COMPUTER LAW & SECURITY REVIEW 51 (2023) 105885

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Algorithms that forget: Machine unlearning and the right to erasure

Bjørn Aslak Juliussen[a,*], *Jon Petter Rui*[b,c], *Dag Johansen*[a]

[a] *Department of Computer Science, UiT The Arctic University of Norway, PO Box 6050 Langnes, N- 9037 Tromsø, Norway*

[b] *Faculty of Law, University of Bergen, Bergen, Norway*

[c] *Faculty of Law, UiT The Arctic University of Norway, Tromsø, Norway*

ARTICLE INFO

Keywords:
Data protection law
General data protection regulation
The right to erasure
The right to be forgotten
Machine learning
Machine unlearning
Privacy

ABSTRACT

Article 17 of the General Data Protection Regulation (GDPR) contains a right for the data subject to obtain the erasure of personal data. The right to erasure in the GDPR gives, however, little clear guidance on how controllers processing personal data should erase the personal data to meet the requirements set out in Article 17. Machine Learning (ML) models that have been trained on personal data are downstream derivatives of the personal data used in the training data set of the ML process. A characteristic of ML is the non-deterministic nature of the learning process. The non-deterministic nature of ML poses significant difficulties in determining whether the personal data in the training data set affects the internal weights and adjusted parameters of the ML model. As a result, invoking the right to erasure in ML and to erase personal data from a ML model is a challenging task.

This paper explores the complexities of enforcing and complying with the right to erasure in a ML context. It examines how novel developments in machine unlearning methods relate to Article 17 of the GDPR. Specifically, the paper delves into the intricacies of how personal data is processed in ML models and how the right to erasure could be implemented in such models. The paper also provides insights into how newly developed machine unlearning techniques could be applied to make ML models more GDPR compliant. The research aims to provide a functional understanding and contribute to a better comprehension of the applied challenges associated with the right to erasure in ML.

* Corresponding author.
E-mail address: bjorn.a.juliussen@uit.no (B.A. Juliussen).

https://doi.org/10.1016/j.clsr.2023.105885

1. Introduction

The General Data Protection Regulation (GDPR)[1] provides a comprehensive framework for data protection in the European Union (EU). Article 17 of the GDPR outlines the right of the data subject to request the erasure of their personal data from the controller. The right in Article 17 is also referred to as the right to be forgotten. Article 17 contains a right for the data subject – with a corresponding obligation for the controller – to obtain the erasure of personal data. Article 17 holds the specific grounds where data subjects can invoke the right to erasure. Although Article 17 sets out the right to erasure of personal data, it does not give any specific technical guidance on how to fully erase personal data from a computer system.

Coinciding with the development of artificial intelligence (AI) and machine learning (ML) methods, the erasure of personal data has evolved from a previously straightforward task to a complex procedure because of the opaque and non-deterministic nature of ML and because ML models learn and develop from previous (personal) data. A ML model is a downstream derivative of the personal data in the training data set. A crucial question to discuss is to what extent full erasure of personal data would require alterations of the derivate of the personal data in the training data set, the ML model.

The practical application of the research question analyzed in the article can be illustrated as follows: Suppose that a lawyer uses a pre-trained transformer model, specifically a chatbot, as an aid in a hectic work environment. Inadvertently, the lawyer includes highly sensitive client information within the prompt of the chatbot and submits it. In such a situation, the lawyer would need assurance that the confidential client data is not integrated into the ML model and subsequently utilized in future chatbot interactions.

Another relevant everyday example of when machine unlearning approaches might be useful is related to employee offboarding. Suppose that an employee leaves a company. During the employee's time working there, several ML model has been personally tailored to the employee and contains personal data, including ML systems for customer relations, employee performance evaluation or ML systems for safety and risk management analysis. Under several national jurisdictions in Europe, personal data of employees needs to be erased when the employee leaves the company.[2] Machine unlearning approaches might be useful in such employee offboarding scenarios.

The right to erasure and the right to be forgotten in Article 17 of the GDPR is often the core motivation behind the development of amnesiac properties of ML algorithms, often referred to as machine unlearning.[3]

The overall aim of this article is to link the legal requirements for erasure in Article 17 of the GDPR with current methods and the technological feasibility of machine unlearning. Through such an interdisciplinary approach, the aim is to interpret the legal scope of the right to erasure under the GDPR. Moreover, the right to erasure of personal data should be interpreted with reference to the available technology, according to Article 17(2) and Recital 66 of the GDPR. It is therefore vital to analyze and conclude how the current state-of-the-art of machine unlearning makes the erasure of personal data in ML possible.

Before delving into the right to erasure and techniques of machine unlearning, it is necessary to present some examples of related works and existing literature. There are several works on machine unlearning, mostly from a technical perspective.[4] There are also several pieces of literature, where the core motivation behind the development of machine unlearning approaches is the right to erasure.[5] However, there are scarce examples in the literature where both the scope of the right to erasure and machine unlearning approaches are analyzed. Most of the technical approaches in the literature presuppose the right to erasure in the GDPR without any further legal analysis of the scope of the right. Hence, this article will attempt to both analyze the scope of the right to erasure in Article 17 GDPR and the technological feasibility of machine unlearning.

In order to explore the interface between the legal scope of the right to erasure and the potential technological feasibility of machine unlearning, it is necessary to first examine and resolve some key research questions.

The rest of this paper is organized as follows. First, Section 2 will analyze the legal status of ML models trained on personal data under the GDPR. The GDPR was adopted in 2016 and entered into force in the EU in 2018.[6] ML research and ML

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

[2] Deletion of personal data of employees when offboarding is a legal requirement under both the purpose limitation principle and the storing limitation principle in GDPR Article 5, and Article 17 (1) litra a) of the GDPR. The deletion requirement in an offboarding situation of an employee might have several exemptions.

[3] Thanh Tam Nguyen, Thanh Trung Huynh, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, Quoc Viet Hung Nguyen, 'A Survey of Machine Unlearning' (2022). ArXiv. Available: https://doi.org/10.48550/arxiv.2209.02299.

[4] Ayush K. Tarun, Vikram S. Chundawat, Murari Mandal, Mohan Kankanhalli, 'Fast Yet Effective Machine Unlearning' *IEEE Transactions on Neural Networks and Learning Systems* DOI: 10.1109/TNNLS.2023.3266233. accessed 15.07.2023; Vikram S. Chundawat, Ayush K. Tarun, Murari Mandal and Mohan Kankanhalli, 'Zero-Shot Machine Unlearning' *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2345-2354, 2023 DOI: 10.1109/TIFS.2023.3265506. accessed 15.07.2023.

[5] Bannihatti Kumar, Vinayshekhar, Rashmi Gangadharaiah, and Dan Roth. 'Privacy Adhering Machine Un-learning in NLP.' arXiv e-prints (2022): arXiv-2212; Kumar, Vinayshekhar Bannihatti, Rashmi Gangadharaiah, and Dan Roth. "Privacy adhering machine un-learning in nlp." arXiv preprint arXiv:2212.09573 (2022); Aloni Cohen, Adam Smith, Marika Swanberg, Prashant Nalini Vasudevan 'Control, Confidentiality, and the Right to be Forgotten.' arXiv preprint arXiv:2210.07876 (2022).

[6] The definition of personal data in GDPR Article 4(1) is similar as the previous definition under 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 [1995] OJ L 281/31 Article 2(1).

techniques have had a vast development since the regulation entered into force. Due to the development in ML, it is important to interpret the definition of personal data in Article 4(1) of the GDPR together with both case law from the Court of Justice of the European Union (CJEU), recommendations from the European Data Protection Board (EDPB) and decisions from various European data protection authorities. Through such an interpretation, it is possible to get the full picture of the legal status of ML models trained on personal data under the GDPR.

The full scope of the right to erasure under Article 17 needs to be interpreted in a ML context with the aim of exploring the extent and scope of the right in relation to ML. Section 3 of the paper will examine the right to erasure with regard to the state-of-the-art machine unlearning techniques to connect the scope of the right in Article 17 with current amnesiac ML methods.

The last section will conclude with some set of recommendations and potential for future research.

2. The legal status of ML models under the GDPR

2.1. Machine learning models

ML is a subset of AI and methods computers apply to make predictions and decisions based on data.[7] In traditional algorithmic-based programming, a programmer instructs the computer on how to compute a desired result. In ML the "machine" learns and recognises patterns from data to predict outcomes or decisions without being explicitly programmed to do so.

A ML model is a weighted function created and trained from the data in the ML process.[8] A ML process starts with a training data set and trains a ML algorithm on the training data. The finished trained and weighted function constitutes the ML model.

The relationship between the ML algorithms and the ML model can be explained using an everyday example. On the Explorer page on Instagram, the users get suggestions for personally tailored content. This Explorer page is created through the use of a ML algorithm similar to word embedding algorithms (Word2vec).[9] Alike as your dictionary in your text message application anticipates that it is likely that the word "you" will be the next word after the phrase "how are", the ML algorithm calculates the probability of the user being interested in specific suggested content on the Explorer page.

The input data in the ML training of the recommender-algorithm in the example above includes data on content that the user has previously liked, content that the followers have engaged with, how long the user has had his or her finger over the picture, metadata from the content and messages, the time, frequency and duration of activities, and other factors.[10] The results from the training are then ranked to suggest content that the user most likely is interested in. In this example, the recommender-algorithm constitutes the ML algorithm and the final recommender version trained on the personal data from the users constitutes the trained ML model. In the context of a recommender system offering personalized recommendations, machine unlearning might enable users to delete or adjust the parameters in the trained ML model. Such unlearning can enable the user to experience new suggestions and recommendations. New recommendations might be preferable for users who want to reset or redefine their social media preferences.[11]

2.2. Could the finally trained ML model be regarded as personal data?

2.2.1. Introduction

In order to conclude the question regarding the right to erasure under the GDPR in a ML context, it is important to clarify the legal status of a ML model trained on personal data under the GDPR.

The GDPR lays down rules relating to the protection of natural persons with regard to the processing of personal data, pursuant to Article 1(1) of the GDPR. A data subject only has the right to obtain the erasure of *personal* data under Article 17 of the GDPR. It is therefore important to examine whether the finished trained and weighted ML model could fulfill the definition of personal data in the regulation. The process of training the ML model described in the former section will fulfill the definition of processing under Article 4(2) of the GDPR.[12] Furthermore, the data used to train the ML model in the example above, including liked content and metadata, will satisfy the definition of personal data under Article 4(1) of the GDPR if it is linked to a specific identifiable natural person.

The crucial question is whether the finished trained ML model fulfills the definition of personal data in Article 4(1) GDPR.

2.2.2. The criterions in Article 4(1) GDPR

Article 4(1) of the GDPR 'personal data' has the following wording: "any information relating to an identified or identifiable

[7] See, Mahesh, Batta. Machine Learning Algorithms- A Review. International Journal of Science and Research (IJSR) 9 (2020)381-386.

[8] Microsoft Learn, 'What is a machine learning model?' Available: https://learn.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learning-modelaccessed 13.07.2023.

[9] Meta AI, powered by AI: Instagram's Explore recommender system. (2019). Available: https://ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/ accessed 13.07.2023.

[10] See, Meta, Privacy policy, available: https://privacycenter.instagram.com/policy/accessed 13.07.2023.

[11] There are also specific requirements for transparency in recommender systems in the Digital Services Act. See, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277/1 Article 27 (1).

[12] See, the definition of processing in Article 4(2) GDPR and Norwegian Data Protection Authority, 'Artificial Intelligence and privacy'. (2018). Page 18. Available: https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdfaccessed 13.07.2023.

natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

The direct identification of a natural person is a straightforward task when the identifier is linked to the data subject, for instance through a name, an identification number or the age or a description of a person. The potential status of a trained ML model as personal data under Article 4(1) would, however, be related to the indirect identification of a natural person under Article 4(1) of the GDPR.

The indirect identifiability criterion in Article 4(1) of the GDPR is elaborated in the non-binding Recital number 26 to the Regulation. According to Recital 26, account should be taken "of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly" to determine whether a natural person is identifiable. Furthermore, to ascertain whether means are reasonably likely to be used to identify a natural person, "account should be taken of all objective factors, such as the cost of and the amount of time required for identification, taking into consideration the available technology at the time of processing and technological developments".

For the remaining legal analysis, it is presupposed that all other criteria in the definition of personal data are fulfilled and that the questionable criterion in relation to a ML model – that is trained and weighted on personal data – is the indirectly identifiable criterion. It is, therefore, presupposed that the trained ML model satisfies the criterion of "any information",[13] and that this information is "relating to" a natural person that fulfills the definition of an "individual". The undecided questions are whether the trained ML model could be applied to indirectly identify an individual and whether the means necessary for such indirect identification are "reasonably likely to be used" by the controller or another person.

The first question under discussion relates to the indirect identifiability criterion and is whether a trained ML model could be applied to indirectly identify an individual natural person.

The identifiability criterion is explained in Article 29 Working Party's non-binding opinion on the concept of personal data.[14] According to the opinion from the working party, a natural person is identified when the person is distinguished from all other members of a group. A natural person is identifiable when the person has not been identified yet, but it is possible to do so.[15] However, it is important to note that this is a non-binding opinion under the previous data protection directive. Furthermore, the opinion has not been adopted by the European Data Protection Board (EDPB). The main focus in the remainder of the section will, thus, be on CJEU case law.[16]

In the Breyer case,[17] the CJEU clarified the understanding of the identifiability criterion related to whether identifiability is assessed from the perspective of the controller or another person or entity.[18]

The request for a preliminary ruling in the Breyer case concerned the interpretation of the definition of personal data under the repealed data protection directive.[19] The definition of personal data is the same under the GDPR as it was under the repealed directive. More specifically, the question under scrutiny in the CJEU concerned whether a dynamic IP address registered by an online media service provider when a natural person accessed the site constituted – with regard to the service provider – personal data under the directive, where only a third party had the information necessary to identify the natural person.

The Court held that the wording "indirectly" suggested that it was not necessary to fulfill the definition of personal data that the information alone allowed for identification.[20] The CJEU then concluded that it was not required for a piece of information to fulfill the definition of personal data that all the information enabling identification was in the hands of one person or entity.[21] It is therefore not necessary that the ML model alone allows for identification of a natural person. It is sufficient that the ML model combined with additional information allows for identification to fulfill the "indirectly" identifiability criterion.

Indirect identifiability from a trained ML model will vary depending on who attempts to identify an individual. The controller that has trained the ML model has, in general, more data to link the results or leaks from the ML model with to identify an individual natural person. If another person or entity outside the controller has access to the trained ML model, the person or entity will have to identify a natural person from

[13] See, Judgement of 20 December 2017 [GC] C-434/16 *Peter Nowak v. Data Protection Commissioner* ECLI:EU:C:2017:994 [34] where the CJEU states that the wording «any information» in the Data Protection Directive «encompasses all kinds of information, not only objective, but also subjective (…) ».

[14] Article 29 Working Party 'Opinion 4/2007 on the concept of personal data' (WP 136, 20 June 2007).

[15] ibid. Pages 12-14.

[16] See further about the identifiability criterion in Nadezhda Purtova, From knowing by name to targeting: the meaning of identification under the GDPR, International Data Privacy Law, Volume 12, Issue 3, August 2022, Pages 163–183, https://doi.org/10.1093/idpl/ipac013 accessed: 15.07.2023.

[17] Judgement of 19 October 2016 [GC] C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* ECLI:EU:C:2016:779.

[18] See also Judgement of 26 April 2023 [General Court, Eight Chamber, extended Composition] T-557/20 Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS) ECLI:EU:T:2023:219 regarding the indirect identifiability criterion and whether identifiability is accessed from the controller or another entity under the similar definition in Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC [2018] OJ L 295/39 Article 4(1).

[19] 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 freedom of such data [1995] OJ L 281/31.

[20] Judgement of 19 October 2016 [GC] C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* ECLI:EU:C:2016:779 [41].

[21] Judgement of 19 October 2016 [GC] C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* ECLI:EU:C:2016:779 [43].

the ML model itself or potential leaks from the trained model combined with openly available sources. These two scenarios differ substantially in how difficult indirect identification is. In a situation where a ML model is trained on personal data, indirect identification is, however, potentially possible both for the controller responsible for training the model and a third party.

The second question under discussion to conclude on the legal status of a ML model trained on personal data is whether indirect identification from a ML model trained on personal data is a mean "reasonably likely to be used". From the Breyer case, we know that the CJEU has concluded that when access to the information necessary to indirectly identify an individual is prohibited by law it does not constitute a mean "reasonably likely to be used".[22] If a controller or third person would need to attack the ML model in a manner that comes under the scope of criminal law, the ML model would thus not constitute personal data.

When assessing whether a mean is reasonably likely to be used for identifying a natural person all objective factors are relevant, including the cost, amount of time required for identification, the available technology, and technological developments.[23] The next paragraphs will assess these relevant objective factors from indirect identification of natural persons in two different examples: Indirect identification from a trained ML model from the perspective of the controller responsible for training the model and indirect identification from a trained ML model in an entity outside the controller that has trained the model.

2.2.3. *Indirect identifiability: two scenarios*

For the first scenario under consideration, assume that a controller has trained a specific ML algorithm on a set of personal data from several different data subjects. The finished trained ML model is ready to be used on previously unseen data. The relevant legal question to consider in such a scenario is whether the trained ML model can indirectly identify a natural person and whether such identification is a mean reasonably likely to be used.

The purpose of training a ML model on a set of training data is to design the algorithms and to tailor the model to the specific data environment represented in the training data. In such a setting, it is common that the model performs better and more accurately on the training data than previously unseen data, a problem known as overfitting.[24]

There have been numerous experimental evaluations on unintended memorization in ML models.[25] Unintended memorization is an issue related to the overfitting problem, but unintended memorization is not completely the same as overfitting. Unintended memorization takes place during model training when a specific data point, for instance, a credit card number, in the training data is assigned a significantly higher likelihood in the model than what would be expected by random chance.[26]

The purpose of ML training is for the finished trained model to perform a prediction or a task. In such a model training process the model must intentionally memorize some data points from the training data. In particular ML approaches such as K-nearest neighbor classification and support vector machines, data points from the data set are applied as direct encoders in the ML model.[27] In such instances memorization of the data points is intentional and personal data becomes an internal part of the ML model. In other types of ML, however, the memorization of (sensitive) personal data in the finished trained ML model is an unintended and accidental side effect of the model training process.

Experimental evaluation of natural language processing has revealed that models for next-word suggestions unintendedly memorized social security numbers and other examples of personal data from the training data set.[28]

When assessing if such unintended memorization from the training data has the consequence that the finished trained model constitutes personal data under the GDPR, the answer is dependent on several factors.

These factors include whether data has been unintentionally memorized by the ML model and the status of the unintentionally memorized data. In a situation as described above where social security numbers or other directly identifiable identifiers have been unintentionally memorized by the ML model, the model would constitute personal data under Article 4(1) GDPR regardless of whether the full training data set is available.

In situations where indirect identifiers, for instance, age, metadata, or other factors in combination have been unintentionally memorized in the inner workings of the ML model and are available through "leaks" from the model, the status of the model as personal data would depend on whether the identification of an individual from this data is a mean that is "reasonably likely to be used". This specific "reasonably likely to be used"-test would differ in situations where the full train-

[22] Judgement of 19 October 2016 [GC] C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* ECLI:EU:C:2016:779 [46-49].

[23] See Recital 26 of the GDPR.

[24] Tom Dietterich, Overfitting and Undercomputing in Machine Learning. ACM Computing Surveys. Vol 27, No 3. September 1995.

[25] See for instance, Nicolas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos and Dawn Song, 'The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks'. In The Proceedings of the 28th USENIX Security Symposium. Vol. 267. 2019. Pages 267-284 and Vitaly Feldman, 'Does learning require memorization? A short tale about a long tail'. In STOC 2020: Proceedings of the 52nd Annual ACM Sigact Symposium on Theory of Computing. Pages 954-959. DOI:10.1145/3357713.3384290.

[26] See, Nicholas Carlini, 'Evaluating and Testing Unintended Memorization in Neural Networks'. Available: https://bair.berkeley.edu/blog/2019/08/13/memorization/ accessed: 15.07.2023

[27] Gavin Brown, Mark Bun, Vitaly Feldman, Adam Smith, Kunal Talwar, 'When Is Memorization of Irrelevant Training Data Necessary for High-Accuracy Training'. In STOC 2021: Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing. DOI:10.1145/3406235.3451131.

[28] Om Thakkar, Swaroop Ramaswamy, Rajiv Mathews, Françoise Beaufays. (2020). Understanding unintended memorization in federated learning. arXiv preprint arXiv:2006.07490. Nicolas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos and Dawn Song, 'The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks'. In The Proceedings of the 28th USENIX Security Symposium. Vol. 267. 2019. Pages 267-284.

ing data set is available or whether the ML model is trained on, for instance, a publicly available and well-known data set.[29]

To conclude on the question of whether a ML model trained on personal data could fulfill the definition of personal data in the GDPR, it would depend on whether identifiers are intentionally or unintentionally memorized by the ML model, the status of such memorized identifiers, and whether the identification of an individual is reasonably likely from the memorized data points.[30] The assessment of these criteria would differ on a case-to-case basis. However, a ML model trained on personal data could in specific circumstances fulfill the definition of personal data in Article 4(1) GDPR.

3. The right to erasure in a ML context

3.1. *Preliminary overview*

The former section concluded that a ML model trained on personal data could, in specific circumstances, constitute personal data. It is also evident from the former section that unintended memorization of personal data is an unresolved concern in ML model training. In situations where the trained ML model fulfills the definition of personal data under Article 4(1) GDPR, it is, therefore, not sufficient to erase personal data points from the training data set to comply with a request for erasure. The following section will first analyze the full scope of the right to erasure under Article 17 GDPR in Section 3.2. Then, the complexity of unlearning in ML is examined in Section 3.3. In Section 3.4, different model retraining and machine unlearning approaches are presented. Section 3.5 analyses and concludes on whether Article 17 of the GDPR would, in some specific scenarios, require full model retraining of the ML model trained on personal data.

3.2. *The scope of the right to erasure under the GDPR*

Machine unlearning is a technological response to the problem of both unintended memorization in ML and the potential for privacy and data protection breaches in ML. Several technological approaches to machine unlearning have the right to erasure in the GDPR as a specific motivation for the development of the different unlearning approaches.[31]

Before the practical enforcement of the right to erasure under the GDPR through methods of machine unlearning could be analyzed, the full scope of the right to erasure in a ML context needs to be explored.

The repealed data protection directive[32] did not have a specific right corresponding to Article 17 of the GDPR. Article 12(1) litra (b) of the Directive contained a right for the data subject to obtain from the controller the erasure of data in situations where the processing of the personal data did not comply with the provisions of the Directive. Furthermore, Article 14(1) litra (a) of Directive 95/46 contained a right for the data subject to object to the processing at any time on compelling legitimate grounds relating to his or her particular situation.

In the Google Spain judgment, the CJEU interpreted these different Articles in Directive 95/46 in relation to a right to obtain erasure of personal data.[33] The case concerned multiple referred questions on both the territorial scope of the Directive and whether data protection rights could be enforced against a search engine. The relevant question regarding the right to erasure concerned search results in a search engine related to an 18-year-old news story about a foreclosure due to social security debt.

A Spanish national had launched a complaint with the Spanish data protection authority to get these search results delisted (erased) from the search engine. The case entered the Spanish courts and was referred to the CJEU. The CJEU interpreted the data subject's rights under Directive 95/46 in light of the fundamental rights in the Charter Articles 7 and 8 and concluded that the data subject had a right to obtain delisting of the search results from Google.

Following the Google Spain judgement, the specific right to erasure has been included in Article 17 of the GDPR. According to Article 17(1), the data subject has the right to obtain erasure of personal data concerning him or her without undue delay in the following circumstances:

(a) If the personal data is no longer necessary for the purpose it was collected for.[34]
(b) If the processing relies on consent as the legal basis and the data subject has revoked the consent.
(c) The data subject has objected to the processing, and the controller's interests in continuing the processing do not override the data subject's interests.

[29] For a similar approach to identifiability in other instances, see the Italian data protection authority in the decision in the case Caffeina Media S.r.l. Available: https://edpb.europa.eu/news/national-news/2022/italian-sa-bans-use-google-analytics-no-adequate-safeguards-data-transfers_enaccessed 13.07.2023.

[30] See further, Michael Veale, Reuben Binns, Lilian Edwards, 'Algorithms that remember: model inversion attacks and data protection law'. In Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences Volume 376, Issue 2133. DOI.10.1098/rsta.2018.0083 and M. R. Leiser, Francien Dechesne, Governing machine-learning models: challenging the personal data presumption. In International Data Privacy Law, Vol. 10, Issue 3, Pages 187-200. DOI:10.1093/idpl/ipaa009.

[31] See for instance, Sebastian Schelter, Stefan Grafenberg, 'HEDGECUT: Maintaining Randomised Trees for Low-Latency Machine Unlearning. In SIGMOD' 21: Proceedings of the 2021 International Conference of Management of Data. DOI: 10.1145/3448016.3457239, Junyaup Kim, Simon S. Woo, 'Efficient Two-Stage Model Retraining for Machine Unlearning'. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2022). DOI: 10.1109/CVPRW56347.2022.00482, and Laura Graves, Vineel Nagisetty, Vijay Ganesh, 'Amnesiac Machine Learning'. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 13, pp. 11516-11524).

[32] 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 [1995] OJ L 281/31.

[33] Judgement of 13 May 2014 [GC] C-131/12 *Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja Gonzáles* ECLI:EU:C:2014:317.

[34] Erasure in such a situation is also a general obligation for the controller under the principles in Article 5(1) litra (b) (purpose limitation), litra (c) (data minimisation), and litra (e) (storage limitation).

(d) The personal data has been processed unlawfully.
(e) Erasure is necessary to comply with a legal obligation to which the controller is subject.
(f) The personal data has been collected in relation to the offer of an information society service to a child pursuant to Article 8(1) GDPR.

Article 17(2) GDPR contains broad exemptions for the right to obtain erasure. The right to erasure does not apply if: I) The processing is necessary for exercising the freedom of expression and information, II) if further processing of the personal data is necessary for compliance with a legal obligation to which the controller is subject, III) for reasons of public interests in the area of public health, IV) for archiving purposes, and V) for the establishment, exercise, or defense of legal claims.

Article 17 contains the grounds where the data subject has the right to obtain erasure of personal data concerning him or her. Article 17 does not explicitly answer how the controller should proceed in the task of erasing the personal data.

The GDPR is a technology-neutral piece of regulation. The technology-neutrality entails that the regulation applies both to simple processing of personal data for instance on a piece of paper that forms part of a filing system, and to complex processing of personal data in training ML models. Generally, the full erasure of personal data is a more straightforward task for simpler processing techniques.

According to Article 17(2), the following principle is relevant when assessing the scope of the right to erasure in situations where the controller has made the personal data public: "(…) the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data".

Subparagraph 2 of Article 17 only directly applies to the situations where the controller has made the personal data public. For the remainder of the assessment of the scope of the right to erasure in a ML context, suppose that a trained ML model is made public. A data subject that was part of the training data set then invokes the right to erasure against the controller that has trained the ML model according to the right set out in Article 17(1) litra b). In such a situation, the full erasure of personal data is dependent on whether full erasure is a reasonable step, where the available technology and cost of implementation are relevant objective factors.

The next section will examine the technological feasibility of machine unlearning in order to conclude on the scope of the right to erasure in a ML context.

3.3. *The complexity of machine unlearning*

In the process of training a ML model on a data set containing personal data, each data point – the information fed into the ML model – may influence the finished trained model. The ML model might both intentionally and unintentionally memorize the personal data during such a model training process depending on the ML algorithms used. When training a complex model, such as a neural network, the training data set might consist of millions if not billions of data points. Moreover, the neural network itself might consist of hundreds or thousands of different nodes and internal layers. GPT- 3 has, for instance, 96 layers and 175 billion trainable parameters.[35] The right to erasure of personal data is a complicated right to enforce in such a model training scenario both due to the vast amount of data points applied to train the model and the stochasticity of the model training. The stochasticity in model training refers to the problem that it is not necessarily known which impact part of the data set has had on the finished trained ML model.[36] The next section will examine whether recent developments in machine unlearning could be an efficient method to invoke the right to erasure in such a setting.

3.4. *From a request of erasure to model retraining*

3.4.1. *Background and context*

In a ML process, data is applied to train a model and the model is then used to make predictions, recommendations, decisions, or to draw inferences. When a data subject invokes the right to erasure against the controller that has trained the ML model, the first step in complying with such a request is to erase the personal data from the data set used to train the model. However, the model itself might contain traces of the personal data in the training data.[37] An unlearning process essentially consists of retraining the model without the erased personal data and some metrics to evaluate and audit whether the retrained model still consists of the personal data erased from the training data set.[38] The unlearned model is then used to make the predictions, recommendations, decisions, or to draw inference.

3.4.2. *Exact retraining*

An inherent problem related to machine unlearning is to define when the unlearned model is sufficiently unlearned. It is an easy task to remove a data point from a data set. However, this specific personal data point might influence the ML model even after a machine unlearning process. The relevant question is therefore: At what point is the impact from the personal data considered inadequate, such that the individual personal data point within the model is deemed sufficiently unlearned? Some definitions of machine unlearning require that a data point is only successfully unlearned if the data point is erased from the training data and the model is fully retrained from the beginning without the erased data point. Such methods are often referred to as exact machine unlearning, perfect machine unlearning, or complete retraining.[39] Fig. 1 illustrates an exact machine unlearning process.

---

[35] Cheng He, GPT-3: The Dream Machine in the Real World. Available: https://towardsdatascience.com/gpt3-the-dream-machine-in-real-world-c99592d4842f accessed:15.07.2023.

[36] See, Lucas Bourtoule, Varun Chandrasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, Nicolas Paternot, 'Machine Unlearning'. In proceedings of the 42nd IEEE Symposium of Security and Privacy. Available: https://arxiv.org/abs/1912.03817.

[37] See section 2.2.

[38] See, Thanh Tam Nguyen, Thanh Trung Huynh, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, Quoc Viet Hung Nguyen, 'A Survey of Machine Unlearning' (2022). ArXiv. Available: https://doi.org/10.48550/arxiv.2209.02299.

[39] Anvith Thudi, Gabriel Deza, Varun Chandrasekaran, Nicolas Papernot, 'Unrolling sgd: Understanding factors influencing machine

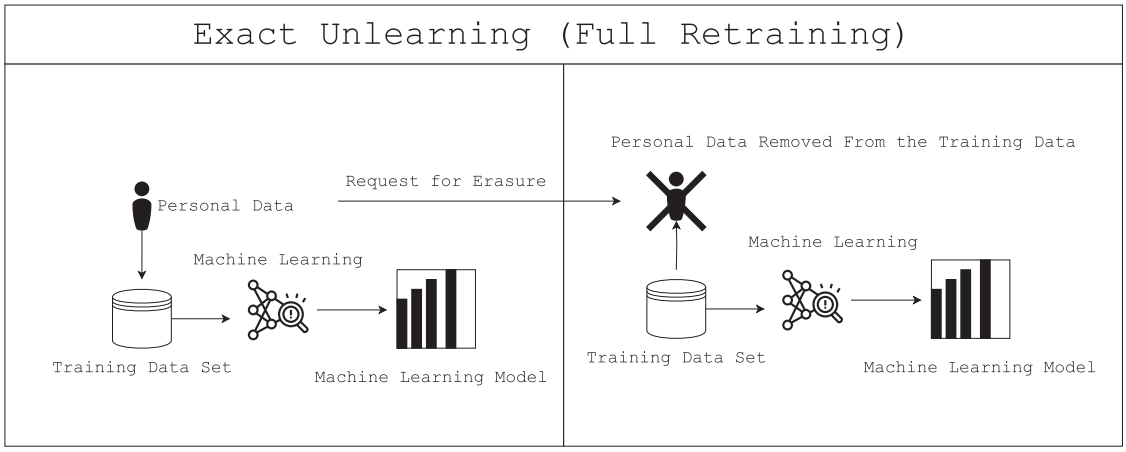

Fig. 1 – Illustration of exact unlearning (full retraining).

An exact machine unlearning process has to be repeated for every request for erasure of personal data and such a retraining process may be both computationally costly and inefficient.

A method to make complete retraining more approachable is through SISA, short for Sharded, Isolated, Sliced, and Aggregated training.[40] In such a ML approach the training dataset is isolated into smaller groups known as shards. Personal data from one data subject would be part of only one shard. The parameters of the trained ML model are stored continuously after the model gets trained on each of the shards. If and when a request for erasure of personal data is submitted, the model parameters from the training of the shards before the affected shard could be retrieved and the model could be retrained on the shards without the personal data. If for instance, the personal data that is requested to be erased is part of the third shard, the stored model parameters trained on shard one and two could be restored and the model could be retrained on shard number three without the erased personal data. The SISA approach is illustrated in Fig. 2.

The SISA approach to model retraining is not without shortcomings. The sharding reduces the amount of data each model is trained on which could reduce the ML model's accuracy. Furthermore, the sharding procedure also increases the need for storage capacity of the training data.

Complete retraining of a ML model, such as the SISA approach, is also associated with other constraints. First, when the personal data is erased from the data set and the model is retrained, it might affect the coherence between the data set and the trained model. The personal data might still affect the model even though the model has been retrained without the personal data due to the non-deterministic nature and the stochasticity in the model training. When the personal data is erased from the data set it is impossible to attribute the model to the specific erased personal data point.

Furthermore, model retraining might be a long and time-consuming process for ML with long training latencies and the overall performance of the model might be affected by the retraining. The concern related to the time and resources full retraining takes is the motivation behind other approaches to machine unlearning, which attempts to get the advantages of full model retraining without the actual retraining.

3.4.3. *Approximation machine unlearning*

An approach to machine unlearning without full retraining is approximation unlearning, also sometimes referred to as bounded or certified unlearning.[41] In approximate unlearning the overall aim is to achieve the advantages of machine unlearning while at the same time avoiding computationally costly retraining of the ML model.

Approximation methods for machine unlearning do not retrain the entire ML model from the beginning but adjusts the weights of the trained ML model to approach the results of unlearning through approximation. In such a process a machine unlearning criterion is first chosen. For instance, such a criterion might be that the new ML model has successfully been unlearnt if the adjusted model has a weaker performance on the data point that is requested to be erased. Another approximation unlearning approach could be to adjust the weights of the trained model to become similar to a ML model that has not been trained on the erased data point.

Unlike full retraining of the ML model, machine unlearning through approximation does not come with any theoretical guarantee that the adjusted ML model is not influenced by the personal data that is requested to be erased. In approximation unlearning the model may still be influenced by the personal data. The ML model is simply approximated to not

unlearning'. In proceedings of the IEEE 7th European Symposium on Security and Privacy (EuroS&P). IEEE, 2022. p. 303-319.

[40] Bourtoule, Lucas, Varu Chandrasekaran, Cristopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie and Nicolas Papernot. (2021, May). Machine unlearning. In 2021 IEEE Symposium on Security and Privacy (SP) (pp. 141-159). IEEE.

[41] See for instance, Laura Graves, Vineel Nagisetty, and Vijay Ganesh. 2021. Amnesiac machine learning. In AAAI, Vol. 35. 11516–11524.

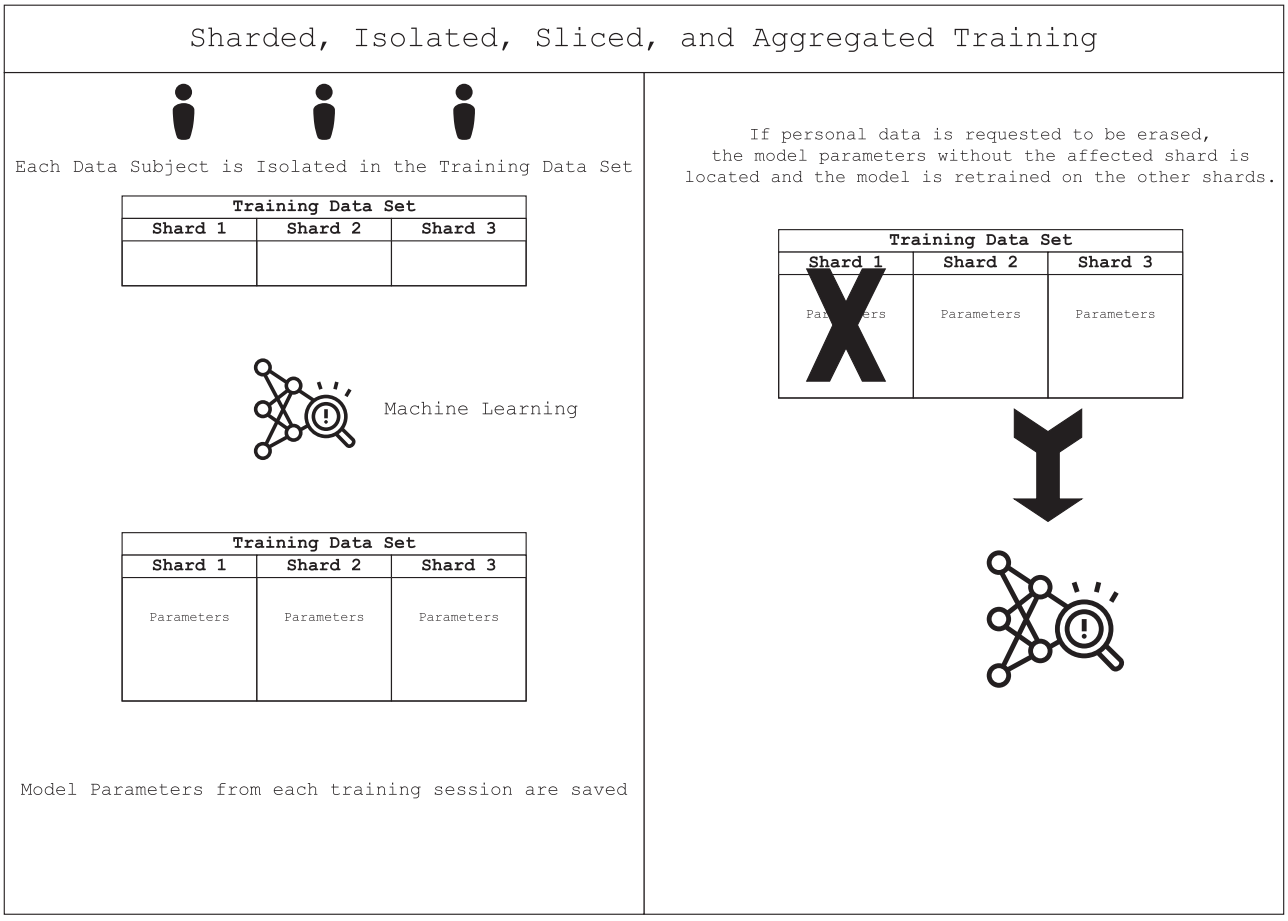

Fig. 2 – Illustration of sharded, isolated, sliced, and aggregated training.

be influenced by the personal data by adjusting the weights of the ML model.[42]

As stated above, machine unlearning approaches are classified as either complete retraining approaches or machine unlearning through approximation. The specific machine unlearning algorithms can be further classified as model-agnostic machine unlearning algorithms, model-intrinsic machine unlearning algorithms, or data-driven machine unlearning algorithms.[43]

3.4.4. *Machine unlearning algorithms: model-agnostic unlearning*

Model-agnostic machine unlearning algorithms consists of unlearning algorithms and approaches that are not ML model-specific. These approaches work on several different ML models. A potential method for model-agnostic machine unlearning is differential privacy.[44]

Differential private algorithms add artificial noise to the data set to prevent the possibility of the output being inferred back to an individual in the data set. Differential privacy provides a mathematical guarantee that the added noise to the data set makes the data point differential private, i.e., that it is unlikely, but not impossible to infer an individual from the output.[45] A guarantee has a 0 % probability to infer an individual would, however, make it impossible for the ML algorithm to learn.[46]

[42] Anvith Thudi, Gabriel Deza, Varun Chandrasekaran, Nicolas Papernot, 'Unrolling sgd: Understanding factors influencing machine unlearning'. In proceedings of the IEEE 7th European Symposium on Security and Privacy (EuroS&P). IEEE, 2022. p. 303-319.

[43] See, Thanh Tam Nguyen, Thanh Trung Huynh, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, Quoc Viet Hung Nguyen, 'A Survey of Machine Unlearning' (2022). ArXiv. Available: https://doi.org/10.48550/arxiv.2209.02299.

[44] Lucas Bourtoule, Varun Chandasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, Nicolas Papernoot. Machine unlearning. In 2021 IEEE Symposium on Security and Privacy (SP) (pp. 141-159). IEEE. DOI: 10.1109/SP40001.2021.00019.

[45] Thanh Tam Nguyen, Thanh Trung Huynh, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, Quoc Viet Hung Nguyen, 'A Survey of Machine Unlearning' (2022). ArXiv. Available: https://doi.org/10.48550/arxiv.2209.02299.

[46] Lucas Bourtoule, Varun Chandasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, Nicolas Papernoot. Machine unlearning. In 2021 IEEE Symposium on Security and Privacy (SP) (pp. 141-159). IEEE. DOI: 10.1109/SP40001.2021.00019.

3.4.5. *Model-intrinsic machine unlearning*

Model-intrinsic machine unlearning algorithms are approaches designed for specific ML models. To explain how a typical model-intrinsic unlearning algorithm works, assume that the machine unlearning method is model-intrinsic for tree-based ML models. In order to explain a machine unlearning model-intrinsic approach for tree-based models, it is first necessary to explain how a tree-based ML model works. Suppose that I want to decide whether or not I want to do an outdoor activity on a given day. To help me decide, I draw a decision tree. The tree consists of parameters with threshold values where values over or under the given threshold make up branches of the tree. For instance, if the temperature is above 10 ° Celsius, I go out and do the outdoor activity, if the temperature is below 10 ° Celsius, I stay inside. Different parameters, for instance, if it is sunny outside, how strong the wind is, and the time of the day makes up the other branches of the decision tree. Using ML, such a decision could be scaled up using many decision trees at the same time, known as a random forest.[47]

Assume that a random forest is applied to make a recommendation and personal data is used in training the random forests. How could a personal data point in the training data be unlearned? In a random forest or other large decision trees, the split features (in the example above sun, wind, and temperature) and the cut-off values (in the example: above or below 10 ° Celsius) are chosen by the ML algorithm. One method for machine unlearning in such a scenario involves measuring how robust the splits are. A split in the decision tree is robust enough if removing some random points from the training data does not affect the decision tree.[48] Note that such a model-intrinsic algorithm is only an approximation of the fact that the data point is unlearned, there is no theoretical framework supporting that a specific personal data point is unlearned in the downstream derivate of the personal data point, the tree-based ML model.

3.4.6. *Data-driven machine unlearning*

Data-driven machine unlearning algorithms include methods that manipulate or make changes to the training data in order to unlearn personal data. The SISA approach mentioned above is a data-driven method for machine unlearning. Other examples include adding noise to the training data in order to change the model from making predictions on the personal data requested for erasure. A further approach adds generative data to the model training in order to make the ML model less dependent on personal data. Generative Adversarial Networks (GANs) might for instance be applied in the model training process to make the model less dependent on personal data. Moreover, the ML model might be retrained on generative data after a request for erasure in order to both fulfill the erasure obligation and, at the same time, not lose necessary ML model accuracy.[49]

Overall, all the methods discussed above, model-agnostic machine unlearning, model-intrinsic machine unlearning, and data-driven methods, do not come with any theoretical guarantee that the requested erased personal data is fully removed from the downstream derivate of the personal data, the trained ML model. Only full retraining of the ML model comes with such a theoretical guarantee. The question that needs to be concluded is whether a request for erasure under GDPR Article 17 requires such a full model retraining or whether an approximation method is sufficient to comply with Article 17 of the GDPR.

3.5. *Does the right to erasure contain a right to full retraining of a ML model?*

The next question under discussion is, therefore, which one of these current broad approaches to machine unlearning – full model retraining or unlearning through approximation – that fulfill the data subject's right to erasure of personal data, and the corresponding obligation for the controller to erase the personal data, under Article 17 of the GDPR.

In situations where the controller has made the personal data public, the controller shall take reasonable steps in informing other controllers who are processing the personal data about the request for erasure from the data subject. Such reasonable steps include technical measures, according to Article 17(2) of the GDPR. A situation where personal data is made public has some similarities with training a ML model. When training a ML model, personal data is used, and the derivative of the personal data is shared. A request for erasure regarding a trained ML model addressed solely by discontinuing its use in relation to the data subject would therefore not be sufficient. In the event that the model incorporates personal data, the processing of said personal data would persist even if the model were employed in relation to other individuals or data subjects.

The steps the controller has to take to erase personal data when personal data is made public shall consider both the cost of implementation and the available technology, according to GDPR Article 17(2). Full retraining of a ML model is, generally, more costly than approximation methods for machine unlearning. However, only full retraining comes with a theoretically founded guarantee that the ML model is not influenced by the erased personal data.

There is little to no guidance to seek in the recommendations from the EDPB on the matter. In the Guidelines 5/2019 on the criteria of the right to be forgotten in search engine cases under the GDPR, there is however one relevant paragraph.[50]

The guidance concerns the delisting of search results in search engines. The EDPB addresses the obligation for search

[47] Leo Breima, 'Random Forests'. *Machine Learning*. 45 (2001). 5-32.

[48] Ayush Sekhari, Jayadev Acharya, Gautam Kamath, Ananda Theertha Suresh. 2021. Remember what you want to forget: Algorithms for machine unlearning. NIPS 34 (2021), 18075–18086.

[49] Deepanjali, S., Dhivya, S., Monica Catherine, S. (2021). Efficient Machine Unlearning Using General Adversarial Network. In: Hemanth, D., Vadivu, G., Sangeetha, M., Balas, V. (eds) Artificial Intelligence Techniques for Advanced Computing Applications. Lecture Notes in Networks and Systems, vol 130. Springer, Singapore. https://doi.org/10.1007/978-981-15-5329-5_45.

[50] EDPB, Guidelines 5/2019 on the criteria of the Right to be Forgotten in search engines cases under the GDPR 5. Available: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_201905_rtbfsearchengines_forpublicconsultation.pdf. Last accessed 13.07.2023.

engines to erase and gives the following guidance: "(…) search engine providers are not exempt in a general manner from the duty to fully erase. In some cases, they will need to carry out actual and full erasure in their indexes or caches. For example, in the event that search engine providers would stop respecting robots.txt requests implemented by the original publisher, they would actually have a duty to fully erase the URL to the content, as opposed to delist which is mainly based on data subject's name".[51]

In the recommendations from the EDPB, the European Data Protection Board recommend that search engine providers in some cases must erase the URL to the content in order to fulfill the duty to fully erase personal data under Article 17 GDPR. A URL is – similarly to a ML model – a downstream derivative, of personal data.

The GDPR is to be interpreted on a case-to-case basis and currently, there is little guidance to seek from the EDPB, the CJEU or national data protection authorities on the question of machine unlearning and the right to erasure.

However, based on the legal analysis in Sections 2 and 3 the following general rule could be established: In situations where it is proven that the trained ML model leaks personal data from the model training and the data subject has requested the erasure of the personal data, it is not sufficient to just erasure the personal data from the training data set. In such a situation there has to be implemented some machine unlearning methods on the trained ML model. The only current machine unlearning method with a theoretically proven guarantee is exact machine unlearning also known as full model retraining. In such a situation, where the ML model leaks personal data and the data subject has requested erasure, the request for full model retraining has a legal basis in Article 17 of the GDPR and should be implemented for full GDPR compliance.

In a situation where the ML model is trained on personal data, but there is not any evidence that identifiability is possible only from the ML model, a request for erasure could be complied with through an approximation unlearning method. In such a situation, the personal data has to be erased from the training data set, but the ML model could be altered through an approximation method to comply with Article 17 of the GDPR. Such a solution to the right to erasure in ML has a legal basis in the relativization of the right to erasure when personal data is made public under subparagraph 2 of Article 17 GDPR.

4. Concluding remarks

Since the right to erasure is dependent on both the cost of implementation and the available technology under subparagraph 2 of Article 17 of the GDPR, a potential for future research would be to experiment on the cost of full model retraining and compare the cost to machine unlearning through approximation. Furthermore, there is a potential for future research in experiments on evaluating the metrics on whether the machine unlearning has been successful, and the cost of such metric testing.

ML models and data protection law have been a discussed topic in legal literature. Leiser and Dechesne argue that data protection law is not an efficient way to govern ML models and that criminal law as deterrence to unlawful model inference is more efficient.[52] The authors also argue that trade secret protection in some cases would be a more efficient governance tool for ML models than data protection law.[53]

When other controllers apply trained ML models through Application Programming Interfaces (APIs), the trade secret perspective comes somewhat short because the outputs of the ML models get shared and come into the hands of another entity. Furthermore, criminal law as deterrence for unlawful model inference also falls short when the model unintentionally leaks personal data without a model attack. Criminal law as prevention of unlawful model inference is also challenged by the international and cross-border nature of training large ML models. We argue that data protection law is a sound governance tool for ML models in some instances and that the right to erasure under Article 17 of the GDPR is relevant in a ML context.

When the Metaverse and autoregressive transformer language models such as Generative Pre-trained Transformers (GPTs) get more commonly used in different parts of society, we will most likely experience a further surge in ML training on personal data. In the Metaverse and other virtual interoperable spaces, this personal data will change from tabular personal information to sensory input and sensory impulses. In such a scenario, the right to erasure under the GDPR will become a crucial right to ensure the fundamental right to privacy and data protection.

The GDPR is technologically neutral. However, it needs to be interpreted in light of available technology. Machine unlearning methods are often developed with the right to erasure as a core motivation. To keep the GDPR as both a living and dynamic instrument and to ensure that the GDPR is interpreted in light of fundamental rights in the Charter of fundamental rights of the European Union and the Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), it is necessary to interpret the right to erasure in a ML context.

The right to erasure in Article 17 of the GDPR is an individual right. An individual remedy has some shortcomings because the enforceability of the right is dependent on individual data subjects making use of the right. However, accountability lies at the heart of algorithmic governance in the GDPR. It is therefore important that the right to erasure is interpreted together with the other rights in the GDPR and the controllers

[51] EDPB, Guidelines 5/2019 on the criteria of the Right to be Forgotten in search engines cases under the GDPR 5. Available: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_201905_rtbfsearchengines_forpublicconsultation.pdf. Last accessed 13.07.2023.

[52] M. R. Leisner and Francien Dechesne, 'Governing machine-learning models: challenging the personal data presumption' 10(3) *International Data Privacy Law*, 187, 187-200. https://doi.org/10.1093/idpl/ipaa009.

[53] M. R. Leisner and Francien Dechesne, 'Governing machine-learning models: challenging the personal data presumption' 10(3) *International Data Privacy Law*, 187, 187-200. https://doi.org/10.1093/idpl/ipaa009.

implement machine unlearning approaches in the design of their systems.

New, exciting, and groundbreaking technology such as ML is important to develop our societies and ML can function as an incubator for both innovation and new research. However, such technological developments should not come with fundamental rights infringement as an expense. It is therefore crucial to both interpret existing fundamental rights in light of technological development and to develop new technologies, such as machine unlearning, as an instrument for fundamental rights compliance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.