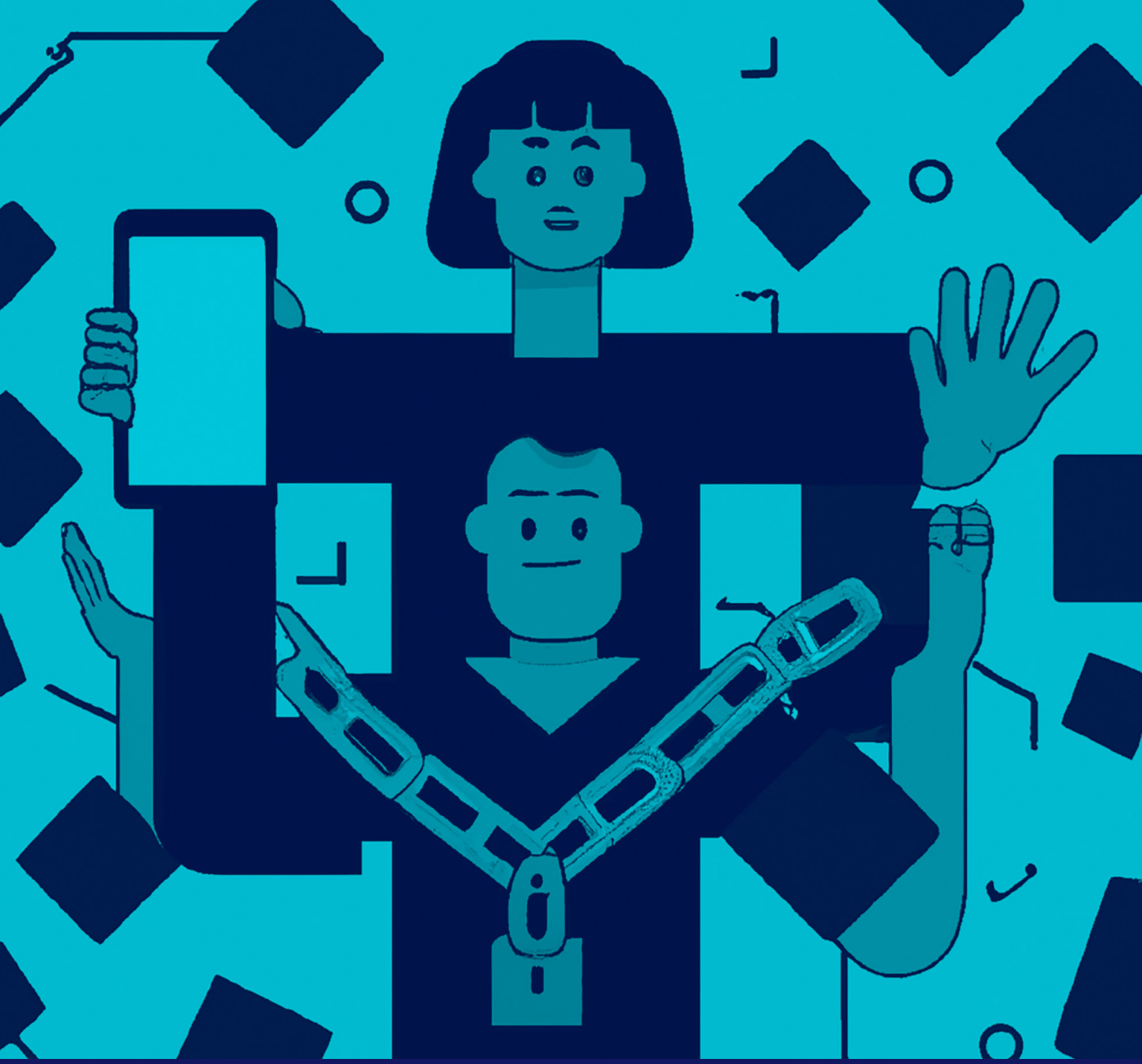


ChatGPT's midlife crisis: trying to forget its past (data)

Research into whether the training data of ChatGPT is compatible with the right to erasure within the meaning of Article 17 GDPR.

Research paper | Jorie Corten

watsonlaw.

1. Introduction

Artificial intelligence has revolutionised the way we interact with technology

Artificial intelligence has revolutionised the way we interact with technology, and one of the most recent and impressive examples of this is the development of language models such as ChatGPT.[1] These models can generate human-like text, making them useful for a wide range of applications, from customer service chatbots to content creation.

However, due to its broad applicability and fast emergence, ChatGPT has created many legal uncertainties. For instance, Italy has temporarily banned ChatGPT over privacy concerns.[2] Although ChatGPT faces various legal challenges under the European legal framework, this research only addresses the General Data Protection Regulation (GDPR).[3] This research aims to explore the following question: Is the training data of ChatGPT compatible with the right to erasure, within the meaning of Article 17 GDPR?

2. ChatGPT

A GPT model is designed to produce text that could have been written by a human

ChatGPT is a Large Generative Artificial Intelligence Model (LGAIM), a technology that is developed using an extensive amount of data.[4] It is built using the Generative Pre-trained Transformer (GPT) architecture, a machine-learning model capable of producing text after "pre-training" on a text database.[5] A GPT model is designed to produce text that could have been written by a human.[6] The GPT-4 model, which was trained on a set of texts published before September 2021, is the most recent model used by ChatGPT.[7] The interface of ChatGPT is similar to a standard chatbot, allowing users to ask questions in natural language (input) and get responses from the computer (output).[8] The training of ChatGPT was carried out in two phases: i) training data, and ii) human input data.[9]

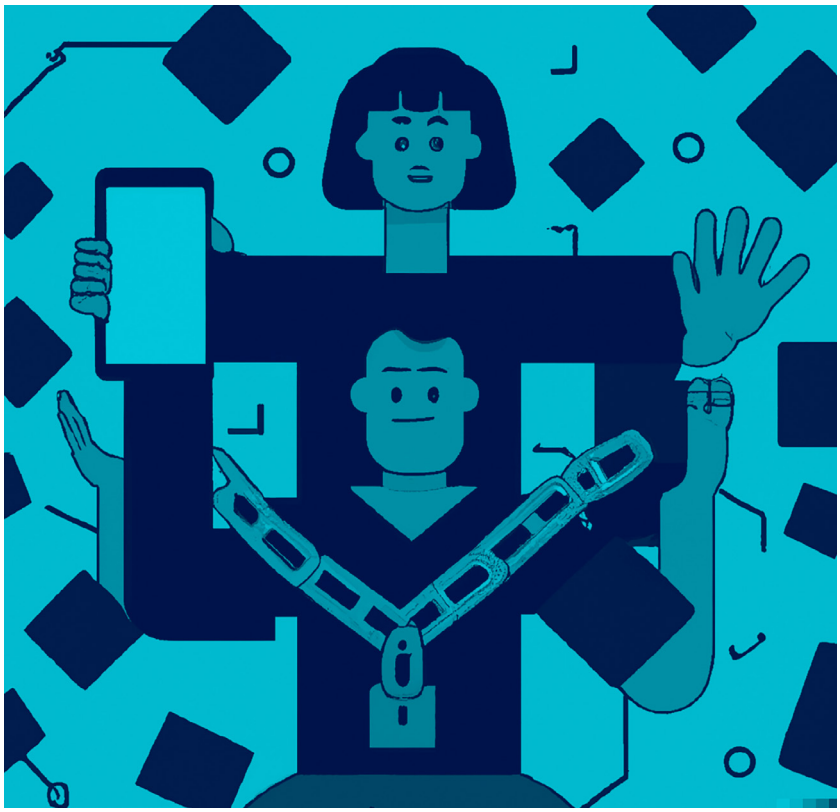
2.1. Training Data

ChatGPT was trained using a dataset of more than 45 terabytes of text from the internet, including books, papers, webpages, and other text-based content.[10] This dataset contains billions of words of text.[11] The training was carried out with the help of a neural network intended for natural language processing.[12] In order to understand the structure and patterns of human language, ChatGPT was trained using this data to anticipate the absence of certain words in a given text.[13] Using probability distributions to determine which sentences would most likely fit together as a solution to the user's question, the output is generated through a process of sampling and combining the training data.[14] The output may be biased and inaccurate due to its probabilistic responses and the imperfect quality of the training data (texts from the internet).[15]

2.2. Human Input Data

After being trained by text from the internet, ChatGPT was refined with human input. Hence, human text input also serves as the foundation for the output produced.[16] For instance, prompts such as *"write a research paper about the compatibility of ChatGPT with the right to erasure"* or users giving feedback by clicking a "thumbs up" button, help ChatGPT to train itself.[17] By merely incorporating the human answer to the output the algorithm can improve itself over time, even without additional training.[18] Therefore, human input data exemplifies the notion of reinforced learning.[19]

Human input data raises legal concerns surrounding users whom (sub)consciously provide personal data as input to ChatGPT. However, this research solely focuses on analysing the right to erasure of personal data within training data. The datasets used to train ChatGPT are typically sources from a wide range of texts available on the internet, which contains personal data that is protected under the GDPR.[20] Hence, I asked ChatGPT if it excluded this personal data from its dataset in its training process. ChatGPT responded that "*(...) ChatGPT itself did not exclude personal data from its dataset during training, (...).*" Therefore, this research continues to assume that the training data of ChatGPT includes personal data within the meaning of Article 4(1) GDPR.



Source: created by the author using DALL • E

3. The General Data Protection Regulation

The GDPR provides a thorough legal framework that harmonises data protection across the European Union (EU) and replaces the Data Protection Directive of 1995 (hereafter: DPD).[21]

The right to erasure is an essential online right of data subjects recognised by EU law

The GDPR is intended to acknowledge the significant technological advances of the last 25 years, especially the advent of networked information spaces like the Internet.[22] However, it was conceptualised and written before LGAIMs like ChatGPT, which are today recognised as one of the most revolutionary upcoming technologies.[23]

The GPDR applies *"to the processing of personal data (...)"*, which constitutes the material scope of the GDPR and is a fundamental notion in data protection law.[24] The data protection rights, such as the right to erasure, only apply when personal data is processed.[25]

3.1. The Right to Erasure

The right to erasure is an essential online right of data subjects recognised by EU law.[26] A data subject is a natural person who is identified or identifiable via personal data.[27] Pursuant to Article 17(1) GDPR *"data subjects have the right to obtain from the controller the erasure of personal data concerning him or her (...)"*. This right to erasure (or 'right to be forgotten') is established by the Court of Justice of the European Union (hereafter: CJEU) in its landmark judgement *Google Spain.*[28] Even though the GDPR constitutes the relevant legal framework today, at the time of this judgement the DPD was still in effect.[29] The DPD did not explicitly state the right to erasure, but in its judgement, the CJEU determined that it can be inferred from the DPD's Articles 12(b) and 14(1)(a), which respectively give the data subject the rights of rectification, erasure, or blocking of their personal data and the right to object to its processing.[30]

The CJEU established that data subjects have the right to erasure, which calls for the removal of links to webpages that contain information about them from the list of results displayed after a search is conducted using the person's name.[31] The fundamental right to data protection may be *"affected significantly"* by the activity of search engines, the CJEU determined.[32] Internet search engines are subject to their enforcement of data protection because they are regarded as *"controllers"* of data.[33] Internet search engines are regarded as controllers because they determine the purposes and means of the processing of personal data.[34] The CJEU ruled that *"a fair balance should be sought"* between the rights of data subjects to data protection and the legitimate interests of searchers.[35] The CJEU continues that the data protection of data subjects *"override, as a rule, not only the economic interest of the operator of the search engine but also the interest of the general public in having access to that information*

upon a search relating to the data subject's name.[36] " Nevertheless, the CJEU adds a crucial nuance. When the interference is justified by the interest of the general public in having access to the information in question, then the rights of the data subject should not prevail.[37]

The GDPR codified *Google Spain* and further elaborated the right to erasure.[38] The GDPR elaborated by enabling the data subject to demand the deletion of personal data relating to them on more grounds,[39] as well as enforcing the responsibility for the controller to erase personal data without undue delay.[40] Moreover, the GDPR lists exemptions for when the right to erasure does not apply, such as the performance of public interest and the right to freedom of expression and information.[41]

ChatGPT is an LGAIM and not an internet search engine

ChatGPT is an LGAIM and not an internet search engine. However, ChatGPT is commonly used as a search engine and was trained on data that internet search engines provide to the general public, so, therefore, I would argue that the same rules apply.

4. Erasing Personal Data on ChatGPT

Personal data from the trained model of ChatGPT can be erased in two different methods: (i) retraining the dataset, or (ii) machine unlearning.

4.1. Retraining the Dataset

On the basis of an amended training dataset, the ChatGPT model could be retrained.[42] Even on the most powerful computational infrastructures in the world, machine learning training is very intensive, making it expensive and time-consuming.[43] Data centres, which are used for machine learning training, absorb between 1.1% and 1.5% of the energy in the world.[44] Therefore retraining a model entails significant energy, time, and labour expenses.[45]

4.2. Machine Unlearning

Another option is amending the model itself after it has been trained ("machine unlearning").[46] However, this is very complicated and hardly ever feasible with existing systems.[47] Techniques for machine unlearning are just now being presented and are underexplored.[48] They are still a long way from being ready to be implemented.[49] The methods that are currently under discussion cannot be added to the existing systems and would necessitate a complete remodel, with unknown results.[50] As a consequence, big tech companies like Google erase results between the model output and the delivered result to searches, instead of removing links from the trained data.[51]

Most of the time, when personal data of a data subject is erased

from the training data, it has little impact on the patterns the model has already learned.[52] Models that base patterns on a single data record are typically seen as being unnecessarily overfitted, as the goal of models should be finding generalisable patterns instead of memorising the training data.[53] Therefore, the right to erasure becomes especially interesting when it is used in a collective action.[54] Depending on the degree of cooperation between data subjects, it might be feasible for data subjects to exercise their right of erasure on ChatGPT.[55] Crowdfunding and online petitions are interesting ways to explore collective actions regarding the right to erasure.[56]

5. Overarching Problems

On the basis of *Google Spain* is it lawful to collect large amounts of personal data for training purposes if the general public has considerable interest in having access to the information in question.[57] Following, the Article 29 Working Party, an independent European advisory body on data protection and privacy, stated that *"there is also an interest of internet users in receiving information using search engines"*.[58] In this regard, the basic right to freedom of expression, as defined in Article 11 of the European Charter of Fundamental Rights as *"the freedom to receive and impart information and ideas"*,[59] must be taken into account when evaluating the requests made by data subjects. Moreover, it is argued in the literature that in cases where someone requests their personal data to be erased, the rights to freedom of expression and data protection must be balanced on a case-by-case basis while taking all relevant factors into account.[60]

The rights to freedom of expression and data protection must be balanced on a case-by-case basis

On the contrary, when in a specific case the right to erasure of a data subject does prevail over the interests of the general public, another dilemma arises. Even though the server of ChatGPT is based in the United States (U.S.), the model is trained on data from all over the world.[61] The CJEU ruled in its landmark case *Google v. CNIL* that search engines should erase personal data only in the EU.[62] So, when a data subject invokes his or her right to erasure, this data is in principle only erased from the ChatGPT model that is accessible in Europe. However, the CJEU continued that EU law does not prohibit the erasure of personal data from all servers.[63] Consequently, a national data protection authority remains competent to order ChatGPT to carry out erasure concerning all models.[64]

This raises the key questions of this paper: (i) Is the interest of the general public regarding ChatGPT's training data less than the right to erasure of data subjects, so as to swing the reasoning of the CJEU in *Google Spain* for ChatGPT? And following, (ii) when a court rules that the right to erasure of a data subject prevails above the interest of the general public, should this data be erased from the server of ChatGPT in the U.S., from the model accessible in Europe, or just from the model accessible in the applicant's country?

6. Conclusion

This research has examined whether the training data of ChatGPT is compatible with the right to erasure, within the meaning of Article 17 GDPR.

This research revealed that the training data of ChatGPT likely includes personal data within the meaning of Article 4(1) GDPR and may, therefore, implicate the right to erasure. While it is technically possible to erase personal data of a data subject from ChatGPT, it has little impact on the patterns the model has already learned. Therefore, it is unclear whether the training data of ChatGPT is fully compatible with the right to erasure pursuant to Article 17 GDPR. Further research is needed to provide clarity on this issue. In the meantime, a potential solution to this issue is for data subjects to start a collective action.

In conclusion, while the GDPR provides a thorough legal framework for data protection, it was not specifically designed to address challenges posed by LGAIMs like ChatGPT. The legal uncertainties surrounding ChatGPT highlight the need to review the GDPR to ensure that it remains relevant and applicable to emerging technologies. Moreover, it highlights the need to ensure that LGAIMs like ChatGPT are developed and used in a manner that is fully compliant with applicable laws and regulations, including the right to erasure. It is crucial to strike a balance between technological innovation and data protection, and the right to erasure must be upheld to protect the privacy rights of data subjects.

Any questions?
Please contact **Jorie Corten** of Watsonlaw.

Footnotes

1 Hacker, P., Engel, A., & Mauer, M. (2023). Regulating chatgpt and other large generative ai models. arXiv preprint arXiv:2302.02337, p. 2-3. (Hereafter: Hacker et al (2023)); Hacker, P. (2023). Understanding and regulating ChatGPT, and other large generative AI models. Verfassungsblog: On Matters Constitutional, p 2. (Hereafter: Hacker (2023)).
2 Garante per la Protezione dei Dati Personali (GDPR). (31 March 2021). Artificial intelligence: stop to ChatGPT by the Italian SA Personal: data is collected unlawfully, no age verification system is in place for children [Press release]. https://www.gpdp.it/home/docweb/-/docweb-display/docweb/9870847#english
3 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).
4 Hacker et al (2023), p. 2-3; Sarel, R. (2023). Restraining ChatGPT, p. 8-9. (Herefter: Sarel (2023)).
5 Europol. (2023). ChatGPT: The impact of Large Language Models on Law Enforcement. Retrieved 7 April 2023, from https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf, p. 3-4. (Hereafter: Europol (2023)); Hacker et al (2023), p. 2-3; Sarel (2023), p. 8-9.
6 Sarel (2023), p. 9.
7 OpenAI. (2023). GPT-4. Retrieved 7 April, 2023, from https://openai.com/research/gpt-4.
8 Sarel (2023), p. 9.
9 Europol (2023), p. 3-4.
10 Europol (2023), p. 3-4; Sarel (2023), p. 9.
11 Europol (2023), p. 3-4.
12 Ibid.
13 Ibid.
14 Hacker et al (2023), p. 2-3.
15 Sarel (2023), p. 8-9.
16 Europol (2023), p. 3-4.
17 Hacker (2023), p 2.
18 Sarel (2023), p. 8-9.
19 Mann, D. L. (2023). Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine: A JACC: Basic to Translational Science Interview With ChatGPT. Basic to Translational Science; Sarel (2023), p. 8-9.
20 Protection of personal data and privacy. Council of Europe. Retrieved 8 April 2023, from https://www.coe.int/en/web/portal/personal-data-protection-and-privacy.
21 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.
22 Panel for the Future of Science and Technology (STOA). (2019). Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law? European Parliament. Retrieved 28 March 2023, from https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf, p 4-5; The European Union Blockchain Observatory and Forum. (2018). Blockchain and the GDPR. Retrieved 28 March 2023, from https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf, p. 7-8.
23 Hacker et al (2023).
24 Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. Law, Innovation and Technology, 10(1). Retrieved 29 March 2023, from https://doi.org/10.1080/17579961.2018.1452176, p. 43-44.
25 Ibid.
26 Tzanou, M. (2020). The unexpected consequences of the EU Right to Be Forgotten: Internet search engines as fundamental rights adjudicators. In Personal Data Protection and Legal Developments in the European Union (pp. 279-301). IGI Global, p. 1-2 of the electronic copy. (Hereafter: Tzanou (2020)).
27 Article 4(1) GDPR.
28 CJEU [GC] 13 May 2014, Google Spain, C-131/12, ECLI:EU:C:2014:317. (Hereafter: Google Spain).
29 Kulk, S., & Borgesius, F. Z. (2014). Google Spain v. González: Did the Court Forget about Freedom of Expression?: Case C-131/12 Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos and Mario Costeja González. European Journal of Risk Regulation, 5(3), 389-398. (Hereafter: Kulk & Borgesius (2014)); Tzanou (2020), p. 1.
30 Google Spain, paras. 70 and 76.
31 Google Spain para. 99 and dictum.
32 Google Spain para. 38.
33 Article 4(7) GDPR; Google Spain para. 33.
34 Article 4(7) GDPR; Google Spain para. 33.
35 Google Spain para. 81.
36 Google Spain para. 99 and dictum.
37 Ibid.
38 Tzanou (2020), p. 1-2.
39 Articles 17(1)(a-f) GDPR.
40 Article 17(1) GDPR.
41 Article 17(3) GDPR.
42 Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: model inversion attacks and data protection law. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180083, p. 9. (Hereafter: Veale et al (2018)).
43 Ibid.
44 Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J. M., & Vasilakos, A. V. (2014). Cloud computing: Survey on energy efficiency. Acm computing surveys (csur), 47(2), 1-36.
45 Ibid.
46 Veale et al (2018), p. 9.
47 Ibid.
48 Cao, Y., & Yang, J. (2015, May). Towards making systems forget with machine unlearning. In 2015 IEEE symposium on security and privacy (pp. 463-480). IEEE.
49 Ibid.
50 Veale et al (2018), p. 9.
51 Google. (2017). Search removals under European privacy law. Retrieved 8 April 2023, from https://perma.cc/8DE4-AXBW.
52 Veale et al (2018), p. 10.
53 Vedder, A. (1999). KDD: The challenge to individualism. Ethics and Information Technology, 1(4), 275-281.
54 See e.g. Ausloos, J. Toh, J., & Giannopoulou, A. (23 November 2022). The case for collective action against the harms of data-driven technologies. Ada Lovelace Institute. Retrieved 7 April 2023, from https://www.adalovelaceinstitute.org/blog/collective-action-harms/.
55 Veale et al (2018), p. 10.
56 Ibid.
57 Google Spain, para. 99.
58 Article 29 Working Party. (2014). Guidelines on the implementation of the Court of Justice of the European Union judgment on "Google Spain and Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González" C-131/12. Retrieved 8 April 2023, from https://ec.europa.eu/newsroom/article29/tems/667236/en, p. 6.
59 Ibid.
60 Kulk & Borgesius (2014), p. 6 of the electronic version.
61 OpenAI. (2023). GPT-4. Retrieved 7 April, 2023, from https://openai.com/research/gpt-4.
62 CJEU [GC] 24 September 2019, Google v CNIL, C-131/12, ECLI:EU:C:2019:772, para. 63. (Hereafter: Google v CNIL).
63 Google v CNIL, para. 72.
64 Ibid.

Questions?

Feel free to ask our specialists.






This paper is written by Jorie Corten, one of the blockchain experts of Watsonlaw. Whether you have questions regarding your crypto-oriented company, or have specific questions regarding the implementation of MiCA, do not hesitate to contact us. Watsonlaw is your partner in blockchain, tokenization and the crypto-market in the Netherlands and abroad, with a broad network of collaborating law firms throughout various countries to provide national as well as international advice. Not only do we possess extensive knowledge regarding the new MiCA regulation, we can also guide your company every step of the way to becoming a leading blockchain enterprise.

watsonlaw.nl

watsonlaw.