THE RIGHT TO ERASURE OF PERSONAL INFORMATION IN LARGE LANGUAGE MODELS: CHALLENGES AND RESPONSES

MO Yunfan

Table of Contents

THE RIGHT TO ERASURE OF PERSONAL INFORMATION IN LARGE LANGUAGE MODELS: CHALLENGES AND RESPONSES

MO Yunfan

Abstract

The "Right to Erasure of Personal Information" (right to erasure), which grants individuals the opportunity to delete their personal information from the internet, has become a necessity in the information age. Academic discourse on the right to erasure has centered on the example of search engines as personal information processor and has been set against the backdrop of the internet age. However, the new era of Generative Artificial Intelligence (Generative AI) has given rise to new formidable problems with the protection of the right to erasure, especially in the case of Large Language Models (LLMs) that are trained on large amounts of data containing personal information. This paradigm shift, therefore, raises a host of questions as to how to protect this legal right properly in the age of Generative AI.

Against the backdrop of Generative AI, this paper argues how the Personal Information Protection Law (PIPL) should be applied to govern the protection of the right to erasure in response to the widespread collection and utilization of personal information by LLMs. Through examining unique technical characteristics of LLMs, such as model opacity, operational autonomy, retraining necessity and hallucination possibility, the paper first discusses the major challenges that arise in protecting the right to erasure within LLMs, such as access barrier, deletion dilemma, retraining difficulty, and object absence. The paper then seeks to propose corresponding legal responses to those challenges. To adapt to the evolving context of Generative AI while upholding the right to erasure, it is necessary to introduce third-party supervision that audit databases and filter personal information, to reinterpret the concept of "erasure" in Article 47 of the PIPL, adhere to the principle of purpose limitation throughout the entire personal information processing activities and allow sufficient flexibility for technological advancement. These measures, as the paper shows, would support the technological innovation and sustainable development in the AI sector while providing adequate protection of the right to erasure.

Keywords: Right to Erasure of Personal Information; Large Language Models; Personal Information Protection Law

I. INTRODUCTION

Sharing personal information and enjoying free Internet services is the keynote of the Internet age. However, when merchants are able to market, nudge and control individuals into loyal customers,[1] and when past disclosures keep haunting back as an indicator of evaluating individual's trustworthiness, individuals gradually realize the brutal truth that "if you are not paying for the product, then you are the product", and "forgetting has become the exception, and remembering the default" in this age.[2] The

[1] NEIL RICHARDS, WHY PRIVACY MATTERS 43 (1st ed. 2021).
[2] VIKTOR MAYER-SCHÖNBERGER, DELETE: THE VIRTUE OF FORGETTING IN THE DIGITAL AGE 1 (2009).

rising social awareness of personal information protection has brought the right to erasure of personal information into the public spotlight. This right aims to grant individuals the opportunity to delete their personal information from the internet, so that the merchants cannot track and market with personalized advertisement, and others cannot judge by past disclosures, in response to the self-determination concerns and the pervasive digital memory issues. Thus, this right is crucial for individuals living "transparent" lives in this digital age, forming a key pillar of personal information protection.

The demand for right to erasure has then made it a prominent issue on the legal agenda. The legal expression of the right to erasure can be dated back to the 1995 European Data Protection Directive,[3] and it really gained significant public attention since 2014, largely due to the rapid development of the Internet and the landmark ruling by the European Court of Justice (ECJ) in the Google Spain v. González case.[4] Following this, the European Union took the lead in recognizing the right to erasure at the legal level, with the General Data Protection Regulation (GDPR) officially coming into force in 2018, where Article 17 enshrines this right under the title "right to erasure (right to be forgotten)".[5] While the underlying attitude of ECJ's judgments and the terminology in the GDPR's Article 17 ignited a worldwide academic debate on the relationship between the right to erasure and the right to be forgotten, China was not bound by this debate and adopted a clearer and more comprehensible expression with the "right to erasure" from a practical perspective. In 2021, with the enforcement of the Personal Information Protection Law (PIPL), the right to erasure of personal information was recognized in Article 47 of the PIPL in China. According to this Article, individuals have the right to request a personal information processor to erase their personal information during processing activities when specific stipulated conditions are met.[6] The right to erasure has also been established in the legal frameworks of the South Korea,[7] Japan,[8] California[9] and other countries and regions, the concept and value of the right to erasure of

---

[3] Directive 95/46/EC, of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, O.J. (L281) 31.

[4] Case C-131/12, Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, ECLI:EU:C:2014:317 (May 13, 2014).

[5] Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), O.J. (L119) 1, 17–18.

[6] Geren Xinxi Baohu Fa (个人信息保护法) [Personal Information Protection Law] (promulgated by the Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021), art. 47 (Chinalawinfo).

[7] GAEINJEONGBO BOHOBEOB [PERSONAL INFORMATION PROTECTION ACT], art. 36 (S. Kor.).

[8] KOJIN JŌHŌ HOGO-HŌ [ACT ON THE PROTECTION OF PERSONAL INFORMATION], art. 35 (Japan).

[9] CAL. CIV. CODE § 1798.192 (2023).

personal information have gradually gained recognition worldwide in law materials.

In 2022, with the surprise launch of ChatGPT, the world officially entered the era of Generative AI. Among Generative AI, Large Language Models (LLMs)—computational models capable of language generation and other natural language processing tasks—have seen particularly rapid and outstanding advancement.[10] Following ChatGPT, numerous LLMs such as Copilot, Gemini, and LLaMA have appeared globally, while Chinese tech companies have also quickly introduced models like ERNIE Bot, ChatGLM, Tongyi Qianwen, and Kimi.[11] These models are characterized by their ability to be pre-trained using a large amount of data crawled from the web, forming foundational models that, through interaction with users, predict and generate contextually relevant, logically sound, and coherent replies in response to the user's input, rather than indexing the source web page content as search engines have done in the past.[12] Given the fact that personal information is inevitably contained in the crawled data and forms part of the training database, new challenges to the personal information protection arise. What behind the LLMs are widespread information collection, increasingly sophisticated algorithmic logic, opaque black box and less predictable outputs, meaning that the application of the right to erasure has become more intricate and unpredictable compares to search engines in the internet age, while legal research related to this issue has struggled to keep pace with the development of technology.

During the theoretical development of the right to erasure, Chinese academic discourses have predominantly focused on theoretical justification and normative construction, largely following the trajectory set by the monumental Google Spain v. González case, with a particular emphasis on search engines as personal information processor, and against the backdrop of internet era.[13] While the past literature has adequately

---

[10] *Large Language Model*, WIKIPEDIA, https://en.wikipedia.org/wiki/Large_language_model (last visited Oct. 2024).

[11] *National Internet Information Office, Sheng Cheng Shi Rengong Zhineng Fuwu Yi Beian Xinxi* (生成式人工智能服务已备案信息) [*Generative AI Services Filed Information*], NATIONAL INTERNET INFORMATION OFFICE, https://www.cac.gov.cn/2024-04/02/c_1713729983803145.htm (last updated Apr. 2024).

[12] *Generative Artificial Intelligence*, WIKIPEDIA, https://en.wikipedia.org/wiki/Generative_artificial_intelligence (last visited Oct. 2024).

[13] *See* Zhang LiAn (张里安) & Han Xuzhi (韩旭至), *Bei Yiwang Quan: Dashuju Shidai Xia de Xin Wenti* (“被遗忘权”：大数据时代下的新问题) [*The "Right to be Forgotten": New Problem in the Big Data Era*], 35(3) HEBEI FAXUE (河北法学) [HEBEI LAW SCIENCE] 35 (2017); Yang Lixin (杨立新) & Han Xu (韩煦), *Bei Yiwang Quan de Zhongguo Bentuhua Ji Falv Shiyong* (被遗忘权的中国本土化及法律适用) [*The Localisation and Legal Application of the Right to be Forgotten in China*], 30(2) FAXUE LUNTAN (法学论坛) [LEGAL FORUM] 24 (2015); Ding Xiaodong (丁晓东), *Bei Yiwang Quan de Jiben Yuanli Yu Changjing Hua Jieding* (被遗忘权的基本原理与场景化界定) [*The Basic Principle and the Contextualized Definition of "Right to be Forgotten"*], 12(6) QINGHUA FAXUE (清华法学) [TSINGHUA UNIVERSITY LAW JOURNAL] 94 (2018); Wan Fang (万方), *Zhong Jiang Bei Yiwang de Quanli——Woguo Yinru Bei Yiwang Quan de Sikao* (终将被遗忘的权利——我国引入被遗忘权的思考) [*The*

dissected how to understand right to erasure in the internet age, the technological context shifted. As a result, few legal scholarships have addressed the practical applications of the right to erasure in the context of the unique data collection and utilization practices of Generative AI models like LLMs, which differ from those of traditional search engines. This thereby leaves a research gap regarding legal responses to the evolving context of exercising the right to erasure.

Thus, this discussion understands right to erasure through a lens of practical legal application, analyzes the challenges that the unique features of the LLM will bring to the practice of the right to erasure, and attempts to provide responses from a legal perspective. The paper will be anchored in the legal text of China's PIPL and its definition of the "right to erasure", in conjunction with the characteristics of LLMs as a representative product in Generative AI era. The paper will structure as follows:

1) the first section will discuss how LLMs involve the processing of personal information and whether the developers of these models can be deemed as personal information processors, laying the groundwork for subsequent discussions on the applicability of the right to erasure in the context of LLMs;

2) the second section will examine the technical attributes of LLMs, using search engines as a comparative reference, and explore the technical challenges and obstacles that may arise when individuals exercise the right to erasure in LLMs;

3) building on the insights from the first two sections, the third section will propose legal solutions in response to the technical challenges, so as to ensure that the right to erasure of personal information can effectively brace itself for the challenges in the new age and accordingly evolve with the times.

II. PERSONAL INFORMATION PROCESSING IN LARGE LANGUAGE MODELS

A. Do Large Language Models process personal information?

Whether and how the LLMs process personal information is the starting point of our discussion on personal information protection in Generative AI age. Clarifying the processing of personal information at each stage of the LLMs will facilitate further discussions on the application of the right to erasure, the possible challenges posed by the LLMs, and other issues related to the personal information protection. The establishment and operation of an LLM can be primarily divided into three

Right to be Forgotten: Reflections on the Introduction of the Right to be Forgotten in China], 34(6) FAXUE PINGLUN (法学评论) [LAW REVIEW] 155 (2016); Wang Liming (王利明), *Lun Geren Xinxi Shanchu Quan* (论个人信息删除权) [*On the Right to Delete Personal Information*], 15(1) DONGFANG FAXUE (东方法学) [ORIENTAL LAW] 38 (2022).

phases, namely, data collection, data training (including pre-training and fine-tuning), and data output interaction. Among these, since the processing of data in the data training phase primarily relies on the previous data collection phase, i.e. if personal information is included in the data collection phase, the training phase, which trains the model based on the collected data, will necessarily involve the processing of personal information. Therefore, the discussion will mainly focus on data collection phase and data output interaction phase.

1. Data Collection Phase

The training of foundational models relies on vast amounts of data, the larger the dataset, the more predictive the model will become in its later applications, allowing it to predict subsequent conversations and generate corresponding texts. As a result, the training of machine learning models is built upon the extensive collection of data. These data sources are highly diverse, ranging from public to specialized sources, including books, articles, websites, posts, and more. Models often crawl large volumes of data with the help of "digital intermediaries" such as network service providers, internet service providers, search engines, and online marketplaces,[14] crawling text from publicly available internet resources for use in model training.[15] These datasets scraped from the internet often contain substantial amount of personal information. For instance, personal information that has been disclosed legally or leaked illegally; personal information that an individual has consented to be processed by other processors or that has been processed for news reporting purpose. The models will crawl personal information of public figures as well as other individuals, regardless of the difference in subject identity and regardless of the authenticity and accuracy of information.[16]

As a matter of fact, the collection of data for LLMs remains a grey area, raising doubts about the legality of data sources and whether the processing is based on lawful grounds. China's Regulation on Network Data Security Management, which will be implemented on 1 January 2025, begins to touch the problems that exist in the data collection phase, stating that "[w]here it is impossible to avoid the collection of unnecessary personal information or an individual's personal information without the consent of the individual in accordance with the law due to the use of

[14] Rita Ghial et al., *Right to Be Forgotten: A Human Rights-Based Approach for Governance in Generative AI*, *in* 1 SMART TRENDS IN COMPUTING AND COMMUNICATIONS: PROCEEDINGS OF SMARTCOM 2024 23, 29 (Tomonobu Senjyu et al. eds., 2024).
[15] Xabier Lareo, *Large language models (LLM),* EUROPEAN DATA PROTECTION SUPERVISOR, https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en (last visited Dec. 21, 2024).
[16] *Id.*

automatic collection technology or any other reason...the network data processor shall delete or anonymize the personal information."[17] While the effective implementation of this regulation is another new story related to the right to erasure of personal information, it is certain that the personal information processing is inevitable in the data collection phase of the LLMs. Thus, it is also clear that the data training phase, including pre-training and fine-tuning based on collected data, also involve the processing of personal information.

2. User Interaction Phase

Besides processing personal information during the data collection phase, the LLM will also involve the collection of personal information during the user interaction phase, i.e., the phase in which users interact with LLM by entering prompts in chatbox and giving feedback based on its answers.

As a pioneer of LLMs, ChatGPT was officially available for public use in November 2022. Within just five days of its release, ChatGPT had gained one million users. This innovative product quickly captivated users, encouraging them to try it by entering prompts to ask different questions.[18] The model's omniscient and human-like response, along with its resemblance to a new form of intelligent being, have created a sense of fascination for users. As a result, users' interactions with ChatGPT tend to be less sensitive and cautious than they would be with other applications or websites. In this context, individuals may inadvertently disclose their personal information through the chatbox. For example, users might share their medical details in hopes of receiving advice, or they may reveal a friend's birthday, hobbies, and family structure to get help planning a birthday party. Similarly, users may disclose details about their marriage or financial situation, seeking assistance in drafting a divorce agreement.[19] Now, if you ask the ChatGPT to "roast me based on your past prompts", you will find that it knows you better than you know yourself. Users enter and share their personal information during conversations with ChatGPT, and ChatGPT collects all

---

[17] Wangluo Shuju Anquan Guanli Tiaoli (网络数据安全管理条例) [Regulation on Network Data Security Management] (promulgated by the St. Council, Sept. 24, 2024, effective Jan.1, 2025) art. 24 (Chinalawinfo).

[18] Shubham Singh, *ChatGPT Statistics (SEP. 2024)-Users Growth Data*, DEMANDSAGE (Sept. 2, 2024), https://www.
demandsage.com/chatgpt-statistics/.

[19] Uri Gal, *ChatGPT is a Data Privacy Nightmare. If You've Ever Posted Online, You Ought to be Concerned*, THE CONVERSATION,
https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283 (last updated Feb. 9., 2023); Sara Reardon, *AI Chatbots Can Diagnose Medical Conditions at Home. How Good Are They?*, SCIAM (Mar. 31, 2023), https://www.scientificamerican.com/article/ai-chatbots-can-diagnose-medical-conditions-at-home-how-good-are-they/.

the content and use it to further train models. According to ChatGPT's Privacy Policy, personal information contained in user input, feedback, or uploaded files will be collected when users use the service, and will be used to improve the service, conduct research and develop new programs, etc..[20] ChatGPT is not the only LLM that is doing this, popular LLMs in China such as Tongyi Qianwen, ERNIE Bot, kimi, all explicitly indicate in their privacy policies that they will collect user inputs and may use them to enhance and iterate their products and services, which shows a common industry practice among LLM developers.[21] Through users' interactions, personal information will be continuously provided to ChatGPT, potentially being used to improve LLM services or becoming part of the pre-training dataset for new projects. While the specific content of this information may hold little value for the developers at the moment, that doesn't rule out the possibility that it could be utilized in the near future, perhaps one day it will be used in personalized advertisement by developers.

In summary, personal information will be provided to LLMs through the user's interaction, the content will be collected and used to improve LLM services or develop new projects by the developers. In short words, LLMs process personal information during the user interaction phase.

B. Are Developers of Large Language Models Personal Information Processors?

The essential question of whether the developer of LLM can be subject to individuals exercising their right to erasure of personal information lies in whether these developers can be deemed as personal information processors.

According to Article 73(1) of the PIPL, a personal information processor refers to an organization or individual that independently decides on the purposes and methods of personal information processing activities.[22] As previously discussed, LLMs do engage in personal information processing activities. The key question then moves to whether the developers of these models independently decide the purposes and methods of processing. The emphasis here is on "independent decision" and "processing purposes and methods," which distinguish a personal

[20] OpenAI, *Privacy Policy* (Nov.14, 2023), https://openai.com/policies/row-privacy-policy/.

[21] *Yinsi Zhengce* (隐私政策) [*Privacy Policy*], TONGYI QIANWEN, https://help.aliyun.com/zh/dashscope/developer-reference/tongyi-qianwen-privacy-policy (last updated Aug. 18, 2023); *Wenxin Yiyan Geren Xinxi Baohu Guize* (文心一言个人信息保护规则) [*ERNIE Bot Personal Information Protection Rule*], WENXIN YIYAN (last updated June 27, 2024), http://yiyanapp.baidu.com/talk/protectionrule/android; *Yonghu Yinsi Xieyi* (用户隐私协议) [*User Privacy Agreement*], KIMI, https://kimi.moonshot.cn/user/agreement/userPrivacy (last updated Dec. 19, 2024).

[22] Geren Xinxi Baohu Fa (个人信息保护法) [Personal Information Protection Law] (promulgated by the Standing Comm. Nat'l People's Cong, Aug. 20, 2021, effective Nov. 1, 2021), art. 73 (Chinalawinfo).

information processor from the party that is commissioned to process personal information. In particular, "independent decision" indicates that the organization or individual autonomously decides without relying on the opinions or requests from other entities, nor with their approval or consent. "Processing purposes and methods" refers to the objectives pursued in the processing of personal information, such as statistical surveys, and the specific techniques employed, such as storage, usage, and retrieval.[23] Using ChatGPT and its developer, OpenAI, as an example, this issue has different answers in different situations.

1. API Service and ChatGPT Enterprise Service

When the collection of data is done by the customer, and the purpose and methods of data processing is decided by the customer, OpenAI only carries out data processing activities according to the instructions. In this case, OpenAI is not a personal information processor, which is the case for the API service and ChatGPT Enterprise service provided by OpenAI.

In the API service, OpenAI's API can be accessed and used, and client companies can integrate ChatGPT into their products and services via the API. In the ChatGPT Enterprise service, the integration is more directly manifested as ChatGPT providing services for the client company's business operations. In these cases, ChatGPT's processing of data is limited to generating reports and analyses on behalf of the customer or complying with the customer's written instructions.[24] OpenAI does not at this point have the autonomy to determine the processing purpose and the processing method. Instead, the customer—whether an organization or individual—acts as the personal information processor that determines the purposes and methods of processing their own data. OpenAI, in turn, processes the data on the instructions of and on behalf of the personal information processor, functioning as a fiduciary of the personal information processor.[25]

2. Individual-Oriented Services

When OpenAI proactively develops AI systems and creates training datasets based on data it selects independently, determining both the purposes and means of data processing, OpenAI is deemed as a personal information processor.[26] The services provided by ChatGPT to individual

---

[23] CHENG XIAO (程啸) & WANG YUAN (王苑), GEREN XINXI BAOHU FA JIAOCHENG (个人信息保护法教程) [PERSONAL INFORMATION PROTECTION LAW COURSE] 37–39 (2023).
[24] *Data processing addendum*, OPENAI, https://openai.com/policies/data-processing-addendum/ (last updated Feb. 15, 2024).
[25] *Id*.
[26] *Determining the Legal Qualification of AI System Providers*, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (June 7, 2024), https://www.cnil.fr/en/determining-legal-qualification-ai-system-providers.

users fall under this category. In such cases, OpenAI, as the developer of ChatGPT, should be regarded as a personal information processor. In other words, developers of LLMs can indeed be the subject of claims for the exercise of the right to erasure under Article 47 of the PIPLin individual service contexts.

Unlike search engines, LLMs are proactive and non-intermediary in their data processing. It would be a futile struggle for developers of LLMs to defend their models in the same manner that Google did in the 2014 Google Spain v. González, claiming that their models are merely brokers of information, are passive, function as intermediaries, and do not process personal data.[27] Though this argument may have been debatable in the days of search engines,[28] the ECJ has also clarified that "it is the search engine operator which determines the purposes and means of that activity and thus of the processing of personal data that it itself carries out within the framework of that activity and which must, consequently, be regarded as the controller."[29] And such a defense is particularly weak in the context of the development of LLMs. LLMs do not crawl data like a search engine by finding, sorting, indexing and displaying the original data source on the search page, but rather extract the content of the data therein, and generate expressions that are consistent with the LLMs through continuous model training and selective feedback, most of the time without indexing to the original data source. Thus, it is evident that developers of LLMs not only process personal data but also autonomously determine the purposes and means of such data processing.

In sum, when an individual user uses a service that is directly open to the public individual by an LLM developer, such as OpenAI, the LLM developer is a processor of personal information, and is thus the request subject for the right to erasure of personal information.

III. The Challenges of Exercising the Right to Erasure in Large Language Models

The era of Generative AI has only just begun, and the research on technical development of LLMs is in full swing. The paradigm shift from search engine and internet era to LLMs and Generative AI era has brought changes in the sophisticated technical environment, thus, we must reexamine the feasibility of the right to erasure in the context of new emerging technologies. This section adopts a more granular approach, focusing on whether the right can be exercised effectively in the LLMs. It

[27] Opinion of Advocate General Jääskinen, Google Spain SL v. Agencia Española de Protección de Datos (May 13, 2014) (Case C-131/12), http://curia.europa.eu/juris/document/document.jsf?text=&docid=138782&doclang=EN [http://perma.cc/Y7C5-65WB].

[28] *Id*.

[29] Case C-131/12, Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, ECLI:EU:C:2014:317 (May 13, 2014).

concentrates on the technical feasibility of exercising the right to erasure, highlighting the challenges posed by the unique characteristics of LLMs compared to search engines.

A. Opacity and Barriers to Access

The opacity of the LLM will have a direct impact on the information accessibility, which is crucial to the right to erasure. Right to erasure is a special right of claim, realized through a claim to the personal information processor. Individuals can only exercise this right once they recognize that their personal information meets the prescribed conditions for erasure and that the data has not yet been deleted by the information processor.[30]

From the perspective of user's actual experience, it is evident that LLMs operate with a degree of opacity and present barriers to access. For instance, when individuals search for their personal information on a search engine, they can simply query relevant keywords or type their own names. Regardless of how search engines have become more sophisticated as a result of technology iterations, in essence, the information is still organized in an indexed format and user's access and search remains relatively convenient.[31] In contrast, it is not that simple to look up information related to individuals in an LLM. On the user end, i.e., the output end of the model, it is difficult for users to find out what personal information has already been collected and trained by directly asking the LLM questions. Firstly, the output of the LLM based on algorithmic predictions is inherently random and selective, and the resulting output is often uncontrolled and unpredictable, which increases the difficulty of reverse access from the LLM's output back to its input.[32] Secondly, although there is a possibility that the LLM may show "cracks" after repeated querying and interrogation and export personal information, the probability of this method is too low to make it a viable, generic and stable access method.[33] Thirdly, even if the afore-mentioned method were feasible, the prompting model does not guarantee that it will output the complete set of personal information regarding the specific user stored in its model for full inspection, which greatly impairs the effectiveness of the reverse access method.[34] Therefore, reverse-checking at the user's end to access the full range of personal information used by an LLM during

---

[30] Wang, *supra* note 13, at 40.
[31] Dawen Zhang et al., *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, AI & ETHICS 7 (Sept. 10, 2024), https://doi.org/10.1007/s43681-024-00573-9.
[32] Busra Bilsin, *Navigating EU Data Protection Law: The Challenge of the Right to Be Forgotten in Al-Driven Text Producers* 31 (Stanford-Vienna Transatlantic Tech. L. Forum EU Law Working Papers, Paper No. 86, 2024), https://law.stanford.edu/wp-content/uploads/2024/02/EU-Law-WP-86-Bilsin.pdf.
[33] Yang Liu et al., *Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models' Alignment*, ARXIV 15–16 (Aug. 9, 2023), arXiv:2308.05374v2 [cs.AI].
[34] Zhang et al., *supra* note 31, at 9.

model training and to target that information to a specific individual is not currently a suitable and practical access solution.

So, is it feasible to switch gears and go straight to the source by examining the original training datasets as well as the supplemental datasets to determine if user's personal information is being collected? The truth is that the LLM is more capable of creating a black box problem than past internet technologies with a more complex and user-unfriendly technical architecture, which makes it more confidential. And LLM developers are subjectively more inclined to hide their training databases and algorithmic techniques for the sake of trade secrets, and possibly also for protection from public scrutiny.[35] In addition, the collection of user inputs by LLMs will also make transparency disclosure of the training databases more complicated and problematic compared to search engines since there are always continuous random updates on supplementary datasets.[36] Thus, there is also low feasibility of directly examining the training database in terms of the real-world business model and LLM ecosystem.

Consequently, the opacity feature of the LLM is distinctive and fundamental, weakening information accessibility, building barriers to user access, and thus leading to the failure of meeting the prerequisite for the claim of the right to erasure. In other words, the transparency problem of the LLM makes it less likely for individuals to recognize or locate eligible personal information that they wish to erase, let alone to claim the right to the personal information processor to erase it so as to exercise the right to erasure.

B. Operational Autonomy and Erasure Dilemma

To exercise right to erasure of personal information, the next challenge would be "how to erase it" once personal information can be successfully accessed and located. Pre-training databases have thus received a considerable amount of attention as the basis for building foundational models and as a collection of stored personal information. It is further concerned that whether the erasure of personal information from LLMs can be achieved effectively by erasing personal information in the pre-training databases.

Compared to a traditional database, which is an organized collection of structured information or data, a pre-training database—despite being referred to as a database—has fundamentally different functions and

[35] *Ghost in the Machine: Addressing the Consumer Harms of Generative AI*, NORWEGIAN CONSUMER COUNCIL 20 (June 2023), https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf; Bilsin, *supra* note 32, at 30.

[36] Bilsin, *supra* note 32, at 31.

objectives, the role and structure have evolved to reflect the distinct demands of training models. The traditional database serves as the operational foundation and support for the product, providing access and query channels, and improving the efficiency of processing and data querying.[37] The pre-training database, in contrast, focuses on the concept of "pre-training", aimed at equipping an LLM with the ability to acquire language and predict words, which is a preliminary and fundamental step in the development of a capable language model.[38] Some scholars, based on their past understanding of the characteristics of the database itself, have pointed out that personal information does not exist in an isolated form at a single point in the database. Instead, it exists in multiple forms in different locations in multiple databases, is stored for a long period of time, or is backed up to prevent system errors and maintain system stability.[39] On this basis, although there are issues such as an overly broad scope of erasure, a threat to database consistency, and excessive erasure costs, erasing personal information from the database remains the preferred and viable option.[40] These understandings of the database are not problematic, and the concerns about deletion are realistic, but they did not recognize the actual role of the pre-training database in the overall model training process: this database is not the same as commonly perceived.

To further clarify, there is a disconnect between the pre-training database and the actual operation of the LLM, as the model functions autonomously once it obtains the capability to predict words based on training with large amount of data. When asking the LLM questions, it will not access the pre-training database through the index and answer after querying in it; rather it generates the answer based on its own linguistic ability. In other words, for the LLM, the pre-training database has already completed its mission and retired, and its subsequent operation has no connection with the pre-training database.

The new challenges of erasure have therefore been introduced. Past approaches that either physically erase, including expropriation and overwriting, destruction and elimination, or place the to-be-erased data in a "garbage offset" space and "mark it for erasure" so that its content is "not visible", "not retrieved", and "not linked"—are no more valuable reference for LLMs.[41] LLMs do not continuously access their pre-training databases

---

[37] *Database*, WIKIPEDIA (Sept. 2024), https://en.wikipedia.org/wiki/Database.

[38] Shayne Longpre et al., *A Pretrainer's Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity*, ARXIV 3–4 (Nov. 13, 2023), arXiv:2305.13169v2[cs.CL].

[39] Eduard Fosch-Villaronga et al., *Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten*, 34(2) COMPUT. LAW SEC. REV. 304, 308–309 (2018).

[40] *Id.*

[41] Peter Fruhwirt et al., *Using Internal MySQL/InnoDBB-Tree Index Navigation for Datahiding*, 462 IFIP ADVANCES INFO. COMMUNICATION TECH. 179, 183 (2015); Zhai Kai (翟凯), *Lun Rengong Zhineng Lingyu Bei Yiwang Quan de Baohu: Kunju Yu Pobi* (论人工智能领域被遗忘权的保护：困局与破壁) [*On the Protection of the Right to be Forgotten in the Field of Artificial Intelligence: Stalemate and Barrier Breaking*], 36(5) FAXUE

to index information, but have learned language patterns that may include personal information. Therefore, deleting personal information physically or logically from the pre-training database does not guarantee the ceasing of generating outputs containing that information. There remains a likelihood that the model will generate content that includes such information.

The characteristics of pre-training of LLMs, along with the paradigm of their pre-training datasets, which differ from traditional databases, create a disconnect between the database and the model's operation, highlighting the operational autonomy of LLMs. This conveys a clear message: the conventional approach of locating and deleting original data is impractical in the LLMs. The law can still continue to mandate the deletion of personal information from databases, which could remain a valid protection method in some circumstances. However, its generic effectiveness will be weakened in practice. Practical considerations suggest a gradual shift in focus from the input ends to the operational mode of LLMs and the issues in the output ends. Specifically, there will be a need to require the "erasure" in the output end instead of the database, which will present a significant challenge for technical experts to find a way to delete just before the final result is generated.

The exercise of the right to erasure is thus caught in a "deletion dilemma", where the act of erasure per se could be achieved, but the erasure is not truly effective. If the deletion of personal information from a pre-training database does not prevent the data from being accessed, read, or used by processors or others, such deletion cannot be considered effective.[42] This then raises critical questions: which data should be deleted and how can effective deletion be achieved?

C. Retraining Dependence and Retraining Conundrum

As mentioned in the previous subsection, due to the autonomous operation of LLMs, simply identifying and deleting the original personal information does not effectively prevent information leakage at the output end. If we still seek to approach the effectiveness of erasure from the perspective of "erasing the personal information from the initial source", it is crucial to take another noteworthy characteristic of these models into account—their reliance on retraining.

LUNTAN (法学论坛) [LEGAL FORUM] 142, 145–147 (2021); Eduard Fosch-Villaronga et al., *supra* note 39.

[42] Since the PIPL does not define the concept of "erasure", this article understands the effective erasure criteria with reference to article 3.10 of the Information Security Technology-Personal Information Security Specification (GB/T35273-2020), defining as "the act of removing personal information from systems involved in the fulfilment of day-to-day business functions so that it remains unavailable for retrieval or access". Although the GB/T35273-2020 is only a national standard document, it is of some value as a reference since the PIPL does not specify the exact meaning of "erasure". To understand effective erasure, see CHENG & WANG, *supra* note 23, at 182.

One key distinction between LLMs and data controllers like search engines in terms of providing information is that the updates of information in LLMs rely on the data from their most recent training. To keep abreast with the times, besides the information extracted from the pre-training databases, LLMs achieve their needs to the up-to-date information through accessing the internet and continually training on new, updated and supplemented datasets. In contrast, search engines can access new information in a more real-time manner by simply indexing it. In the year that ChatGPT-3.5 has been open publicly for service, it was noticed that the ChatGPT sometimes responded by stating that the collection of information for the training database ends in September 2021, and that it cannot provide information beyond that date. Even now that LLMs like ChatGPT have gradually started to be able to update relatively new information,[43] they still crawl the data first, train accordingly and then equip the capability of outputting the new content, rather than indexing the links directly and updating them instantly like search engines.

As a result, the changes to the content of the training database will not be reflected directly in the outputs unless the LLM is retrained on the updated database. It can then be inferred that the content related to personal information will not be ceased outputting unless the LLM is retrained on the database that has deleted relevant personal information. In other words, if aiming to solve the personal information outputting problem from the training database side, both the removal of personal information in the original databases and the subsequent retraining to the model are two essential elements due to LLM's reliance on retraining.

However, retraining on the databases that have erased relevant personal information is more complicated and demanding than adding information to train supplementarily. Full model retraining, i.e. re-creation, is one of the representative retraining methods, which was initially proposed as a simple and straightforward approach. The full model retraining refers to the periodic retraining or renewing of the LLM using a "training database that has deleted relevant personal information" so that it no longer learns the relevant content, thus stop generating content related to deleted information.[44] However, if every deletion can trigger a comprehensive retraining of the whole model, this process would be not only costly but also disruptive with the normal function. Accordingly, it would not only places an unreasonable burden on the LLM developers, but also stifles the development of technology innovation in AI. Several other

---

[43] Antoinette Radford & Zoe Kleinman, *ChatGPT can Now Access up to Date Information*, BBC NEWS (Sept. 28, 2023), https://www.bbc.com/news/technology-66940771.

[44] Debbie Reynolds, *Strategies for Managing Data Deletion and the Right to be Forgotten with LLMs*, LINKEDIN (June 1, 2024), https://www.linkedin.com/pulse/data-privacy-age-artificial-intelligence-ai-large-models-reynolds-shqqc/.

solutions to cease the output of personal information are also proposed by machine unlearning scholars, represented by SISA training, Differential Privacy, etc..[45] While these solutions may be preferable than the complete re-creation, it has been noted by some scholars that both still require retraining the entire model, which is extremely expensive and time-consuming given the size of the LLM, and potentially affects the performance of the entire model.[46]

While it remains to be seen whether machine unlearning experts will find a suitable erasure solution and retraining method, it is not the focusing argument of this paper. In contrast, the paper clarifies that if one relies on erasure of personal information from the initial database to achieve erasure effects, retraining is currently an inevitable path for the LLM to cease generating deleted personal information. However, a number of relevant challenges to erasure and training, including cost issue, technical feasibility issue, follow, leading to the troubling retraining conundrum. Whether the effectiveness of erasure can be fully achieved remains a question mark.

D. Uncontrollability of Hallucinations and Absence of Objects

Another essential and indispensable factor to consider when exercising the right to erasure is the identification of the object of the right, as the exercise of any right necessarily requires the existence of a corresponding object. However, the exercise of the right to erasure is troubled by the unique and unexpected feature of the LLM, the phenomenon of "hallucination", which leads to the absence of object.

Hallucination refers to the phenomenon wherein LLMs generate false or misleading information and present as facts. The models may generate content that does not align with the user's prompt, contradicts previously generated responses, or is inconsistent with pre-established knowledge of the world.[47] It has already been found that LLM may not be as smart as we thought it would be, they may be trapped by the tricky question "9.11 and 9.9, which number is bigger?",[48] may generate misinformation and

---

[45] Lucas Bourtoule et al., *Machine Unlearning*, 42 IEEE SYMP. SECUR. & PRIV. 141, 141–143 (2021); Da Yu et al., *Differentially Private Fine-tuning of Language Models*., ARXIV (July 19, 2022), arXiv:2110.06500.

[46] Jiaao Chen & Diyi Yang, *Unlearn What You Want to Forget: Efficient Unlearning for LLMs*, *in* PROCEEDINGS OF THE 2023 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING 12041, 12043 (Houda Bouamor et al. eds, 2023); Joel Jang et al., *Knowledge Unlearning for Mitigating Privacy Risks in Language Models*, *in* 1 PROCEEDINGS OF THE 61ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS 14389, 14389–392 (Anna Rogers et al. eds, 2023).

[47] Joshua Maynez, *On Faithfulness and Factuality in Abstractive Summarization*, *in* PROCEEDINGS OF THE 58TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS 1906 (Dan Jurafsky et al. eds, 2020); *Hallucination (artificial intelligence)*, WIKIPEDIA (Aug. 2024) https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence); Yue Zhang et al., *Siren's Song in the Ocean: A Survey on Hallucination in Large Language Models*, ARXIV 3–4 (Sept. 3, 2023), arXiv:2309.01219 [cs.CL].

[48] *Why 9.11 is larger than 9.9.....incredible*, OPENAI COMMUNITY (July 2024), https://community.openai.com/t/why-9-11-is-larger-than-9-9-incredible/869824.

disinformation, saying that "the mother of Afonso II was Queen Urraca of Castile instead of Dulce Berenguer of Barcelona".[49] Consequently, LLMs may also have the possibility of outputting completely incorrect and fabricated content related to personal information. Accordingly, various speculations and theories have been proposed to explain the causes of hallucinations in LLMs. Some attribute the problem to a lack of relevant knowledge or internalization of incorrect information. Others argue that hallucinations may arise from flawed alignment processes, low coverage of alignment examples, inherent ambiguity in the supervision data,[50] or suggest that hallucinations can happen when LLMs adopt potentially risky generation strategies and problematic training mechanisms.[51] However, what underlies the problem of hallucination is an extreme sense of loss of control. Even if there isn't incorrect or relevant information within the training dataset, large models still produce erroneous or fabricated outputs, and developers of these models are currently unable to keep the generation of such hallucination fully under control. While safeguards, sophisticated prompts, and innovative methods developed by technical experts help mitigate hallucination to some extent, its occurrence can only be minimized—not entirely eliminated.[52]

Thus, when focusing attention on personal information and right to erasure, LLMs are fully capable of generating incorrect or fabricated personal information due to hallucinations, supporting by experimental evidence.[53] What's worse, if there isn't problematic information in the training dataset and the hallucination happens simply due to some mysterious reasons, how can an individual exercise the right to erasure? In that case, incorrect personal information is generated and exported without the original, collected personal information being present. Even if the individual has access to the training database, there is nothing that can be done about it. That is, the object we are trying to erase is missing, the object of exercising the right is missing, and the developers of the LLMs are currently unable to solve and completely eliminate this problem, leaving it an open problem and an ongoing area. In light of the issue of

---

[49] Dipto Barman et al., *The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination*, 16 MACH. LEARNING WITH APPLICATIONS 1, 1 (2024); Zhang et al., *supra* note 47, at 4.

[50] Shen Zheng et al., *Why does chatgpt fall short in answering questions faithfully?*, ARXIV (Dec. 3, 2023) arXiv:2304.10513; Hannah Rashkin et al., *Increasing faithfulness in knowledge-grounded dialogue with controllable features*, *in* 1 PROCEEDINGS OF THE 59TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS AND THE 11TH INTERNATIONAL JOINT CONFERENCE ON NATURAL LANGUAGE PROCESSING 704 (Zong et al. eds., 2021); Zhang et al., *supra* note 47, at 9–10.

[51] Emily M Bender et al., *On the dangers of stochastic parrots: Can language models be too big?*, *in* PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 610–623, (2021); Zhang et al., *supra* note 47, at 9–10.

[52] Zhang et al., *supra* note 31, at 6; *Our approach to Al safety*, OPENAI (April 5, 2023), https://openai.com/index/our-approach-to-ai-safety.

[53] Zhang et al., *supra* note 31, at 12.

hallucinations and the consequential absence of object, the right to erasure of personal information runs into the problem of not adapting well to the Generative AI age.

IV. Legal Responses to the Exercise of the Right to Erasure in the Context of the Large Language Model

The complexity and unpredictability of Generative AI technologies introduce new obstacles to the protection and exercise of individual rights to personal information. The special natures of LLMs and the issues surrounding the technical feasibility of applying the right to erasure within them epitomize a representative picture. In the context of the unique technological characteristics of LLMs, is there a need for a new understanding and corresponding revision to the right to erasure? While the technical experts are perseveringly exploring better technical solutions, this section aims to propose legal solutions that are more compatible with the characteristics of the LLMs, offering new understanding of the right to erasure from a legal perspective. In a broader scope, this section also aims to uphold and implement the right to erasure while ensuring adequate space for the innovative development of Generative AI.

A. Introduce Third-party Regulation

Transparency is always the theme when it comes to the Generative AI that even sounds a little bit cliché these days. The transparency problem of the database of LLMs not only makes it difficult for individuals to figure out whether information is being collected and processed, but also makes it sophisticated for individuals to locate personal information in the training database that needs to be deleted. The lack of accessibility thus sets up a front barrier and threshold for individuals to exercise their rights. In such a situation of unequal position and asymmetric information between the two parties, to better protect the exercise of individual rights, the introduction of a third party to regulate can, to a certain extent, bridge the gap.

Database auditing and control have been widely implemented in practice, with regulations as well as internal security policies requiring the provision of transparency mechanisms, compliance management with granular auditing of database operations, and real-time warnings of possible risky behavior in the database.[54] As for the black box problem of algorithms, which is the infrastructure of AI development, there have long been many scholars who have proposed solutions such as filing, disclosure and supervision. These solutions have already been reflected in practice.

[54] Ronald Caldwell, *What is a Database Audit?*, Liquid Web (March 2, 2023), https://www.liquidweb.com/blog/database-audit/; *Shuju Ku Shenji* (数据库审计) [*Databases Audit*], Baidu Baike (Feb. 2023), https://baike.baidu.com/item/数据库审计/7882064?fr=ge_ala.

For example, according to the provisions in the Guiding Opinion on Regulating the Asset Management Business of Financial Institutions, financial institutions are required to report the main parameters of the AI model to the financial supervisory and regulatory authorities, and the guidance also emphasizes the importance of artificial intervention.[55]

The underlying similarity between the two aforementioned options that are already in practice and the regulation of training databases for LLMs can provide inspiring insights for seeking to regulate training databases. This paper argues that due to the challenges individuals face in accessing the database to exercise the right to access, it would be more effective to designate the personal information protection department under the PIPL as a third-party supervisor, granting it the authority to regularly audit and supervise the content related to personal information in the training database and to manage unified access and screening. Thus introducing human and third-party intervention into the LLM's process. The supervisory authority shall audit pre-training databases and newly added supplementary databases, specifically screening for three categories of personal information: sensitive personal information, personal information collected beyond the purpose of processing and personal information processed in violation of the law. The authority should supervise and ensure that developers have removed such information. However, the frequency of screening should be controlled, with efforts made to minimize duplicate audits of the same dataset without significant updates, so that the database auditing will not be destructive to the LLM ecosystem.

This approach can not only preliminarily filter the data and narrow the scope of personal information that users need to request so as to mitigate the opacity issue, enhance user trust in LLMs and facilitate the exercise of individual rights, but also prevent the heavy administrative burden of processing numerous user requests for access to the training database on developers, balancing the protection of personal information with the development of LLMs.

B. Reinterpret "Erasure"

The operational autonomy of the LLM and the corresponding possible ways of erasure bring new challenges for "erasure", in addition to the concerns about the feasibility of erasure and the efficacy of erasure from the technical perspective, we might consider reinterpreting the concept of

[55] Guanyu Guifan Jinrong Jigoou Zichan Guanli Yewu de Zhidao Yijian (关于规范金融机构资产管理业务的指导意见) [Guiding Opinion on Regulating the Asset Management Business of Financial Institutions] (promulgated by the Bank of China, China Banking and Insurance Regulatory Commission, China Securities Regulatory Commission, State Administration of Foreign Exchange, Apr. 27, 2018, effective Apr. 27, 2018), art. 23 (Chinalawinfo).

"erasure" in the context of the Generative AI era to face the challenge from a legal perspective.

According to paragraph 2 of Article 47 of the PIPL, "... [w]here it is difficult to realize the deletion of personal information technically, the personal information processor shall cease the processing of personal information other than storing and taking necessary security protection measures for such information.", it can be seen that "deletion of personal information is technically difficult to realize" could be an exemption for developers of LLMs. So, what kind of situation would be considered "technically difficult to realize"? Cheng Xiao has cited an example, "in the case of cloud storage, personal information cannot be deleted unless the user's data is completely emptied, this is the case of technically difficult to achieve." That is to say, "technically difficult to realize" applies when personal information cannot be deleted by existing technology by all means, or can only be deleted by paying an unreasonable cost under the existing technological conditions.[56] Accordingly, this paper suggests that "technically infeasible" as an exemption should be interpreted with abstinence rather than broadness, particularly as new technologies related to machine unlearning are emerging and evolving. A broad interpretation would diminish the effectiveness of Article 47 and weaken the right to erasure of personal information.

Therefore, this paper contends that the technical challenges discussed in section 3 cannot yet be classified as "technically infeasible" and obtain the exemption of the right to erasure, since various developed and applied technical solutions driven by technology advancements are demonstrating effective methods of erasure and reducing the cost of deletion. There remains significant potential to explore erasure methods within Generative AI and LLMs. Most importantly, we should not proactively and easily narrow the scope and weaken the effectiveness of the right.

However, the presentation of these technical evolution reminds us to revisit the concept of "erasure". The idea of solving the deletion problem from the input side has encountered the opacity and retraining problems of LLMs. At the same time, the nature of LLMs' autonomous operation, predicting words rather than retrieving stored data, also prompts us to shift our attention to the output side of the model. In this context, prioritizing the effect of erasure over the act of erasure itself aligns more closely with the operational characteristics of LLMs. Practical trends have already responded to this a step forward, with new techniques favoring approaches like model editing, guardrails, unlearning layers, negating and so on to achieve erasure, rather than being limited to the traditional "physical

[56] CHENG & WANG, *supra* note 23, at 181.

erasure" in the LLM.[57] Therefore, this paper suggests revisiting the concept of "erasure" and endorsing the new concept and understanding of "logical erasure", i.e., instead of aiming for the complete elimination of data in the physical state and the complete removal of data from the storage medium, the effect of data erasure should be pursued based on the technological means to achieve a near-elimination effect that ceases personal information output. So as to keep abreast with the current technological development trend, to spare space for the advancements of machine unlearning technology, and to balance the protection of the right to erasure of personal information with minimizing interference in technological progress.[58]

In sum, although the development of erasure technologies faces technical challenges and the erasure within LLMs encounters significant difficulties, the erasure within LLMs should not be interpreted as "technically infeasible" exemption under Article 47 of the PIPL. On the contrary, there is a need to re-evaluate the concept of "erasure" to encompass broader erasure methods to align with practical demands. Therefore, to meet the requirement of "erasure" in the PIPL, physical erasure is the principle, and logical erasure is the exception, which both fall into the comprehensive concept of erasure. Specifically, when considering how erasure should be carried out, we evaluate whether physical erasure can take place in conjunction with technical and other practical factors first. When physical erasure is technically difficult to achieve, logical erasure should be considered as an alternative, aiming to achieve the effect of rendering personal information inaccessible. Both erasure methods satisfy the requirements for erasure under the PIPL and will not automatically trigger the exemption of "technically infeasible erasure" merely because physical erasure fails. For subjects such as LLMs, which are inherently and technically not suitable for physical erasure, logical erasure should be considered an acceptable alternative. As long as "erasure" is achieved in effect, i.e., the personal information no longer has the possibility of being exported and the personal information will henceforth be excluded from the data processing process, whether or not the personal information is actually and completely eliminated from the database should not be a troubling problem for us.

---

[57] Ronen Eldan & Mark Russinovich, *Who's Harry Potter? Approximate Unlearning in LLMs*, ARXIV (4 Oct 2023), arXiv:2310.02238 [cs.CL]; Jang et al., *supra* note 46, at 8–10; Zhang et al., *supra* note 31, at 10–11; Liu et al., *supra* note 33, at 16; Chen & Yang, *supra* note 46, at 12043–12044.

[58] This new understanding of "logical erasure" also aligns with the provisions and tendencies outlined in Article 3.10 of the Information Security Technology-Personal Information Security Specification (GB/T 35273-2020), a national standard document of significant reference value.

C. Adhere to the Purpose Limitation Principle

The continuous and up-to-date operation of LLMs cannot be achieved without constant training, with databases regularly updated and expanded, new information is constantly processed. While technical experts in machine unlearning are focusing on finding feasible solutions of retraining within LLMs, the law should reinforce adherence to the purpose limitation principle and ensure that it underpins all personal information processing activities.

This proactive approach would ease the burden of subsequent oversight and regulation from the outset. It would also fundamentally minimize placing pressure on individuals to protect their own personal information through exercising the right to erasure and lessen the scenarios in which the right to erasure needs to be applied. Adhering to the purpose limitation principle throughout the process, alongside the exercise of the right to erasure, will contribute to the protection of personal information while alleviating the burden and feasibility issues associated with exercising such right.

The purpose limitation principle is contained in the provisions of Article 6 of the PIPL, which states:

"Personal information processing shall be for a clear and reasonable purpose, directly related to the processing purpose and in a manner that has the minimum impact on the rights and interests of individuals. Collection of personal information shall be limited to the minimum scope necessary for achieving the processing purpose and shall not be excessive."

The provision specifies two important stages covered by the purpose limitation principle: first, the information collection stage, where the collection of personal information shall be limited to the "minimum scope necessary to achieve the processing purpose", and data shall be collected in a restrained manner, subject to the limitations of an explicit, reasonable, and specified purpose;[59] and second, the information processing stage, where the processing of personal information shall be "directly related to the processing purpose", and no further processing shall be carried out beyond the original purpose of the processing.

More specifically, in the information collection phase, LLM developers should adhere to the purpose limitation principle when initially building their pre-training databases, and proactively implement data governance and data management under the guidance of this principle. For example, algorithms should avoid blind crawling of data as much as possible, and the collected data should be analyzed and filtered to refrain

[59] *See* GDPR, art. 5.

from excessive personal information collection beyond the "minimum scope necessary to achieve the processing purpose". For new supplementary datasets, developers of LLMs should also comply with this principle. To enforce this principle practically and effectively, this paper proposes that the LLM developers are encouraged to actively develop data categorization and filtering systems. Though establishing such a system may introduce initial costs for developers in the early stages of model development, it is worth the investment based on current legal compliance requirements and the practical enforcement of individuals' right to erasure. In the long run, a data categorization and filtering system can shoulder the burden of increasingly rigorous compliance demands, lighten the load of user requests for data erasure, avoid complex deletion processes, and lower the costs associated with information erasure—ultimately promoting the sustainable and healthy operation of the entire LLM ecosystem.[60]

In the information processing phase, LLM developers should also consistently adhere to the purpose limitation principle and refraining from processing that goes beyond the original processing purpose, limiting uncontrolled processing in the algorithmic black box. Filing requirement on the algorithmic side as well as compliance regulation of overall data may be possible solutions to limiting excessive processing. In addition, the LLM industry should be encouraged to adhere to the purpose limitation principle through industry self-regulation. For developers, adhering to this principle will set up a healthier industry standard without affecting the operation and development of the LLM, and strengthen the user trust in the LLMs in this personal information-sensitive age, which would be conducive to the development of the industry.

To conclude, the purpose limitation principle is a principle that needs to be adhered to throughout the data processing process, which intends to reduce the involvement of personal information in the data processing process of LLMs from an ex-ante point of view. Adhering to purpose limitation principle will mitigate the subsequent compliance and unlearning technique development pressure on developers, and, more importantly, will reduce the demand for individuals to exercise their right to erasure, which is also a solution to the feasibility issue of the right to erasure in the LLM from another perspective.

D. Leave Adequate Space for Technological Development

The issue of hallucinations renders the right to erasure practically inapplicable. When personal information is leaked due to the hallucination, it is hard for individuals to remedy through the right to erasure, as the

[60] Reynolds, *supra* note 44.

exercise of this right lacks a clear object in such cases. Hence law makers seek to fill up the corresponding gaps and solve the hallucination problem to protect personal information. In parallel, the hallucination remains one of the most important problems plaguing LLM developers, LLM developers and deep learning experts are still working to solve it. In this respect, legal regulation aligns with the interests pursued by LLM developers, who are motivated to proactively address the hallucination problem.[61] After all, users of LLMs will not support a model that outputs false or non-existing information, and competitors in the LLM market will not be lenient to a less competitive model that is trapped in the hallucination problem, business is business.

Against this background, this paper argues that since the technical attributes of the hallucination problem are so salient and impossible to ignore as mentioned in section 3 part D, increasing investment in technological research and awaiting breakthrough advancements to address the issue rather than exhausting efforts to devise legal regulatory solutions may be a more prudent approach.

In this case, what the law can do is to leave developers time and space to iterate on the hallucination problem, refrain from over-regulation. Promoting the industry norm and standard would also be helpful, technical communications within the industry should be strengthened and corresponding technical specifications should be formed, so that the whole industry can have one consistent goal in solving the hallucination problem and make efforts to this end.

E. Interim Conclusion

One thing to admit, several pitfalls remain in the legal solutions proposed in the previous subsection. For instance, the introduction of third-party regulation proposes a general filtering of the database on a regular basis, but for specific scenarios such as the consent has been withdrawn by the individual, or the processing purpose has been achieved, it is necessary for the individual concerned to identify and exercise their own rights, and at this point the opacity of the LLM still makes it difficult for individuals to access their personal information. However, this paper believes that the solution to the problem requires the efforts of all parties from different realms. Legal solutions are not the only response. Technological development, social norms, and cognitive adjustments, can also help us better build and solidify responses to the current challenges.

The legal approaches proposed in this paper are clear. It recognizes the salient nature of LLMs as fundamentally different from previous

[61] Zhang et al., *supra* note 52.

internet-era products, and advocates for a reinterpretation of the right to erasure's practical application, taking into account the unique characteristics of LLMs, to ensure the effective application of the right to erasure in the era of LLMs. At the same time, the legal responses in this paper also aim to maximize the protection of individual information rights without imposing excessive interference in the normal Generative AI progression, thereby allowing time and space for technology to evolve in a compliant manner.

V. CONCLUSION

The rapid pace of technological innovation has ushered us, with unstoppable momentum, from the Internet era into the AI era. The age of AI is so dazzling and fascinating that the plots in those science-fiction films are already becoming a reality, and one day we may have our own Jarvis just like Iron Man. However, down to the earth, personal information protection in AI era is still the public's core concern. Though the emergence of LLM, the representative product of Generative AI, is changing our lives; what remains unchanged is that the LLM collects users' personal information just as its fellow data controllers, and the need for users to control, to erase their personal information persists. The right to erasure is still a necessity in Generative AI era.

However, this right will now be exercised in a different context. The applicability of the right to erasure granted to individuals in China's PIPL enacted in 2021 in the new era thus requires further review. This paper examines the right to erasure of personal information as a representative example of personal information rights and approaches this analysis through the lens of LLMs—a hallmark of Generative AI. By adopting a more practice-oriented perspective, it assesses the applicability of the right to erasure in the Generative AI age, in the LLM.

This paper first clarifies that the processing of personal information is involved in the LLMs throughout the whole process, and developers can be classified as personal information processors in the individual-user scenarios, thereby making them subject to the PIPL.

The paper then sets out to look into the LLM's characteristics and the corresponding challenges that it will pose for the exercise of the right to erasure. First, this paper points out the opacity problem in the LLM's database, and argues that, unlike search engines, it is difficult for users to access the LLM's database in a convenient way, and therefore they are unable to realize that their personal information has been collected and which personal information has been collected, which fail to meet the prerequisite of exercising right to erasure; Second, this paper notes the autonomous operation of the LLM, i.e., the LLM functions autonomously through word prediction following the completion of pre-training without

indexing the database, and therefore preventing LLM from generating outputs related to personal information simply by deleting the contents in the database does not work, and thus leads to the challenge of the effective erasure; Third, the paper further discusses the retraining issue, noting that the updates and changes in the outputs are based on the database retraining and supplemental training, however, the feasibility of retraining is limited, it is not only costly but also destructive to the LLM ecosystem, the effectiveness of exercising the right to erasure is thus doubted; Fourth, this paper also explores the unique hallucination problem of the LLM, including the fact that non-existent personal information in the database will be falsely output like fake news, and states that the hallucination problem may lead to the absence of the object of the right to erasure, making the exercise of the right to erasure impossible.

Recognizing the challenges LLMs pose for the exercise of the right to erasure, this paper seeks to provide legal responses. First, this paper proposes that third-party supervision should be introduced to regularly audit the personal-information-related contents in databases to filter and screen personal information in response to the transparency problem, to compensate to a certain extent for the accessibility issue, which is the prerequisite for the right to erasure; Second, this paper argues that the concept of "erasure" should be reinterpreted to include the logical erasure as an acceptable alternative, which focuses on achieving the effect of erasure rather than insisting on absolute physical elimination in the context of LLMs, and thus encompasses broader erasure methods to align with practical demands and enhances the feasibility of right to erasure in LLMs; Third, this paper reiterates the purpose limitation principle and advocates for its consistent application throughout all stages of personal information processing, to mitigate subsequent regulatory and compliance pressures from the outset, and to further reduce the scenarios in which individuals need to claim right to erasure for remedy, alleviating the right's feasibility issue in the LLM from another perspective. Finally, this paper contends that regarding technical problems such as hallucination, providing time and space for the development of technological solutions is a wiser approach than struggling to come up with specific legal solution, and that legal approaches should prioritize guidance and avoid excessive regulation.

In conclusion, many challenges remain to be solved in exercising the right to erasure within LLMs. This paper offers legal responses tailored to the unique characteristics of LLMs, aiming to provide a fresh perspective on exercising the right to erasure effectively in the Generative AI age and to prevent this right from becoming an unattainable ideal. This paper also expects that the law, while safeguarding individuals' rights over their

personal information, can support technological innovation and the sustainable development of AI.