation and information exchange among data protection authorities regarding possible enforcement actions.[46] Its ultimate goal is to coordinate the GDPR enforcement on generative AI technology across the bloc. However, in the short term, early adopters of DPAs such as Italy and Spain may conclude their investigations and take enforcement action before the Board is able to provide any harmonising recommendations.[47]

The EDPB is currently working towards a unified approach to regulating AI technologies, such as ChatGPT, across the EU. However, it is important to note how individual companies, like OpenAI, are navigating these

[43] OpenAI reinstates service in Italy with enhanced transparency and rights for european users and non-users (2023), https://www.gpdp.it/home/docweb/-/docweb-display/docweb/9881490#english (last visited Feb. 12, 2024).

[44] Is ChatGPT safe for all ages? (2023), https://help.openai.com/en/articles/8313401-is-chatgpt-safe-for-all-ages (last visited Dec. 25, 2023).

[45] Spain Asks EU Data Protection Board to Discuss OpenAI's ChatGPT (2023), https://www.reuters.com/technology/spains-data-regulator-asks-eu-data-protection-committee-evaluate-chatgpt-issues-2023-04-11/ (last visited Jun. 28, 2023).

[46] EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT (2023), https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en (last visited Dec. 25, 2023).

[47] *Ibid*.

regulations. In particular, OpenAI's use of a subsidiary in Ireland as a data controller under the GDPR has implications for where and how user data is processed. According to the GDPR, establishments (companies) may imply the effective and real exercise of activity through stable arrangements, whether through a branch or a subsidiary with legal personality.[48] In the Privacy Policy, which is addressed to users in the EEA and Switzerland,[49] OpenAI has appointed its subsidiary company, OpenAI Ireland Limited, as the controller. This means that data processing will occur in the EU. However, since OpenAI's primary establishment for training AI models is located in the US, data transfer is inevitable. Data transfers between the EU and the US pose privacy risks that are challenging to mitigate. This issue is further discussed in the following chapter of the paper.

It is evident that, due to its non-compliance with the GDPR, ChatGPT has been found to have the potential to cause serious legal issues in Europe. As such, it is essential to identify specific areas where ChatGPT does not comply with GDPR requirements. Solutions to bring ChatGPT into compliance with GDPR requirements can be found through a thorough understanding of these issues.

A. Compliance with the Data Protection Rights of the Users

ChatGPT users, who can be anywhere in the world, are the primary concern regarding the processing of personal data by OpenAI. Once again, the processing of personal data of natural persons in the EU is subject to the GDPR. These individuals are accorded certain rights under the GDPR, including the right to rectification and the right to be forgotten, which must be upheld by the data controller. In addition to these rights, the GDPR outlines fundamental principles for data processing, such as lawfulness, purpose limitation, transparency, and data retention period. OpenAI's existing Privacy Policy is examined in this section in light of these crucial GDPR rights and principles.

1. Right To Rectification

Data subjects have the right to request the controllers to rectify inaccurate and/or incomplete personal data concerning them.[50] In the Privacy Policy, the right to correct or update personal information is listed under the rights of data subjects. However, OpenAI states that they may not be able to correct the

[48] *Supra* note 9, recital 22.

[49] Europe privacy policy (2023), https://openai.com/policies/eu-privacy-policy (last visited Dec. 25, 2023).

[50] *Supra* note 9, art. 16.

inaccuracy due to the technical complexity of ChatGPT.[51] They point out to users that if they cannot correct the data, the users can always request to remove the personal data which is discussed in the following body paragraph.

2. Right To Be Forgotten

One of the main issues is ChatGPT's compliance with the right to be forgotten. This right is stipulated in Article 17 of the GDPR and allows individuals to request the erasure of their personal data from an organisation's records under certain conditions. The concept of the right to be forgotten is based on the fundamental need of the individual to be able to determine the course of his life autonomously, without being permanently or periodically stigmatised as a result of a specific act committed in the past.[52] Especially when these events occurred many years ago and are not related to the current context.[53] That is why, if the data is no longer necessary for the purposes for which it was collected or processed, or if the individual withdraws consent or objects to the processing, the organisation must take reasonable steps. These include notifying other parties who have access to the data of the request and deleting any links, copies or replicas of the data.

The right to be forgotten pertains to the legitimacy of data processing as time passes. Processing operations that were once legitimate may become illegitimate. The importance of this right to EU lawmakers stems from the understanding that information and events from the past can have stigma and consequences for individuals even many years later.

For example, a Spanish citizen filed a complaint against Google Spain and Google Inc. after his name in Google's search engine led to newspaper pages about his past social debt recovery proceedings. He requested the removal or modification of these pages and the removal of his personal data from Google's search results. Google refused, but the courts ruled in favour of the citizen, stating that the search results were "*inadequate, irrelevant or no longer relevant or excessive*" since the debt had been paid long ago. This case highlights the importance of the right to be forgotten to prevent individuals from being indexed in search engines for past actions such as revenge porn, petty crimes, or unpaid debts.[54]

However, applying the right to be forgotten to ChatGPT is not straightforward as the model is trained on millions of pages of data scraped from the web. This data includes personal information about individuals who may not have given their consent or may not be aware of how their data is

[51] *Supra* note 12, section 4.

[52] Cécile de Terwangne, The Right to be Forgotten and the Informational Autonomy in the Digital Environment, 1 (2013).

[53] *Id.*, 6.

[54] Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, C-131/12, Judgment, para. 92 (2014).

being used.[55] Furthermore, deleting data from a complex AI system such as ChatGPT may not be technically feasible or effective, as the data may be embedded in the model's parameters or may continue to influence its outputs even after removal. In other words, once data is processed to train an AI, it is complex to get the AI to unlearn the processed data.[56]

For many standard machine learning models, complete removal of personal data would require re-training the entire model from scratch on the remaining data.[57] The research on data erasure in machine learning suggests that only the surface has been scratched in understanding the efficiency of erasure in learning systems. Moreover, that is done through a set of simplified assumptions on a single model and a single database. In contrast, ChatGPT is a further development of a class of machine learning models for natural language processing known as Large Language Models (LLMs). Therefore, deleting data from ChatGPT's training database is quite complex.

OpenAI has introduced a feature in ChatGPT to turn off the chat history, which contains logs and other data that might be used for AI training. Turning off the chat history will stop new data from being stored. The user's current data will be deleted in 30 days. However, this does not have any retroactive effect. Previously processed data for AI training purposes may not be completely deleted for technical reasons discussed in the previous body paragraph. Therefore, other solutions must be found to ensure the right to be forgotten, as this method alone is insufficient.

One method introduced in the literature to ensure the right to be forgotten is machine unlearning. Machine unlearning is an evolving area of artificial intelligence that makes a model forget or unlearn specific parts of its training data, essentially reversing the process of machine learning.[58]

For example, a machine learning model predicts movie recommendations based on user ratings. This model was trained on a dataset that includes ratings from a specific user. However, if that user decides to delete their account and all associated data from the platform, the platform must respect their privacy and remove their data not only from the database but also from the recommendation model. To avoid the computational expense of retraining the model from scratch without the user's data, machine unlearning techniques can be employed to make the model forget the user's data. This results in a model that provides recommendations as if it had never

[55] *Supra* note 24.

[56] Antonio Ginart, Melody Y. Guan, Gregory Valiant and James Zou, *Making AI Forget You: Data Deletion in Machine Learning*, Proceedings of the 33rd International Conference on Neural Information Processing Systems 3518, 3521-22 (2019).

[57] *Id*., 3527.

[58] Announcing the First Machine Unlearning Challenge (2023), https://blog.research.google/2023/06/announcing-first-machine-unlearning.html (last visited Dec. 25, 2023).

encountered user's data, thus respecting their decision to remove their data from the platform.

Machine unlearning techniques are being developed that could potentially help AI developers comply with the right to be forgotten under the GDPR. These techniques aim to adjust the already-trained model to remove the influence of the requested data, without needing to retrain the entire model from scratch.[59] While these techniques are promising, they are still in the early stages of development and face several challenges, including the need for a universally accepted gauge to determine the effectiveness of machine unlearning and the difficulty of ensuring complete data point forgetting for larger, more complex deep learning models.[60] Therefore, machine unlearning has the potential to assist OpenAI in complying with the right to be forgotten under the GDPR. However, it is a complex issue that requires further research and development.

Another technique introduced in AI development is prompt engineering. This involves asking the appropriate questions to obtain optimal output from an AI.[61] To ensure the right to be forgotten, prompts should be designed to avoid topics or questions that may lead the model to generate responses based on forgotten data.

For example: If you enter a prompt such as "Has person (A) visited (C) website?", ChatGPT will provide detailed information about it. However, through prompt engineering, the user can request adherence to the right to be forgotten under the GDPR, and ask to erase all data regarding person (A)'s visits to websites. Then, AI will understand it must not answer questions based on that erased information and decline to respond due to privacy considerations.[62] However, prompt engineering is not a guaranteed solution. Although it can assist in directing the model's responses, it cannot entirely prevent the model from generating responses based on its training data.[63]

Another technique that can be used to protect the right to be forgotten is differential privacy. Differential privacy is a mathematical framework that quantifies the privacy guarantees provided by an algorithm. It introduces randomness in the responses to queries to provide robust privacy

[59] Dawen Zhang et al., Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions, 11-12 (2023). Available at: https://doi.org/10.48550/arXiv.2307.03941 (last visited Dec. 25, 2023).

[60] Siva Sai et al., *Machine Un-learning: An Overview of Techniques, Applications, and Future Directions*, 16 Cognitive Computation 482, 483 (2024).

[61] Google Machine Prompt Engineering for Generative AI, https://developers.google.com/machine-learning/resources/prompt-eng (last visited Dec. 25, 2023).

[62] Zhang et al., *supra* note 59, 13.

[63] *Ibid*.

assurances.[64] This provides privacy while sharing information about a group of individuals. For example, consider an AI generative model trained on medical records to predict health results. The model's purpose is to generate synthetic patient profiles for research purposes. However, to safeguard the privacy of individuals in the training data, the model must be differentially private. By introducing a small amount of random noise[65] into its learning process, the model achieves differential privacy. This noise does not significantly affect the model's ability to generate realistic synthetic patient profiles. However, it ensures that the output cannot be traced back to any specific individual in the training data. Therefore, even if someone had access to all the synthetic patient profiles generated by the model, they would not be able to determine whether a specific individual's medical record was part of the training data.

Differential privacy is achieved by making small, arbitrary changes to individual data points that do not alter the relevant statistics.[66] In essence, an algorithm is considered differentially private if an observer who sees its output cannot determine whether the computation involves the information of a specific individual.

Differential privacy offers a quantified measure of privacy loss and an upper bound, enabling curators to choose the explicit trade-off between privacy and accuracy. The system is resilient against privacy attacks that are not yet known.[67] However, it promotes increased data sharing, which if not done properly, can increase privacy risks.

Implementing differential privacy in practice can be challenging and may not be sufficient on its own to ensure the right to be forgotten. Therefore, a combination of methods, including machine unlearning and prompt engineering, may be necessary.

It is therefore unclear whether ChatGPT can comply with the right to be forgotten, as set out in Article 17 of the GDPR. Extensive research and regulation will be required to address this issue and to ensure that individuals' data privacy rights are respected and protected.

3. The Lawfulness of Data Processing

[64] Michael B. Hawes, *Implementing Differential Privacy: Seven Lessons From the 2020 United States Census*, 2 Harvard Data Science Review 2, 6 (2020). Available at: https://doi.org/10.1162/99608f92.353c6f99 (last visited Feb. 4, 2024).

[65] Random noise refers to small random variations that are added during the learning process in order to ensure differential privacy. This allows the model to learn overall patterns and produce realistic outputs, while preventing the model from memorising individual data points, thereby protecting individual privacy.

[66] Joseph Ficek et al., *Differential Privacy in Health Research: A Scoping Review*, 28 Journal of the American Medical Informatics Association 2269, 2270 (2021). Available at: https://doi.org/10.1093/jamia/ocab135 (last visited Feb. 4, 2024).

[67] *Ibid*.

The lawfulness of data processing could be another issue in relation to GDPR compliance, as highlighted by the GPDP. The GDPR sets out six legal bases for processing personal data, which are:[68]

a) Consent – the data subject has given clear and informed consent for a specific purpose.
b) Contract – the processing is necessary for the performance of a contract to which the data subject is a party or to take steps at the request of the data subject prior to entering into a contract.
c) Legal obligation – the processing is necessary for compliance with a legal obligation to which the controller is subject.
d) Vital interests – the processing is necessary to protect the vital interests of the data subject or another natural person.
e) Public interest – the processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
f) Legitimate interests – the processing is necessary for the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject.

Depending on how an AI is used and by whom, different legal bases may apply. For example, if an AI is used by a company to provide customer service chatbots, the legal basis may be a contract (if the chatbot is part of the service contract) or legitimate interest (if the chatbot is used to improve customer satisfaction or loyalty). If an AI is used by a researcher to generate academic content, the legal basis may be public interest (if the research is funded by a public body or serves a public purpose) or consent (if the researcher obtains the consent of the data subjects whose data is used to train or fine-tune the AI).

OpenAI must ensure that it has a valid lawful basis for each of its processing activities and that it documents and communicates this basis to the data subjects - the users of ChatGPT. Whenever an individual wishes to use the services of OpenAI by any technical means, he/she must agree to the terms of use and privacy policy of the company: *"(…) Performance of a contract with you when we provide and maintain our Services. When we process Account Information, Content, and Technical Information solely to provide our Services to you, this information is necessary to be able to provide our Services. If you do not provide this information, we may not be able to provide our Services to you"*.[69] This implies the necessity of the data process for the performance of the agreement between the user and OpenAI. This falls under Articles 6 (a) and 6 (b) of the GDPR, as the data subject (in this case, the user) also consents to the

[68] *Supra* note 9, art. 6 (1).
[69] *Supra* note 12, section 9.

processing of his/her personal data for one of the specific purposes specified in OpenAI's privacy policy by agreeing to the terms of service. It should be noted that in this case, both the conclusion of the agreement and the giving of consent require the individual to have either legal capacity under national law or the consent of a guardian. Therefore, OpenAI's data processing activities, governed by lawful bases under Articles 6 (a) and 6 (b) of the GDPR, are established through a contractual agreement and the user's explicit consent outlined in OpenAI's privacy policy, with the user's legal capacity or guardian's consent ensuring both the conclusion of the agreement and compliance with legal standards for data protection.

In conclusion, OpenAI's data processing activities are multifaceted and grounded in various legal bases. This ensures both operational efficiency and adherence to privacy regulations.

4. Purpose Limitation

This section will thoroughly examine OpenAI's Privacy Policy, with a focus on the purpose limitation principle outlined in Article 5 (1) (b) of the GDPR. The analysis will then extend to the provisions of Article 89, with an emphasis on pseudonymisation as a key safeguard. The definitions of aggregation, de-identification, and pseudonymisation will be explored in depth to evaluate OpenAI's compliance with the GDPR. The study will then examine the effect of ChatGPT's commercial development on the applicability of Article 89. This will provide a thorough understanding of OpenAI's data processing practices and potential consequences.

Article 5 (1) (b) of the GDPR states that personal data must be "*collected for specified, explicit and legitimate purposes and not further processed in a manner incompatible with those purposes*". However, further processing may be legitimate for specific purposes, such as a) archiving purposes in the public interest, b) scientific research, c) historical research, and d) statistical purposes. Article 89 (1) of the GDPR provides safeguards for the processing of data by endorsing pseudonymisation as a protective measure to ensure data minimisation.

In comparison, in the Privacy Policy, OpenAI states that it aggregates and de-identifies personal data for the purpose of conducting research and other similar purposes.[70] Thus, in order to determine whether OpenAI's policy complies with Article 89, it is necessary to clarify and compare the definitions of "aggregation" and "de-identification" with "pseudonymisation".

Recital 28 of the GDPR states that the application of pseudonymisation to personal data may reduce the risks to data subjects and assist the controller in complying with its data protection obligations. This statement clearly indicates that the use of pseudonymisation is encouraged by the legislator. According to the GDPR, pseudonymisation means to process personal data

[70] *Id.*, section 2.

so as to make it unattributable to any specific data subject, unless additional information has been provided, subject to such additional information being stored separately and subject to technical and organisational measures to ensure it is not attributed to any specific or identifiable natural person.[71] An example of pseudonymisation is the replacement of a customer's phone number with a random number or symbol. For instance, a telephone number could be displayed as "###-###-####". It is important to note that the legislator does not intend to exclude other methods of protecting personal data.

While the GDPR encourages the use of pseudonymisation to protect personal data, the CCPA goes a step further by defining de-identified information and outlining specific steps to ensure such information is not re-identified or disseminated. Both regulations highlight the importance of protecting personal data, but they approach it in slightly different ways.

This paragraph will examine the CCPA's approach. De-identified information is defined in the CCPA as information that cannot reasonably be identified, related, described, or associated, directly or indirectly, with a particular consumer. To ensure that such information is not re-identified or disseminated, a company that uses de-identified information must take four operational and organizational steps.[72] According to the steps listed by the legislator, the controller must: i. implement technical safeguards to prevent re-identification of the data subject; ii. implement business processes that explicitly prohibit re-identification of the information; iii. implement business processes to prevent inadvertent disclosure of deidentified information; iv. not attempt to re-identify the information. A straightforward method of de-identification involves replacing personal identifiers such as names and email addresses with random numbers or codes. This enables businesses to maintain customer records while protecting their personal information.

In a similar manner, aggregation of personal information refers to the removal of individual consumer identities so that they cannot be linked to a consumer or household. This process effectively generalizes the information to a group or category of consumers under the CCPA. The aggregated data will then be processed for statistical purposes only, and will not be used to take action or make decisions about specific individuals under the GDPR.[73] To better understand aggregation, consider a company using a generative AI model to uncover trends in the music industry. The company has amassed a vast collection of data from a variety of musical genres, such as rock, pop, classical, and jazz. By aggregating this data, the company can identify overarching trends in music. Under the CCPA, the company must ensure that this data is de-identified and cannot be linked back to any specific individual.

[71] *Supra* note 9, art. 4 (5).

[72] *Supra* note 36, § 1798.140.

[73] *Supra* note 9, recital 162.

In summary, these de-identification processes which are more effective than pseudonymisation can be used in the US to comply with the CCPA. In contrast, anonymisation, similar in nature to de-identification, is used in Europe as an alternative for GDPR compliance.

The main difference between anonymisation and pseudonymisation under the GDPR lies in the reversibility of the process. Anonymisation removes or alters personal data in a way that it cannot be linked back to a specific individual, making the process irreversible. Conversely, pseudonymisation replaces personal data with pseudonyms or identifiers. This process is reversible if the additional information necessary for re-identification is available. Therefore, pseudonymised data is still subject to GDPR protection, while anonymised data is not.

According to the GDPR, anonymisation is the modification of personal data in such a way that the person behind the individual data can no longer be identified. The EU legislator defines anonymous information as "*...information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable*". In general, under the GDPR, information can be directly or indirectly identifiable if it relates to an identified or identifiable natural person.

The following types of information are all considered direct identifiers under the GDPR: name, address, postal code, phone number, Internet Protocol (IP) address, photograph or image, and any other unique personal characteristics. Indirect identifiers can be used by third parties in conjunction with other sources of information to identify an individual. For example, information about work location, job title, salary, equipment ID, etc. According to the legislator, once information is fully anonymised - it does not contain any direct or indirect identifiers - it no longer falls under the requirements of the GDPR.

In conclusion, both methods aim to protect personal data, but they differ in terms of reversibility and the level of data protection they offer under the GDPR. Thus, pseudonymisation cannot be an alternative to de-identification under CCPA.

To sum up, pseudonymisation does not completely prevent information from being attributed to a natural person. It has limitations, although it can be a good way to protect the security and privacy of personal information. Pseudonymised data can be relatively easily identified indirectly, even if it does not directly identify an individual. Anonymisation, on the other hand, makes the information completely unattributable to any natural person. Moreover, the GDPR's concept of anonymisation is stricter than the CCPA's de-identification requirement, as the GDPR requires that an individual's identifiable information be "*irreversibly prevented*" from being used.

In contrast, the potential risks associated with the re-identification of previously de-identified information are significant, as the data can easily be misused and abused. Re-identification can result in privacy violations by revealing sensitive information about individuals that was meant to be anonymous. It can also lead to a loss of trust among individuals whose data was collected and potentially harm the reputation of the organization responsible, such as OpenAI and the organizations that oversee their data protection policy. These actions can harm individuals, ranging from identity theft to discrimination or stigmatization, particularly if the data is sensitive, such as health data.[74] As the GDPR requires stricter measures under the definition of anonymisation, OpenAI may not comply with this requirement, while its privacy policy complies with the CCPA in terms of de-identification or de-personalisation.

In addition to the above, Article 89 applies to research or statistical purposes. Although the development of earlier versions of ChatGPT was funded solely as a research project, OpenAI now offers a monthly subscription for "ChatGPT Plus", which gives access to the latest updates before the free version of the chatbot. Over time, ChatGPT is becoming more commercial than just an AI research project. In 2019, OpenAI announced the creation of OpenAI LP, a distinct entity that operates as a 'capped-profit' corporation.[75] This marked a transformative shift in its foundational structure. As a result, Article 89 may not apply to newer ChatGPT models due to their commercial nature.

5. Time Period of Data Processing and Transparency

The duration of processing should take into account the reasons why the controller needs to process the data, as well as any legal obligations to keep the data for a fixed period of time. The GDPR gives individuals the right to be informed about the collection and use of their personal data, which leads to a variety of information obligations on the part of the controller.[76] With regard to the content of the information, the controller is obliged to provide information on its identity, whether a data protection officer has been appointed, his contact details, the purposes of the processing, the legal basis, etc. In addition, the right to information also includes information about the duration of storage, the rights of the data subject, the possibility to withdraw consent, the right to lodge a complaint with the authorities, and whether the provision of personal data is a legal or contractual requirement.

[74] Khaled El Emam, Elizabeth Jonker, Luk Arbuckle and Bradley Malin, *Correction: A Systematic Review of Re-Identification Attacks on Health Data*, 6 Plos One, 2-3 (2015). Available at: https://doi.org/10.1371/journal.pone.0126772 (last visited Feb. 4, 2024).

[75] OpenAI LP (2019), https://openai.com/blog/openai-lp (last visited Jan. 25, 2024).

[76] *Supra* note 9, art. 13-14.

When reviewing the Privacy Policy, the section where OpenAI addresses the rights of data subjects does not make it clear for what period of time or activity they need to process and hold the personal data. For such situations, the GDPR emphasises that where it is not possible to determine the period for which personal data will be retained, the criteria used to determine the period should be clear to data subjects.[77]

However, analysis of the purposes that OpenAI lists in its privacy policy, shows that it is hard to guess the time needed for processing. For example, the data processed for the purpose of improving ChatGPT and conducting research is already integrated into the AI system, which is complicated for AI to forget. This means that the data can be used indefinitely. Therefore, there are no clear criteria for determining the time period. OpenAI models, thus, do not meet the duration of processing requirement of the GDPR.

Although OpenAI's models may not meet the GDPR's requirements for data processing duration due to their inherent design, the organization's approach to transparency further complicates matters. OpenAI provides some information about its data collection and training methods, but it remains guarded about the specifics. This adds another layer of complexity to the discussion around data privacy and protection.

OpenAI's transparency is still controversial, as they are protective when it comes to disclosing the data they collect and use for training. The technical report recently published by OpenAI states that the GPT-4[78] is pre-trained to predict the next token in a document using both publicly available data, such as internet data, and data licensed from third-party providers. The model is also fine-tuned using reinforcement learning from human feedback. However, due to the competitive environment and security implications, the report does not provide further details about the model, including dataset construction, training methods, and model size (which presumably includes the size of the data processed).[79]

Another aspect of transparency is governed by Article 14 of the GDPR which applies to data collected from the internet and used in training. This provision determines information which should be provided where personal data has not been obtained from the data subject.[80] However, in many cases, providing the data subject whose data were part of the training data with the information required by Article 14 of the GDPR may require a disproportionate effort and therefore may not be necessary. Regarding the definition of such effort, the Article 29 Working Party considers that a

[77] *Id.*, art. 13 (2) (a).

[78] A transformer-style model, further developed version of GPT-3.5 model. Both are the basis architectures for ChatGPT.

[79] OpenAI, GPT-4 Technical Report, 2 (2023).

[80] *Supra* note 9, art. 14.

disproportionate effort arises in particular when the data are collected from a large number of individuals whose contact details are unknown.[81]

In conclusion, OpenAI's models and methods offer significant advancements in AI technology. However, they also present complex challenges in terms of data privacy and protection. The organization's approach to transparency, particularly regarding the specifics of its data collection and training methods, remains a contentious issue. Additionally, the application of the GDPR provisions to data collected from the internet for training purposes adds another layer of complexity. Balancing the need for data to improve and innovate AI technology with the imperative of protecting individual privacy rights is a nuanced issue that requires ongoing attention and careful consideration.

6. Data Transfer

This section will comprehensively explore OpenAI's data transfer practices in accordance with the GDPR's stringent requirements. The analysis will focus on scrutinising EU-US data transfers involving OpenAI. The analysis emphasises the explicit consent requirement under Article 49 (a) of the GDPR and assesses the Privacy Policy's compliance with GDPR standards. The assessment sheds light on potential gaps and risks associated with the lack of transparency in OpenAI's data transfer practices.

When it comes to data transfers, the GDPR lays down strict rules that must not be compromised under any circumstances: "*Any transfer of personal data (...) shall take place only if (...) the conditions laid down in this Chapter are met by the controller and the processor, (...)*".[82]

As OpenAI is a company registered in the US, the EU-US data transfers should be analysed to determine whether or not the transfer of personal data processed, or to be processed, by OpenAI complies with the GDPR.

Data transfers between the EU and the US were previously regulated by the Adequacy Decision on the EU-US Privacy Shield. It was adopted on 12 July 2016 and allowed the free flow of data to companies certified in the US under the Privacy Shield. In its judgment, the Court of Justice of the European Union annulled the Adequacy Decision for not providing an adequate level of protection for the transfer of personal data from the EU to the US. The primary reasons were surveillance programs, a lack of judicial redress for EU individuals, and inadequate supervisory mechanisms.[83] The EU-US Privacy Shield is therefore no longer a valid mechanism for transferring personal data from the European Union to the United States. After more than a year of

[81] Article 29 Data Protection Working Party, Guidelines on Transparency under Regulation 2016/679, para. 58 (2018).

[82] *Supra* note 9, art. 44.

[83] Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems, C-311/18, Judgement, para. 168, 178 (2020).

negotiations, a new EU-US Data Privacy Framework was announced on 25th March 2022 which implements the commitments made by the US in favour of Europeans.[84] The new framework re-establishes an important legal mechanism for transferring EU personal data to the US. However, the details of the framework are still being translated into legal documents.[85]

In the absence of an adequacy decision, organisations can use SCCs approved by the European Commission for data transfers. SCCs are pre-approved model data protection clauses that allow data exporters to transfer data to countries outside the European Economic Area (EEA) that the European Commission has identified as providing an 'inadequate' level of data protection.[86] There are two sets of SCCs: (1) SCCs for the relationship between controllers and processors; and (2) SCCs as a tool for data transfers.[87] OpenAI does not have a disclosed processor in the EU. They collect the data directly (as both controller and processor). Therefore, SCCs cannot be applied in this case.

Another option may be BCRs.[88] Multinational organisations can establish BCRs, which are internal rules for cross-border data transfers within the same group. BCRs must be approved by the relevant data protection authorities to ensure that they provide adequate safeguards for the protection of personal data. However, as set out in recital 110, the BCRs may be applied by a group of companies (enterprises) engaged in a joint economic activity based on international transfers from the EU to organisations within the same group of companies.

In the complex field of data transfer mechanisms, it is important to consider OpenAI's potential use of BCRs or SCCs, especially given the unique structure of the organization. The establishment of OpenAI Ireland Limited in 2023 allows for the application of BCRs. However, the lack of updates in their privacy policy regarding the use of either method introduces uncertainty. The lack of clarity in OpenAI's communication about its approach to data transfers raises concerns about its commitment to transparency. This is particularly significant when examining the specifics outlined in the Privacy Policy related to user data transfer to OpenAI's U.S. facilities. It is important to note that the user's express consent is required for this transfer, as governed by Article 49

[84] European Commission, EU-U.S. Data Privacy Framework (2023).

[85] United States and European Commission Joint Statement on Trans-Atlantic Data Privacy Framework (2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/25/united-states-and-european-commission-joint-statement-on-trans-atlantic-data-privacy-framework/ (last visited Dec. 26, 2023).

[86] *Supra* note 9, recital 168, art. 46 (2) (c) (d).

[87] Standard Contractual Clauses (SCC): Standard Contractual Clauses for Data Transfers between EU and non-EU Countries (2021), https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en (last visited Dec. 13, 2023).

[88] *Supra* note 9, art. 46 (2) (b) and art. 47.

(a) of the GDPR. It highlights the importance of providing detailed and current information on the mechanisms and safeguards used. It is crucial for OpenAI to comply with GDPR standards and ensure user trust.

Regarding data transfers by OpenAI, the Privacy Policy states that personal information will be transferred from the user's location to OpenAI's facilities and servers in the United States. OpenAI assures users that appropriate safeguards will be applied when transferring personal data outside the EEA and that personal data will only be transferred pursuant to a lawful transfer mechanism.[89] By accepting the Privacy Policy, users expressly consent to the transfer of their data outside the EU. This situation is governed by Article 49 (a), according to which the transfer may be lawful if the data subject has expressly consented to the proposed transfer after having been informed of the possible risks of such transfer to the data subject due to the absence of an adequacy decision and of appropriate safeguards.

The details of data transfers, when and where they are required, what the appropriate safeguards are, how a transfer mechanism is considered legally valid, and, most importantly, the information on the potential risks of such transfers to the data subject due to the lack of an adequacy decision and appropriate safeguards, are not included in the Privacy Policy. Thus, it appears that OpenAI's data transfer policy is not GDPR compliant at present.

On the whole, the decisions and orders of the data protection authorities rightly point to the legitimate interests and rights of the data subjects to be informed about how their personal data is used in the training and fine-tuning of large language models and generative AI models, including ChatGPT. It should be seen as a directive from the authorities to the developer community to share crucial information about training, personal data and associated risks with the general public, rather than keeping secrets under a misnomer.

B. Compliance with the Data Protection Rights of the Third Parties

The personal data of third parties could be processed by ChatGPT in several cases. These cases may be as follows:

a) User input – when a user enters data belonging to a third party in the prompt;

b) APIs – with the help of APIs, ChatGPT can integrate with external systems and services. Authorised API connections allow the chatbot to retrieve data from third parties, such as user account information or relevant data from external databases;

c) Web scraping – by scraping publicly available information, ChatGPT can access data from websites or web services. Web scraping involves

[89] *Supra* note 12, section 9.

extracting data from HTML pages or structured APIs provided by the website owner.

d) Database Integration – ChatGPT may be given access to certain databases that store third-party data. This may involve integration with an external database or permission to access specific data sources maintained by third parties.

1. User Input

User content, in particular user input or prompts, is collected and stored by the OpenAI. This input may include the personal data of third parties. This can be done either on purpose or unintentionally.

Under the Terms of Use, if a User intends to process personal data by using ChatGPT, he/she must provide legally adequate privacy notices and obtain the necessary consent. The User shall be responsible for the lawfulness of such processing. OpenAI therefore disclaims any responsibility for the processing of third-party data entered by the user.

In contrast, the GDPR clearly states that the controller has the primary responsibility for ensuring that personal data is processed in accordance with the principles and requirements of the GDPR: "*(...) the controller shall implement appropriate technical and organisational measures to ensure and be able to demonstrate that processing is carried out in accordance with this Regulation*".[90] While certain responsibilities may be shared or delegated (e.g. with processors, joint controllers, and sub-contractors), the controller always retains overall responsibility and should maintain a level of supervision and control over the processing activities. Thus, OpenAI's disclaimer does not comply with the requirements of the GDPR.

2. APIs

As discussed earlier, OpenAI's API allows for the customization and integration of various models, including ChatGPT, into different platforms. The use of ChatGPT becomes particularly challenging from a privacy perspective when companies integrate it into their website or build their own applications based on the language model technology available through the API and offer them to their customers.[91]

Before the integrated services are made available to customers, appropriate data protection agreements, such as a Data Processing Agreement (DPA) or Joint Controller Agreement (JCA), must be in place with OpenAI. OpenAI provides a Data Processing Addendum[92] for API users. However, according

[90] *Supra* note 9, art. 24.

[91] Xiaodong Wu, Ran Duan and Jianbing Ni, *Unveiling Security, Privacy, and Ethical Concern of ChatGPT*, Journal of Information and Intelligence, 7-8 (2023). Available at: https://doi.org/10.1016/j.jiixd.2023.10.007 (last visited Feb. 9, 2024).

[92] *Supra* note 30.

to OpenAI, it will not sign any DPA provided by the user or any amendments to its own DPA.

Moreover, companies need to be aware that signing a DPA with OpenAI is by no means sufficient. As data will be transferred to the US, the EU standard contractual clauses (SCCs) will also need to be included in the agreement between the EEA-based company and OpenAI. Companies should be aware of any additional protections under DPAs and will need to conduct a transfer impact assessment[93] under SCC requirements.

Another problematic aspect of API arises when they are used in the recruitment process, as well as the assessment, promotion or retention of individuals in work-related contractual relationships. This concern arises from AI's potential to perpetuate historical patterns of discrimination, for example against women, certain age groups, people with disabilities, or people of certain racial or ethnic origin or sexual orientation.[94] So the companies must ensure that no automated decision-making in this context leads to unlawful data processing without human intervention if they intend to integrate ChatGPT into their recruitment process, such as the selection of applicants.[95]

Additionally, the staff of the companies should be made aware that they should not enter personal information about a client, supplier, business partner or colleague in ChatGPT. If they do, they could violate confidentiality obligations, including those arising from the GDPR, which are regularly required of them by their employers or business partners, as the case may be.

This is particularly important because, unlike OpenAI's non-API consumer products, the use of data submitted by customers via API to train or improve AI models is denied by default. Therefore, data will be accepted only if the user opts into data sharing by completing OpenAI's Data Sharing Opt-In form.[96] This means that, in the event of a data breach, the responsibility falls on the user who inputs the data, depending on the specific circumstances. Thus, it is crucial for both the users and companies to understand the potential risks and use AI responsibly.

Furthermore, any data sent via the API will be retained for up to 30 days to help monitor for abuse and misuse. Except where otherwise required by law, the data will be deleted at the end of this period. A limited number of authorized OpenAI employees and specialist third-party contractors who are subject to confidentiality and security requirements can access this data only

[93] Transfer impact assessment is an evaluation process that companies undertake to assess the potential risks and impacts associated with transferring personal data from one jurisdiction to another.

[94] J. Stewart Black & Patrick van Esch, *AI-enabled Recruiting: What is It and How should a Manager Use It?*, 63 Business Horizons 215, 223 (2020).

[95] *Supra* note 9, art. 22.

[96] *Supra* note 29.

to investigate and review suspicions of abuse. If OpenAI suspects that data contains platform abuse, it may still flag the data using content classifiers. In conclusion, OpenAI's APIs present challenges for GDPR compliance. Although OpenAI provides DPAs, it does not sign user-provided DPAs or modify its own. For EEA-based companies, agreements must include EU standard contractual clauses to facilitate data transfer to the US. When using ChatGPT, it is important to avoid discriminatory outcomes, and users must be aware of their responsibility for data breaches. OpenAI's policy prohibits the use of customer data for training without explicit consent, but potential risks of data sharing still exist. The 30-day data retention policy and OpenAI's control over non-training data processing duration highlight the importance of user understanding and responsible AI use. Therefore, due to these complexities and challenges, OpenAI's API policies are not GDPR compliant.

3. Web Scraping

Data scraping is a technique for the extraction of data from the Internet.[97] Conventional web scraping involves writing code that traverses web pages, locating specific data elements, and extracting those elements. AI-powered web scraping takes this one step further, using machine learning and natural language processing algorithms to automate the extraction process. AI-based web scrapers can be trained to recognize patterns, understand web content semantics, and extract data from even complex and dynamically changing web pages.[98] This activity must be conducted in accordance with legal requirements. Websites may have terms of service or policies that restrict or prohibit web scraping activities.

OpenAI's policies require individuals who extract data from a website using ChatGPT to confirm that they are processing such data in accordance with applicable law, including respecting websites' terms of service or policies that may restrict or prohibit web scraping activities. According to OpenAI, they provide the tools for data extraction, but it is the responsibility of the users to ensure compliance with all relevant laws and regulations, including data protection, privacy laws, and the terms of service of the websites being scraped.[99] OpenAI aims to balance the benefits of AI-powered data extraction with the need to respect individual privacy and data protection rights. However, according to Articles 24 and 25 of the GDPR OpenAI, as the data controller, remains responsible for data processing.

[97] Data Scraping and Data Mining from Beginner to Pro with Python Sciences AI (2021), https://www.oreilly.com/videos/data-scraping-and/9781801818483/ (last visited Jun. 14, 2023).

[98] Kumar Swarn Avinash, Moustafa M. Nasralla, Iván García-Magariño and Kumar Harsh, *A machine-Learning Scraping Tool for Data Fusion in the Analysis of Sentiments about Pandemics for Supporting Business Decisions with Human-Centric AI Explanations*?, 7 PeerJ Computer Science, 2 (2021).

[99] *Supra* note 29.

4. Database Integration

Among AI models, ChatGPT stands out for its ability to seamlessly integrate with databases, automating and improving data handling processes. By training ChatGPT on business data, individuals can create a personalised AI chatbot that can handle a wide range of tasks, from customer service to sales. Businesses can use a cloud-based platform such as GPT-3, or use the API to integrate the model into their existing systems.[100] In all cases, individuals and companies provide access to the data, which is likely to include not only their customers' data but also contracts with other companies, employee data or other information that may contain trade secrets and confidential data.

OpenAI disclaims responsibility to third parties when it comes to processing third-party data in all of the above cases, emphasising that the data is entered by the user and the user must ensure that the process complies with applicable law, in the EU with the GDPR. At the same time, for their API (commercial) models, they do not process data belonging to third parties unless explicitly requested by the user. OpenAI's approach to disclaiming responsibility for third-party data processing is not uncommon in the tech industry. The idea is that the user, who inputs the data, is responsible for ensuring that the data complies with applicable laws. For example, Salesforce states that they are not responsible for the privacy or data security practices of their customers, which may differ from those explained in their Privacy Statement.[101]

In contrast, Article 28 (1) of the GDPR places responsibility on organizations providing tools for data processing to ensure that their systems are designed and used in a way that respects data protection principles. Therefore, the GDPR mandates that all businesses and their partners are responsible for protecting user data.

As previously discussed, AI has the potential to perpetuate historical patterns of discrimination. As such, AI developers may need to take more responsibility for ensuring that their technology is fair and unbiased. To sum up, despite the general attitude of Open AI, there are valid arguments for why ChatGPT developers should be more responsible for third-party data processing.

III. What Can be Done by OpenAI to Ensure the GDPR Compliance?

The following section explores the measures OpenAI can implement to achieve GDPR compliance for its AI tool, ChatGPT. These steps are crucial not

[100] OpenAI, Introducing ChatGPT Enterprise. Available at: https://openai.com/blog/introducing-chatgpt-enterprise (last visited Feb. 5, 2024).

[101] Salesforce Privacy Statement, https://www.salesforce.com/company/privacy/full_privacy/ (last visited Dec. 27, 2023).

only for securing user privacy and upholding legal regulations but also for fostering a more trustworthy and transparent legal environment.

By establishing clear data collection and usage practices, OpenAI can contribute to a legal sphere where individuals have greater control over their information and businesses operate within a clear and predictable framework. This, in turn, can lead to increased user confidence in utilizing AI-powered tools within legal processes, ultimately benefiting the legal system as a whole.

To address the issue of data minimisation and retention, OpenAI must implement mechanisms to examine the information gathered by ChatGPT, primarily through prompts. To achieve this goal, it is essential to develop a reliable prompt analysis system in ChatGPT that minimises unnecessary data collection. OpenAI should create an advanced algorithm that actively analyses prompts in real time and extracts only essential information to fulfil the intended purpose of ChatGPT. In accordance with the GDPR's principle of data minimisation, the algorithm should exclude any personal data that is not necessary.[102]

Additionally, OpenAI must ensure that ChatGPT does not retain data for longer than required. A more detailed time frame should be set for retaining personal data in accordance with the GDPR, which requires specific time periods to be established wherever possible. It is recommended that ChatGPT retains personal data for the duration of a conversation or session and deletes it afterwards unless there is a legal or contractual obligation to keep it longer. Alternatively, ChatGPT may retain personal data for a longer period, such as a month or a year, for research or statistical purposes, provided appropriate safeguards are applied, such as anonymisation, encryption, or pseudonymisation.

Moreover, OpenAI should implement anonymization or pseudonymization of user data before storing or processing it to protect the user's identity and privacy. As discussed earlier, complete anonymization of user data will exempt the process from the GDPR, and pseudonymization is sufficient to ensure GDPR compliance under Article 89 for research purposes. Furthermore, it is essential to provide users with clear and transparent information regarding the data collected, its usage, and how they can exercise their rights. This will increase user awareness and trust.

Furthermore, OpenAI has listed the purposes of processing under Sector 8 in its Privacy Policy. However, the GDPR requires more than just listing the purposes of data processing. The GDPR requires organizations to process personal data in a transparent manner. However, the training of ChatGPT lacks transparency in this regard. OpenAI should provide clearer information

[102] Abigail Goldsteen, Gilad Ezov, Ron Shmelkin, Micha Moffie and Ariel Farkash, *Data Minimization for GDPR Compliance in Machine Learning Models*, 2 AI and Ethics 477, 481-482 (2022). Available at: https://doi.org/10.1007/s43681-021-00095-8 (last visited Feb. 9, 2024).

on how they reduce the amount of personal information in their training datasets before using such data to improve their models.

OpenAI needs to demonstrate that it has the necessary organisational and technological safeguards in place to prevent a data breach from happening. It should also have mechanisms in place for the handling of any data breaches that do occur. In accordance with Article 55 of the GDPR, data breaches that could put individuals at risk must be reported to the data protection authority and the individuals concerned within 72 hours, without undue delay. Additionally, to ensure transparency and user understanding of data collection and usage, the chatbot should provide a simple and easily accessible explanation form. This form must include privacy regulations that govern the collection of users' data. OpenAI can provide this information through a link in the chatbot's conversational flow or a condensed version in the welcome message and dialogue.

Finally, OpenAI has already listed its current subprocessors involved in third-party data processing, demonstrating ongoing compliance with the GDPR. To further comply with Article 24 (1) of the regulation, appropriate technical and organizational measures must be implemented to safeguard user data throughout its lifecycle. These measures include encryption, access controls, and robust data retention policies. Establishing clear procedures for handling data breaches is crucial to minimize potential risks to users and uphold their right to be informed. By comprehensively addressing these aspects, OpenAI will further strengthen its commitment to GDPR compliance and user data privacy.

By implementing the proposed solutions and incorporating the suggested improvements into its Privacy Policy, OpenAI can demonstrate a stronger commitment to GDPR compliance. This will contribute to building a more trustworthy and transparent legal environment, fostering user confidence in utilizing AI-powered tools within legal processes and ultimately benefiting the legal system as a whole.

Conclusion

In the current digital landscape, safeguarding data has become a crucial aspect of business, particularly in light of recent advancements in AI. AI chatbots, such as ChatGPT, primarily function by collecting data, which necessitates compliance with data protection regulations. This is because AI developers require data to construct a chatbot that can truly comprehend context and facilitate meaningful conversations.

The GDPR is a set of regulations provided by the EU to grant legal control to its residents over the sharing, updating, and removal of their private information online. The law aims to promote transparency in the relationship between companies and users when collecting and storing data and to protect online privacy.

Currently, evaluating ChatGPT's compliance with the GDPR is challenging due to incomplete operational information. The extent of personal data in ChatGPT's original dataset remains unclear, but it is reasonable to assume that the vast amount of data used to train it contains personal information. This information is still present in the dataset used by ChatGPT. When questioned about the use of data for training, the chatbot claims that all data has been anonymised and scrubbed to remove any identification. However, verifying this claim is fundamentally difficult, even for knowledgeable users. As discussed in Chapter I, OpenAI specifies certain categories of purposes that are limited by the Privacy Policy when it comes to the use of personal data. However, it is difficult to determine whether or not ChatGPT complies with the purpose limitation principle because OpenAI's privacy policy is constantly changing. This is particularly true when OpenAI publicly states that they may use personal data to improve their models.

According to OpenAI's policy, user content is not shared with third parties for marketing purposes. The Policy does not, however, provide explicit information on how data sharing with third parties for non-marketing purposes is handled.

In addition, a wealth of personal information is made available to the computer through interactions. OpenAI states that the company collects personal data through using services, through interactions, including the type of content users are engaging with. The 'right to erasure' is the ability, under the GDPR, for an individual to request the complete removal of their data from the files of an organisation. Natural language processing technologies, such as ChatGPT, process personal data and transform it into a complex data structure, making it challenging to extract an individual's data. Therefore, the removal of such data poses a significant challenge. It is unclear whether ChatGPT can comply with the right to be forgotten as set out in Article 17 of the GDPR, as discussed in Chapter II.

Therefore, ChatGPT does not appear to be GDPR compliant. There is an apparent lack of transparency, the possibility of unlawful collection and use of personal data, and difficulties for data subjects to exercise their rights, including the right to information and the right to forget.

OpenAI should apply mechanisms to ChatGPT to examine the information it gathers, ensuring that ChatGPT collects only the minimum necessary information and does not retain it for longer than required. Anonymization or pseudonymization of user data should be implemented before storing or processing it to protect the user's identity and privacy. Additionally, transparency in processing personal data for the training of ChatGPT should be ensured. Further specific measures are listed in Chapter III of the paper.

In conclusion, given the increasing significance of AI in the modern world, it is crucial to regulate its use effectively. Therefore, analyzing the GDPR compliance of ChatGPT, one of the most prominent AI chatbots, and ensuring its adherence to GDPR standards is essential for safeguarding privacy and building trust in AI technologies.