15

Generative AI and Data Protection

Hannah Ruschemeier

15.1 PRELUDE: RISKS AND CHALLENGES OF GENERATIVE AI

Now that the initial hype around generative AI in the form of large language models and image generators has subsided, legal issues are coming to the fore. In addition to discussions about generative AI and copyright, there is an increasing focus on the friction between generative models and the requirements of data protection law. In the United States, several lawsuits are underway against Google and OpenAI regarding potential privacy violations by generative models.[1] Regulators are currently active in the European Union[2] and the European Data Protection Board (EDPD) has set up a task force to deal with ChatGPT.[3] The Italian data protection authority Garante had already opened a case against OpenAI in 2023, which led to a temporary national ban on ChatGPT.[4] After the proceedings were concluded, the authority found violations of the General Data Protection Regulation (GDPR).[5] Investigations into data protection violations are also underway in Poland.[6] Other countries such as Germany have

[1] See the class action against OpenAI: <https://storage.courtlistener.com/recap/gov.uscourts.cand.425482/gov.uscourts.cand.425482.1.0.pdf> accessed 29 April 2024.

[2] For an overview of ongoing investigations of data protection authorities outside the European Union, see Gabriela Zanfir-Fortuna, 'How Data Protection Authorities Are De Facto Regulating Generative AI – Future of Privacy Forum', <https://fpf.org/blog/how-data-protection-authorities-are-de-facto-regulating-generative-ai/> accessed 29 April 2024. Regarding the Italian investigation: EDPD European Data Protection Board, 'Report of the Work Undertaken by the ChatGPT Taskforce' (2024), p. 5 <www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-chatgpt-taskforce_en> accessed 29 April 2024.

[3] European Data Protection Board (EDPB), 'EDPB Resolves Dispute on Transfers by Meta and Creates Task Force on Chat GPT | European Data Protection Board' (13 April 2023) <www.edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en> accessed 29 April 2024.

[4] Garante per la protezione dei dati personali (GPDP), 'Intelligenza artificiale: il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori' <www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870847> accessed 29 April 2024.

[5] Garante per la protezione dei dati personali (GPDP), 'ChatGPT: Garante privacy, notificato a OpenAI l'atto di contestazione per le violazioni alla normativa privacy' <https://gpdp.it:443/web/guest/home/docweb/-/docweb-display/docweb/9978020> accessed 29 April 2024.

[6] Urząd ochrony danych osobowych (UODO), 'Aktualności – UODO' <https://uodo.gov.pl/pl/138/2823> accessed 29 April 2024.

issued requests for information,[7] and the French data protection authority has developed an action plan.[8] In the case of Maximilian Schrems, the data protection non-governmental organization (NGO) NOYB filed a complaint with the Austrian data protection authority in April 2024,[9] centred on incorrect information about an individual provided by ChatGPT, which OpenAI did not correct and nor did it respond to the request for information about what data was processed. These cases make it clear that data protection authorities are already AI regulators and generative AI is a core issue for data protection.[10]

From a legal perspective, generative models introduce a range of distinct issues, which are well documented across various scholarly sources.[11] In particular, the foundation models on which the popular large language models (LLMs) are built pose new security risks and vulnerabilities that need to be addressed. This then gives rise to the need for a socio-technical assessment, including legal and ethical aspects, to understand these risks and the necessary safety mechanisms. Understanding the risks posed by LLMs requires a contextual approach: normative rules, like law, always operate in context.

A major concern is the protection of personal data and privacy. Different experiments have shown that it is possible to extract personal and sensitive information about individuals from LLMs.[12] Researchers have proven that LLMs are able to memorise training data, either through over-application of abundant parameters to small datasets, which reduces the capacity to generalise to new data, or through optimisation for generalisation in long-tailed data

[7] Landesbeauftragte für Datenschutz und Informationssicherheit NRW (LDI NRW), 'Prüfung von ChatGPT geht in die nächste Runde' (30 October 2023) <www.ldi.nrw.de/pruefung-von-chatgpt-geht-die-naechste-runde> accessed 29 April 2024. The State Commissioner for Data Protection and Freedom of Information Rhineland-Palatinate, LfDI asks about ChatGPT, 24 April 2023. The State Commissioner for Data Protection Schleswig-Holstein sent out a list of questions: <www.datenschutzzentrum.de/uploads/chatgpt/20230419_Request-OpenAI_ULD-Schleswig-Holstein_IZG.pdf> accessed 29 April 2024.

[8] Commission Nationale de l'Informatique et des Libertés (CNIL), 'Artificial Intelligence: The Action Plan of the CNIL' (16 May 2023) <www.cnil.fr/en/artificial-intelligence-action-plan-cnil> accessed 29 April 2024.

[9] None of Your Business (NOYB) – Europäisches Zentrum für digitale Rechte, 'ChatGPT Provides False Information about People, and OpenAI Can't Correct It' (29 April 2024) <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it> accessed 29 April 2024.

[10] Until the beginning of 2024, OpenAI had not been established in Europe, meaning no member state data protection authority was responsible under Art. 56, 60 GDPR. Therefore, the supervisory authorities of the member states could all act in accordance with Art. 55 GDPR within their areas of competence. OpenAI now operates an office in Dublin, which is designated as the data controller.

[11] Damien Charlotin, 'Large Language Models and the Future of Law' [2023] *SSRN Electronic Journal* <www.ssrn.com/abstract=4548258> accessed 24 April 2024; Siyin Chen, 'Potential Applications and Safety of Large Language Models in Healthcare' (2024) 1 *Interdisciplinary Humanities and Communication Studies* <www.deanfrancispress.com/index.php/hc/article/view/561> accessed 24 April 2024; Pier Giorgio Chiara, 'Italy. Italian DPA v. OpenAI's ChatGPT: The Reasons behind the Investigation and the Temporary Limitation to Processing' (2023) 9 *European Data Protection Law Review* 68; Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 5 *UC Davis Law Review* 2657; Julian Hazell, 'Spear Phishing with Large Language Models' (14 December 2023) <www.governance.ai/research-paper/llms-used-spear-phishing> accessed 1 October 2024; Henrique Marcos and Melina Pullin, 'Large Language Models and EU Data Protection: Mapping (Some) of the Problems – The Digital Constitutionalist' (11 October 2023) <https://digi-con.org/large-language-models-and-eu-data-protection-mapping-some-of-the-problems/> accessed 5 April 2024; El-Mahdi El-Mhamdi et al., 'On the Impossible Safety of Large AI Models' (arXiv, 9 May 2023) <http://arxiv.org/abs/2209.15259> accessed 10 August 2023; Matthew Sag, 'Copyright Safety for Generative AI' (4 May 2023) <https://papers.ssrn.com/abstract=4438593> accessed 2 August 2023; Laura Weidinger et al., 'Taxonomy of Risks Posed by Language Models', *2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2022) <https://dl.acm.org/doi/10.1145/3531146.3533088> accessed 24 April 2024.

[12] Nicholas Carlini et al., 'Extracting Training Data from Diffusion Models' (arXiv, 30 January 2023) <http://arxiv.org/abs/2301.13188> accessed 25 April 2024; El-Mhamdi et al. (n. 11); Maanak Gupta et al., 'From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy' (2023) 11 *IEEE Access* 80218; Milad Nasr et al., 'Scalable Extraction of Training Data from (Production) Language Models' (2023) *arXiv.org* <https://arxiv.org/abs/2311.17035v1> accessed 1 October 2024.

distributions.[13] Although this phenomenon most often occurs where duplicates exist in the training data, it still appears where training data has been partially deduplicated. Larger models with more parameters 'remember' more data than smaller models.[14] Violations of people's privacy and right to data protection result from both incorrect information and correct information they do not want published.[15] These risks are exacerbated by unregulated and therefore uncontrolled secondary downstream use of the models. In the case of popular LLMs operated by global technology companies, commercial resale seems remote, as the companies have no interest in giving up their exclusive option for commercial exploitation. The situation is different for smaller, but in some cases no less risky models: Mixtral 8x7B competes with and surpasses GPT 3.5 in some respects, due to smart architecture that combines eight different expert models, and has been recently made open source.[16] This only highlights the need for an overview of the purposes for which these models are used and a categorisation to enable a context-based risk assessment.[17]

Data protection law gives rise to its own particular frictions, from the general function and technical specificities of big data applications and generative AI on the one hand and, on the other, the particularities of generative models. Generative models are used in different contexts for different purposes, to generate text, code, video, images, audio, and so on. In this chapter, I will focus on LLMs, which generate text by calculating the probability of word order. Data that is linguistically translatable – that is, can be understood by human recipients – can clearly also contain personal data as covered by data protection law. For this reason, they are a good example of the problems of how data protection law works in relation to AI-generated content.

This chapter first outlines the overarching lines of conflict between data protection law and generative AI (Section 15.2). It then goes into the specific legal issues of the GDPR: the scope and legal basis for authorisation of different steps of data processing by generative AI (Section 15.2), the principles of data processing (Section 15.3), the rights of data subjects (Section 15.4) and questions of responsibility (Section 15.5). Section 15.6 discusses the transferability of the argument to models that create images, audio, and video. The chapter concludes with an outlook (Section 15.7).

15.2 STRUCTURAL CHALLENGES OF GENERATIVE AI FOR DATA PROTECTION LAW

Data protection law in the EU is primarily addressed by the GDPR.[18] The current system of the GDPR is rooted in the Data Protection Directive adopted in 1995,[19] the right to data protection

[13] Claudio Novelli et al., 'Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity' (14 January 2024) <https://papers.ssrn.com/abstract=4694565> accessed 15 January 2024.

[14] Nicholas Carlini et al., 'Quantifying Memorization across Neural Language Models' (*arXiv.org*, 15 February 2022) <https://arxiv.org/abs/2202.07646v3> accessed 29 April 2024.

[15] Rainer Mühlhoff and Hannah Ruschemeier, 'Predictive Analytics and the Collective Dimensions of Data Protection' (2024) 16 *Law, Innovation and Technology* 261.

[16] Philipp Hacker, 'What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models' (2023) *Verfassungsblog* <https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/> accessed 6 January 2024.

[17] Rainer Mühlhoff and Hannah Ruschemeier, 'Regulating AI with Purpose Limitation for Models' (2024) 1 *Journal of AI Law and Regulation* 24.

[18] 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) (Text with EEA relevance) OJ 2016 No. L119, 4 May 2016.

[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 free movement of such data OJ L 281, 23 November 1995, 31–50.

(Article 8 Charter of Fundamental Rights of the European Union, CFR), the right to privacy (Article 7 CFR),[20] and the primary legal foundations in Article 16 Treaty on the Functioning of the European Union (TFEU).[21] Article 1 GDPR sets the matter and scope as the processing of personal data with the objective of protecting fundamental rights and freedoms of natural persons, Article 1 GDPR. Correspondingly, the understanding of 'processing' in terms of personal data is very broad and takes an all-encompassing approach to cover practically any interaction.[22] For this reason, when personal data is involved, all stages in the lifecycle of an AI model may fall within the scope of the GDPR.

From a regulatory perspective, the various steps of data processing in the lifecycle of an AI model are therefore important, and for generative models can be differentiated as follows.[23] The first step is the collection of training data, made up of many data points. These may comprise personal or non-personal information. In certain instances, this process utilises extremely large datasets, making it challenging, if not impossible, to differentiate between various categories of data. For instance, ChatGPT was developed using copious amounts of data freely available on the internet. The second step is the actual training of the model using the collected data, resulting in a configured model. The third step is model application, meaning that the trained model is applied to specific cases or individuals, making the model a tool that computes a specific output in response to input data.[24] This breadth of data and the training process mean model output contains information about cases or individuals as well as of 'third parties' that were not part of the training data.

15.2.1 *Quantity*

The first problem area relates to how the training of powerful AI models, or the processing of large amounts of data, works in relation to the amount of data processed.[25] The sheer quantity of data processed by AI models is the core, as yet unresolved, problem of AI and data protection.[26] Generative AI models are typically trained on billions, if not hundreds of billions, of parameters and require large amounts of training data and computing power.[27] Data protection law on the other hand is based on the idea that the individual steps of data processing and the data processed can be identified. This concept applies the idea of individual control to empower individuals by allowing them to manage their own personal information.[28] But models trained on unprecedentedly large datasets make it impossible to manually identify or even review

[20] I understand privacy as one of the protected interests under data protection law. For more about the relationship, see Johannes Eichenhofer, *E-Privacy: Theorie und Dogmatik eines europäischen Privatheitsschutzes im Internet-Zeitalter*, Vol. 301 (Mohr Siebeck, 2021) 47f. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6638955>.

[21] Additionally, Article 39 TEU lays down specific provisions for the area of the Common Foreign and Security Policy.

[22] See V. P. Hert and P. D. Hert, 'GDPR Art. 4 Abs. 2 Processing' in I. Spiecker gen Döhmann, V. Papakonstantinou, G. Hornung, and P. de Hert (eds.), *General Data Protection Regulation: Article-by-Article Commentary* (Baden-Baden: Nomos Verlagsgesellschaft, 2023), para. 3.

[23] These steps are developed in: Mühlhoff and Ruschemeier (n. 15); Mühlhoff and Ruschemeier (n. 17).

[24] Mühlhoff and Ruschemeier (n. 17).

[25] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2016) 47 *Seton Hall Law Review* 995.

[26] Generative models are based on neural networks, usually Generative Adversarial Networks and Transformer Networks: Weidinger et al. (n. 11).

[27] Tom B. Brown et al., 'Language Models Are Few-Shot Learners' (2020) Neural Information Processing Systems', 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea June 21– 24, 2022 <https://dl.acm.org/doi/proceedings/10.1145/3531146>.

[28] Daniel J. Solove, 'Artificial Intelligence and Privacy' (1 February 2024) <https://papers.ssrn.com/abstract=4713111> accessed 25 March 2024.

whether data processing complies with legal requirements, and thus harbours potential for privacy and data protection violations.[29] Furthermore, this approach conflicts with the principle of data minimisation laid down in Article 5(1)(c). This mode of operation reveals the problems of governance arising from the systematic design of the GDPR, which, for example, envisages individual consent as the basis for authorisation and presupposes the identification of individual data subjects and the data to be attributed to them.

15.2.2 *Purposes*

Privacy and data protection also seem to be at odds with the general concept of generative AI when it comes to the relevance of purposes. Data protection is highly contextual, and its level of protection depends on the type of data processed, by whom, in which settings, and for which purposes (Article 5(1)(b) GDPR). LLMs on the other hand, cover a wide range of purposes, applications, and operating environments. According to Article 3(63) of the new regulation on artificial intelligence[30] (AI Act), a general-purpose AI model includes AI models trained with a large amount of data using self-supervision at scale, which display significant generality, are capable of competently performing a wide range of distinct tasks regardless of how the model is placed on the market, and that can be integrated into a variety of downstream systems or applications. It does not include AI models used for research, development, or prototyping activities before they are placed on the market. This definition is a good description of the current market situation; OpenAI, for example, now offers a wider variety of different GPTs for specific tasks: the laundry buddy for laundry-specific questions about stains and laundry settings, the sous chef that provides users with recipes, or the negotiator that helps a user to argue in their favour.[31] These downstream applications will gain more relevance, as it can be expected that the foundation models will not continue to be used primarily as isolated applications as has been the case to date, but will be integrated into other models as modular building blocks. This will increase both desirable and undesirable effects due to the possible scaling of model output. Here, even the design aspect of LLMs is difficult to reconcile with legislation, seemingly conflicting with the GDPR's purpose limitation principle. In particular, when models are made available to numerous third parties via an interface, ensuring compatibility for that model and its data with the purposes for which the data was originally collected (Article 6(4) GDPR) becomes difficult, if not impossible.[32]

15.3 SCOPE OF THE GDPR AND LEGAL BASIS

The GDPR is applicable in terms of material and geography – that is, it extends to the processing of personal data for activities within the EU, even when that processing takes place elsewhere (Article 3(1) GDPR), and where goods or services are offered to data subjects within the Union (Article 3(2) GDPR). It therefore applies to all generative models in use in the Union

[29] On the relationship between privacy and data protection: Raphael Gellert and Serge Gutwirth, 'The Legal Construction of Privacy and Data Protection' (2013) 29 *Computer Law & Security Review* 522; Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 *International Data Privacy Law* 222.

[30] P9_TA(2024)0138 (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[31] These are available under ChatGPT 4o at chatgpt.com with a paid subscription.

[32] Mühlhoff and Ruschemeier (n. 17).

15.3.1 *Scope of Application*

15.3.1.1 Personal Data

The GDPR applies to the processing of personal data (Article 2(1)) if none of the exceptions in paragraphs 2–3 apply. This processing includes both the collection of training data and the training of the models, as well as the storage and use or sale of the model to generate output based on user requests.

The processing of personal data begins with step one, the collection of vast amounts of data with which to train an LLM. As the effectiveness of LLMs is directly linked to the breadth and variety of their datasets, this data is obtained by scraping content from numerous websites. Inevitably this often includes personal data (Article 4(1) GDPR) such as names, dates of birth, or other identifying information.[33] As personal data also includes incomplete or indirect details which may result in an individual being identified through additional information,[34] this processing is covered by the GDPR, even before the model is trained or released.

In the second step of data processing, the training of the model, identifying personal data becomes more challenging, as the final trained model may differ from its training data. An artificial neural network is represented by a large matrix of numbers, determined by weights and other parameters such as activation thresholds.[35] While the training data may include personal information, the data in the model may not necessarily retain that characteristic: personal data may be anonymised where advanced techniques such as differential privacy and federated machine learning are used during the training process to remove references to the training data.[36]

A trained model resulting from such anonymisation that makes the reconstruction of training data impossible or highly unlikely is not considered to constitute personal data. However, the current popular large language models tend to persist in producing identifying information, whether by design or by accident. It cannot therefore always be assumed that model data has been fully anonymised: research into this 'remembering' phenomenon is ongoing. This is critical from the GDPR standpoint as the storage of the model also constitutes data processing under the GDPR if the model data is not properly anonymised. In addition, many authors[37]

[33] European Data Protection Board (EDPB) (n. 3); Hannah Ruschemeier, 'Squaring the Circle' (*Verfassungsblog*, 7 April 2023) <https://verfassungsblog.de/squaring-the-circle/> accessed 25 March 2024. The Hamburg data protection authority, on the other hand, believes that LLMs do not store personal data, without explaining how this relates to the scraping of information that undoubtedly constitutes personal data: <https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/240715_Diskussionspapier_HmbBfDI_KI_Modelle.pdf> accessed 25 March 2024.

[34] In its judgment on the *Breyer* case (CJEU C-582/14, ECLI:EU:C:2016:779) the ECJ also considered whether personal data is involved in the context of indirect identifiability.

[35] Mühlhoff and Ruschemeier (n. 17) 27; Martín Abadi et al., 'Deep Learning with Differential Privacy' (2016) *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security – CCS'16* 308 <https://dl.acm.org/doi/proceedings/10.1145/2976749>.

[36] Abadi et al. (n. 35); Cynthia Dwork, 'Differential Privacy' in Michele Bugliesi et al. (eds.), *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II*, Vol. 2 (Springer 2006); Khaled El Emam, Sam Rodgers and Bradley Malin, 'Anonymising and Sharing Individual Patient Data' (2015) 350 *BMJ* (Clinical research ed.) h1139.

[37] K. Emam and C. Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2014) 5 *International Data Privacy Law* 73; Wim Jan Schreurs et al., 'Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector' in Mireille Hildebrandt and Serge Gutwirth (eds.), *Profiling the European Citizen* (Springer 2008); Alexander Roßnagel,

argue that the anonymising of personal data itself is also a processing operation that requires justification under the GDPR.[38]

In processing step number three, the production of output, the models or applications using them can produce personal data. Whether the information provided is correct or not is immaterial: when an LLM produces outputs that contain the names and bibliographical information of real people, they are processing personal data. Additionally, individuals can often be easily identified from the context of the text prompt or text output, or by using search engines. LLMs linked to search engines may also facilitate identification. Particularly in the case of public LLMs, it is likely that many data subjects can be identified for the reasons mentioned above.[39] It is important to note that the people in the training data are not theoretically the same as those produced in the output data even where they have the same name, as LLMs can also generate names of existing people, for example by producing information to users can then assign to individuals.

15.3.1.2 Territorial Scope of Application

Article 3(1) of the GDPR states that the Regulation applies 'to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not'. Thus, the processing of personal data does not have to take place in the European Union itself, but can be performed on servers that are for example based in the United States or other third countries. As mentioned above, *lex loci solutionis* (Article 3(2)) means the requirements apply if the data processor offers its services to EU citizens, even where the processor is not located in the EU. This therefore brings global technologies including LLMs and other AI models such as ChatGPT, Bard, and Gemini squarely under the GDPR where these are accessible from the European Union.

15.3.2 *Legal Basis for Data Processing*

All processing of personal data within the scope of the GDPR requires a legal basis (Article 6(1) GDPR). The question of the legal basis for data processing across the life cycle of a generative AI system poses different problems, as it depends on the stage of the data processing. As argued before, it is essential to distinguish between the different steps of data processing when analysing AI and data protection.[40]

15.3.2.1 Collection of Training Data

The first step in the life cycle of a generative model is the collection of training data In the case of LLMs like GPT-4 or Bard, this step consists of scraping data from the internet. The indiscriminate scouring of almost the entire internet logically excludes the legal basis of consent (Article 6(1)(a)). In the absence of legal obligations or contractual relationships between the

'Datenlöschung und Anonymisierung. Verhältnis der beiden Datenschutzinstrumente nach der DSGVO' (2021) 11 *ZD* 188.

[38] Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP 216) 0829/14/EN. On this dispute in the context of predictive analytics: Mühlhoff and Ruschemeier (n. 15).

[39] Paulina Pesch and Raine bec Böhme, 'Verarbeitung personenbezogener Daten und Datenrichtigkeit Bei Großen Sprachmodellen' (2023) 26 *Multimedia und Recht* 917.

[40] Mühlhoff and Ruschemeier (n. 15).

operators of LLMs and all internet users worldwide, the scraping of training data can only rely on the legal basis of legitimate interest provided in Article 6(1)(f) GDPR.[41]

Article 6(1)(f) GDPR states that data processing is lawful if it is necessary for the purposes of pursuing the legitimate interests of the data controller or by a third party, provided these interests do not override the fundamental rights and freedoms of the data subjects requiring protection.[42] The ECJ has clarified that this provision lays down three cumulative conditions: (1) the pursuit of a legitimate interest by the controller or by a third party; (2) the processing of personal data must be necessary to pursue that legitimate interest; and (3) that the legitimate interest of the controller or of a third party are not outweighed by the interests or fundamental freedoms and rights of the data subject.[43]

The fact that this constitutes the only plausible legal basis exposes the structural problem of data protection law in relation to data-intensive technologies, not least because whether Article 6(1)(f) provides a sufficient legal basis must be determined on a case-by-case basis.[44] There are indications that general interest may outweigh the purpose of processing, or that this can be assumed, if data subjects could reasonably expect their data to be processed for training purposes. However, the nature of mass data scraping makes it almost impossible to identify individual interests, and therefore cannot provide satisfactory answers in terms of current legal doctrine and legal systems.

15.3.2.2 Legitimate Interest

The term 'legitimate interests' is deliberately broad, to encompass legal, economic, or idealistic interests, excluding only hypothetical and public interests. The collection of data to train a generative model for commercial use is initially a legitimate economic interest protected by the freedom to conduct a business under Article 16 CFR. The argument that the European Court of Justice (ECJ) also cited freedom of information (Article 11(2) ECFR) as a legitimate interest transferable to generative model training[45] in the Google Spain case[46] does not apply to models that are only accessible for a fee. Furthermore, search engines and generative models operate differently, and are therefore not comparable. Sources referenced in search engines can be deleted or corrected, whereas LLMs generate a unique new text for each question for which a new probability is calculated. If the output text is incorrect, it cannot be corrected for future outputs.

15.3.2.3 Necessity

The necessity test under Article 6(1)(f) requires that the processing of personal data be a proportionate means of achieving legitimate interests. Processing is considered necessary if the processing of personal data is essential to achieve the objective of the processor's legitimate interest – in this case, training and putting an AI model on the market – and that these interests

[41] Zarsky (n. 25); Frederik J Zuiderveen Borgesius et al., 'Tracking Walls, Take-It-or-Leave-It Choices, the GDPR, and the ePrivacy Regulation' (2017) 3 *European Data Protection Law Review* 353.

[42] Ruschemeier (n. 33).

[43] Case C-252/21, Meta ECLI:EU:C:2023:537, para. 106; Case C-597/19, EU:C:2021:492, para. 106.

[44] Mary Donnelly and Maeve McDonagh, 'Health Research, Consent and the GDPR Exemption' (2019) 26 *European Journal of Health Law* 97; Novelli et al. (n. 13).

[45] Christoph Krönke, 'Attention Is All You Need: ChatGPT und die DSGVO' (2023) *Verfassungsblog* <https://verfassungsblog.de/attention-is-all-you-need/> accessed 29 April 2024.

[46] Case C-131/12, Google Spain, ECLI:EU:C:2014:317.

do not outweigh the rights of the data subject. In rare cases, where only anonymised data is sufficient to train the model, that training may not require personal data. However, anonymised data alone is generally not adequate for training generative models, even were such anonymisation possible in the training phase.

15.3.2.4 Balancing of Interests

This balancing of interests between processor and data subject must also consider the rights of data subjects under Articles 7 and 8 ECFR. Their interests are particularly affected when AI collects, combines, and contextualises personal data available on the internet in response to user queries.

There is a valid argument for the interests of the processor in providing a large amount of training data to ensure powerful generative models able to generate word sequences that correspond to human language. Nevertheless, it is not absolutely necessary to scrape data on a scale that covers almost all publicly accessible resources on the internet to develop generative models: datasets can be generated in other ways, such as through data donations, effective consent solutions, or data collection by the data controller itself. However, none of these alternative options would be able to create the required breadth of data. The question is therefore *which* specific interest of the processor is worth protecting. Meta, for example, publicly admitted that the main difficulty the acquisition of licences for copyrighted material would have imposed on the development of generative models was expense. The same argument applies to collecting training data in a manner compliant with privacy regulations: the approach would have required considerable resources. However, it is unlikely that cost savings can constitute a legitimate interest, and at any rate, an interest based on structural infringements has considerably less protective value.

In the Meta case,[47] the ECJ also ruled that the personalisation of content – Meta's core business model – was not necessary for the operation of a social network. The ECJ went on, stating that legitimate interests do not adequately justify Meta's practices of tracking and profiling individuals for the purpose of conducting its behavioural advertising business across its social platforms:

> it is important to note that, despite the fact that the services of an online social network such as Facebook are free of charge, the user of that network cannot reasonably expect that the operator of the social network will process that user's personal data, without his or her consent, for the purposes of personalised advertising. In those circumstances, it must be held that the interests and fundamental rights of such a user override the interest of that operator in financing its activity through personalised advertising, with the result that the processing by that operator for such purposes cannot fall within the scope of point (f) of the first subparagraph of Article 6(1) of the GDPR.[48]

This raises substantial doubts about the ability of companies like OpenAI to defend the processing of vast amounts of personal data to establish a commercial generative AI enterprise, particularly given that such tools pose numerous emerging risks to identified individuals, including issues like disinformation, defamation, identity theft, and fraud.

Context is therefore critical in safeguarding privacy and data protection. The public accessibility of data on the internet, even where disclosed by the data subjects themselves, does not completely

[47] Case C-252/21, Meta ECLI:EU:C:2023:537, para. 102 et seq.
[48] Ibid., para. 117.

negate their legitimate interest in its protection. As Recital 47 states, the interests and fundamental rights of the data subject may in particular override the interest of the data controller where personal data is processed in circumstances or in ways that data subjects do not reasonably expect. Although it is now public knowledge that data posted on the internet may be processed in ways other than initially thought, it is also a question of the specific purposes of the processing. For example, the legitimate expectation of privacy means that decades-old or deleted posts, personal websites, and entries cannot be used in perpetuity to train commercial models. It is reasonable to suggest that the typical internet user does not expect, or intend, their data to be utilised as training material for LLMs for the financial gain of others. Therefore, the use of the data for training these models represents a secondary purpose. In most instances, it is unlikely that a data subject made their data publicly available to serve as a dataset for the financial gain of LLM providers, making the use of such publicly available data an infringement on contextual privacy.[49]

Moreover, a legitimate interest must be determined within the broader European and national regulatory context. The broad scope of scraping also means that an unmanageable number of people are affected, which opens the claim of legitimacy to questions of proportionality. According to the German constitutional doctrine of the Federal Constitutional Court, a particularly large number of people being affected without cause can impact the claim of legitimacy. This impact is referred to as 'scatter width', and is a line of argumentation used by the ECJ.[50] Such effects also arise in the case of universal data processing, as almost all internet users are affected.

Additionally, the legitimate interest must also be lawful, meaning it should conform to all applicable laws and regulations, including the principles and other provisions of data protection law. This includes ensuring that processing aligns with the expectations of the data subject based on their relationship with the controller, adheres to the principles of data minimisation, and implements appropriate safeguards (Recital 50 of the GDPR). In the case of broad-scale internet scraping, individual interests are difficult to identify. However, there were concerns about the legality of scraping from the outset, including with regard to potential copyright infringements.[51] An interest pursued through structural infringements cannot be legitimate.

Compatibility with the principles of data protection law under Article 5 GDPR also plays a role in the balancing of interests. This requires legitimate interests be evaluated in terms of the fairness of processing (Article 5(1)(a)), purpose limitation (Article 5(1)(b)), data minimisation (Article 5(1)(c)), and data accuracy (Article 5(d)).

Therefore, legitimate interests cannot be assumed across all training data. The matter is made more complex given it is extremely difficult, if not impossible, to comprehensively exclude personal data pertaining to minors or special categories of personal data according to Article 9(1) from training data. To complicate this matter further, the point at which the processing of personal data 'reveals' special categories of personal data under Article 9(1) GDPR has not yet been conclusively clarified.

15.3.2.5 Training of the Model

It is worth undertaking a chronological review of the various data processing operations used for training the model. A key consideration is whether data anonymisation occurs during model

[49] Helen Nissenbaum, 'A Contextual Approach to Privacy Online' (2011) 140 *Daedalus* 32.

[50] Case C-511/18 inter alia, ECLI:EU:C:2020:791, marg. no. 143 f. – La Quadrature du Net; Case C-293/12, ECLI:EU:C:2014:238, para. 57 ff. – Digital Rights Ireland Ltd; Case C-203/15, ECLI:EU:C:2016:970, para 105 f. – Tele2; Case C-140/20, ECLI:EU:C:2022:258, marg. No. 66 – Commissioner of An Garda Síochána.

[51] Ruschemeier (n. 33).

training. If the data is anonymised in the course of training, a further processing of anonymous data would fall outside the scope of the GDPR. The prevailing view is that a legal basis is required for the anonymisation of data under Article 6(1).[52] In this context, anonymisation should be understood as normative rather than technical, in line with the ECJ ruling that holds that data has been anonymised. This is because the ECJ considers anonymisation has occurred even if it is technically possible, but unlikely, that the controller could carry out identification with the means available including additional information.[53] According to the court, data is considered anonymous under the GDPR if re-identification is illegal.[54]

In principle, the anonymisation of personal data is generally easy to justify under Article 6 GDPR. The practice aligns with the principle of data minimisation and storage limitation. Effective and permanent anonymisation can serve the interests of both data subjects and data controllers: the former are protected from unauthorised interference with their fundamental data protection rights, while the latter are freed from some of the perceived burdens of complying with the stringent requirements of data protection law.[55] However, this argument struggles to hold in light of the volume of data, as effective consent from the data subjects pursuant to Articles 6(1)(a) and 7 GDPR cannot be obtained in practice.

While it may be possible to institute legal obligations to anonymise training data under Article 6(1)(c), this is not yet relevant in practice. This means that the legal basis of legitimate interest in Article 6(1)(f) may also apply to anonymisation. Generally speaking, this provision may provide adequate results as anonymisation will typically be in the interest of the data subjects themselves, so that at the very least a conflicting interest is improbable. In the case of larger LLMs, it is equally unlikely that a data subject will have an individual interest in non-anonymisation and, moreover, even where such interest of an individual data subject exists, it would outweigh other relevant interests, such as of the other data subjects. As a result, anonymisation is permissible.

Assessing the processing of special categories of personal data under Article 9 is more difficult: anonymisation would require a case to be made under Article 9(2). As described above, the processing of special categories of data cannot be ruled out for LLMs. Although the hurdles of Article 9(2) are high, the 'made public' provision of Article 9(2)(e) can also be considered here. Others argue for a teleological reduction of Article 9(1) for anonymisation.[56] Neither variant constitutes an infringement of data subjects' rights where training data has been anonymised. These complex considerations alone show that there are gaps between the individual-based approach of the GDPR and the tools required for adequately regulating generative AI.

15.3.2.6 Generating Output

The output of generative language models may constitute the processing of personal data. Here, a distinction must be made between the processing of scraped training data and the processing of user data in the form of prompts entered while using the model. There is no legitimate interest

[52] Emam and Alvarez (n. 37); Roßnagel (n. 37); Schreurs et al. (n. 37) 241–270; Article 29 Working Party Article 29 Working Party, 'Opinion 05/2014 on Anonymisation Techniques (WP 216) 0829/14/EN' (Article 29 Working Party 2014) 29.

[53] Case C-582/14, ECLI:EU:C:2016:779 – *Breyer*, para. 45–49.

[54] Ibid. This, however, contradicts the protective function of the GDPR, which aims to protect data subjects from unlawful processing of their data. See Philipp Hacker, 'A Legal Framework for AI Training Data – From First Principles to the Artificial Intelligence Act' (2021) 13 *Law, Innovation and Technology* 257.

[55] Gerrit Hornung and Bernd Wagner, 'Anonymisierung als datenschutzrelevante Verarbeitung? Rechtliche Anforderungen und Grenzen für die Anonymisierung personenbezogener Daten' (2020) *ZD* 223.

[56] Ibid.

in processing user data, for example in the context of input prompts when using LLMs. Instead, effective consent, pursuant to Article 6(1)(a) must be obtained, using a tool that needs to be critically evaluated in the digital space.[57] OpenAI had to update its privacy policy for EU users after being investigated by the Italian data protection authority. It now states: 'We use the content you provide to improve our services, such as to train the models that run our services. Read our instructions on how to opt out of the use of your content to train our models.'[58] However, consent is only a valid basis for prompts containing personal information about the user themselves. If users create prompts that include personal data from other persons, they cannot validly consent on their behalf.[59]

When a generative model is capable of producing output which includes personal data, the issue of how training data was collected remains relevant throughout its lifecycle. If there was no legal basis for collecting the training data, there is no legal basis for using it to generate output. Theoretically, legitimate interest could also be considered here under Article 6(1)(f), but must be assessed on a case-by-case basis according to the criteria described above. However, LLMs make individual assessments difficult because of the quantity of data they process. In addition, generative models are scalable in terms of their output, which means false information can be disseminated to a large number of users and third parties.

Output processing is also problematic in cases where models infer or disclose special categories of personal data under Article 9(1) GDPR. It has been shown that models can memorise and reproduce private and personal information such as phone numbers, addresses, and medical documents.[60] In the age of big data, it is now potentially possible to infer sensitive information from almost any data, especially if one includes the boundless category of political opinions, covered by Article 9(1) GDPR. This means that 'normal' personal data can reveal special categories of personal data covered by Article 9(1), although the criteria for distinguishing between general and sensitive data remain contested.[61] One proposed criterion goes to the intention behind the data processing. Scenarios involving context-specific information could lead to the generation of sensitive data depending on the purpose of evaluation.[62] Court rulings tend to support this assumption: the ECJ seemed to interpret 'revealing' broadly in the *Meta* case,[63] and in another ruling, the Court decided that the disclosure of a spouse, partner, or

[57] Omri Ben-shahar and Carl E. Schneider, 'The Failure of Mandated Disclosure' (2011) 159 *University of Pennsylvania Law Review* 647; Borgesius et al. (n. 41); Sourya Joyee De and Abdessamad Imine, 'Consent for Targeted Advertising: The Case of Facebook' (2020) 35 *AI & Society* 1055; Trung Tin Nguyen, Michael Backes and Ben Stock, 'Freely Given Consent? Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android Apps', *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (Association for Computing Machinery 2022) <https://dl.acm.org/doi/10.1145/3548606.3560564> accessed 13 July 2023; Mühlhoff and Ruschemeier (n. 17) 31.

[58] OpenAI, 'Datenschutzerklärung' (15 December 2023) <https://openai.com/de-DE/policies/eu-privacy-policy/> accessed 4 October 2024.

[59] Novelli et al. (n. 13).

[60] Carlini et al. (n. 12).

[61] Mühlhoff and Ruschemeier (n. 15).

[62] Michael Matejek and Steffen Mäusezahl, 'Gewöhnliche vs. sensible personenbezogene Daten. Abgrenzung und Verarbeitungsrahmen von Daten gem. Art. 9 DS-GVO' (2019) *ZD* 551; Sebastian Schulz, 'Art. 9 DS-GVO', in Peter Gola and Dirk Heckmann (eds.), *Datenschutzgrundverordnung* (Munich: C. H. Beck, 2018); Ludmila Georgieva and Christopher Kuner, 'Art. 9 Processing of Special Categories of Personal Data' in Christopher Kuner et al. (eds.), *The EU General Data Protection Regulation (GDPR)* (Oxford: Oxford University Press, 2020) 373; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Do Large Language Models Have a Legal Duty to Tell the Truth?' <https://papers.ssrn.com/abstract=4771884> accessed 30 April 2024. Question No. 2 in ECJ Case C- 252/21, *Meta Platforms and Others* [2021] (ECLI:EU:C:2022:704), dismissed by the Advocate General in his Opinion, para. 41.

[63] Case C-252/21, *Meta*, ECLI:EU:C:2023:537, para. 73.

cohabitee's name could potentially indicate the sexual orientation of the applicant.[64] The Court has established minimal criteria for what constitutes the 'revealing' of sensitive data: the act of an 'intellectual operation involving comparison or deduction' is deemed sufficient to extend the special protection regime meant for sensitive data to personal data that is not inherently sensitive. However, this judgment was not directly related to big data, leaving the distinction somewhat ambiguous.[65]

Consequently, in many instances involving big data, merely being able to potentially infer sensitive information may subject processes such as AI training to the provisions of Article 9. and there is little likelihood that LLMs satisfy the exceptions in Article 9(2). For instance, the research exemption under Article 9(2)(j) is restricted to the development of models for research purposes and does not permit their commercial exploitation, as indicated in Recitals 159 and 162.[66]

Another important distinction is whether LLM output can be used to infer sensitive information about individuals that they have not made public themselves. Even if certain indicators, for example of political orientation, are available on the internet, LLM output may aggregate this information. As such, Article 9(2)(e) does not constitute a legal basis for this type of derivation.[67]

Data accuracy requirements (Article 5(1)(d) GDPR also apply to LLM output. Language models have been shown to 'hallucinate' and produce incorrect information, including incorrect personal data.[68] Under the GDPR, operators are responsible for ensuring data accuracy (Articles 5(2), 24, 25(1) GDPR).[69] Although all popular applications provide disclaimers to inform users that the models may not always be correct, the effect of such notifications are questionable given *automation bias*.[70] Even if the current error rate of LLMs[71] does not justify generally prohibiting such applications on the basis of ensuring data accuracy, it does affect data subjects' rights. The right to data accuracy becomes even more significant if the right to rectification or erasure cannot be effectively enforced.

15.4 DATA SUBJECTS' RIGHTS

As with other areas of data-intensive technology application, there are problems with the enforcement of data subjects' rights in the case of generative models.[72] In general, the actors involved mean many data-driven AI technologies are developed, promoted, sold, and used by a

[64] Case C-184/20, ECLI:EU:C:2022:601.
[65] Mühlhoff and Ruschemeier (n. 15).
[66] Novelli et al. (n. 13).
[67] See also Case C-252/21, *Meta*, ECLI:EU:C:2023:537, para. 73.
[68] Wachter et al. (n. 62).
[69] As the ECJ already stated in 2008 with regard to the maintenance of data in the Central Register of Foreigners – similar to an actor acting under public authority, in order to delete or correct incorrect information without delay. Case C-524/06, 2008 I-09705. Krönke (n. 45).
[70] Hannah Ruschemeier, 'The Problems of the Automation Bias in the Public Sector – A Legal Perspective' in Weizenbaum Institute (ed.), *Weizenbaum Conference Proceedings: AI, Big Data, Social Media and People on the Move* (2023) <www.weizenbaum-library.de/items/4bc59488-2fa5-4609-9db8-47f2bf1d57f7> accessed 22 August 2023; Lukas Hondrich and Hannah Ruschemeier, 'Addressing Automation Bias through Verifiability' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4521411> accessed 4 October 2024; Hannah Ruschemeier and Lukas J. Hondrich, 'Automation Bias in Public Administration – An Interdisciplinary Perspective from Law and Psychology' (2024) 41 *Government Information Quarterly* 101953.
[71] Cade Metz, 'Chatbots May "Hallucinate" More Often Than Many Realize' *The New York Times* (6 November 2023) <www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html> accessed 30 April 2024.
[72] Daniel Solove, 'The Limitations of Privacy Rights' (2023) 98 *Notre Dame Law Review* 975.

handful of big tech companies, which establishes an informational power asymmetry between the powerful processors and the users. As a result, privacy rights alone are insufficient to address the issue of data disempowerment.[73] Individuals typically are not able to fully manage their personal data, as there is a fundamental limit to the control they can exert. While rights can afford a modest degree of influence in certain isolated cases, this influence is too sporadic and disjointed to significantly safeguard privacy. Ultimately, rights function primarily as a minor element within a broader framework.[74]

The sheer quantity of the data processed from various sources seems to make it impossible to identify and inform individuals of the processing, or of the processor, to enable data subjects to assert their rights with regard to the processing of their data. Therefore, the data quantity realistically rules out compliance with the data subject's right to information. In practice, reporting indicates that companies such as OpenAI and Midjourney have not responded to requests for information from people who found themselves in the training data.[75]

The prerequisite for the exercise of the data subjects' rights under the GDPR provided in Articles 13–22 is, first and foremost, that the data subject is aware of the data processing. Users providing input into an AI model in the form of prompts are covered by Article 13 GDPR. However, Article 14 GDPR also comes into play where data has not been collected from the data subjects themselves. According to both standards, data subjects must be informed about who has processed which data (categories of data), for what purposes, on what legal basis, and whether this data has been disclosed to third parties. These transparency provisions have the specific purpose of enabling data subjects to exercise their other rights, such as the right to erasure or rectification. Article 14(5) states the exception that the transparency obligation does not apply where and insofar as the provision of such information proves impossible or would involve a 'disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes … In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available'.

Again, it depends on the individual case, although it seems doubtful whether LLM operators can invoke unreasonableness if they already knew before the model was developed that individual requests for information could not be enforced. In any case, the principle of responsibility in Article 5(2) GDPR means a failure to respond to such requests or a reference to general impossibility is not sufficient.

Practice has shown that LLMs and possibly other generative AI models that produce content operate almost universally, not just at an individual level. This near-universal infringement reflects the profound mismatch between data-intensive models and the individual rights approach to data protection taken by data protection laws. As a result of this universality, other rights of data subjects such as the right to rectification (Article 16 GDPR)[76] and the right to erasure (Article 17 GDPR) exist on paper, but become unenforceable in practice.[77]

[73] We have explained in detail elsewhere that systemic, supra-individual solutions and a collective understanding of data protection are needed: Mühlhoff and Ruschemeier (n. 15); Mühlhoff and Ruschemeier (n. 17); Rainer Mühlhoff, 'Predictive Privacy: Collective Data Protection in Times of AI and Big Data' (2023) 10 *Big Data & Society* 1.

[74] Solove (n. 72).

[75] Elisa Harlan and Katharina Brunner, 'We Are All Raw Material for AI' (*BR*, 7 July 2023) <https://interaktiv.br.de/ki-trainingsdaten/en/> accessed 4 October 2024.

[76] Novelli et al. (n. 13).

[77] Novelli et al. (n. 13) rightly point out that invoking the right to erasure (and correction under Article 21 GDPR) depends on whether the LLM itself is personal data. This depends whether the training data is anonymised or not, see Section 15.3.2.3.

Furthermore, removal requests from an individual data subject cannot produce the intended outcome, particularly in cases where the same information has been disseminated by multiple users interacting with the LLM.[78] In essence, simply deleting data from a training dataset offers only a superficial remedy, as it does not guarantee the elimination of the ability to retrieve that data or extract related information embedded within the model's parameters.[79] As the output of certain machine learning models is shaped by the data used during the training phase, the original training data or information related to removed data may be deduced or 'leaked'.[80]

15.5 RESPONSIBILITY

In addition to the various steps of data processing in generative models, multiple parties could potentially be considered data controllers under the GDPR due to their levels of involvement. The GDPR establishes three categories of responsibility for data processing in relation to the data subject: controller, processor, and third parties.

The data controller is primarily responsible for compliance with the provisions of the GDPR (Article 5(2)). As defined by Article 4(7) GDPR, the data controller is a 'natural or legal person … which, alone or jointly with others, determines the purposes and means of the processing of personal data'. Article 4(8) goes on to define the processor as a 'natural or legal person … to which the personal data are disclosed, whether a third party or not'. Third parties, on the other hand, are actors other than the data subject, controller, or processor (Article 4(10)).

Prima facie, the legal companies that develop and deploy generative models are data controllers. However, a differentiated picture emerges in the various steps of data processing. Indisputably, companies like OpenAI and Google act as data controllers in relation to the processing steps involved in establishing the parameters for foundational training and storing of the model, given they exclusively determine the modalities of data processing, such as the decision to release a freely accessible LLM. However, in terms of output production, generative models process data based on the prompts from their users. Whether this can be used to establish providers and users as joint controllers within the meaning of Article 26 GDPR remains an open question.

Joint controllership under Article 26 GDPR refers to the situation where two or more controllers jointly determine the purposes and means of data processing. In contrast, the relationship between a controller and a processor (Articles 4 (7, 8) and 28 GDPR) is different, as in this constellation, the processor processes data *on behalf* of, and subject to the instructions of, the controller. Joint controllership is thus a relationship of equality, whereas the data processor operates as a contractor, following instructions issued by the data controller. Whether this structure can be transferred to the relationship between providers of generative models and those who use them is questionable.

Users are not considered processors within the meaning of Article 28 GDPR, as while there is a contract between them and the providers, they do not have the obligations of a processor, especially those imposed by Article 28(3) GDPR, as they may generate prompts at will, and are not processing data according to instruction. The purpose of generative models is to enable users to freely use the model for their own defined purposes, for example to formulate letters, find cooking recipes, or revise texts, free from instructions from the provider.

[78] Novelli et al. (n. 13).
[79] Ibid.
[80] Ibid.

Users and providers could therefore be joint controllers, but this would require them to jointly define the purposes of data processing, and set out transparent mutual obligations. This classification is supported by the fact that users and providers both influence the purposes of data processing: the providers of generative models set the basic framework within which their models are used, while users specify the purposes according to their individual needs. Consequently, both users and providers are interdependent and have a reciprocal effect on data processing. However, this is contradicted by the fact that users also tend to be data subjects, and according to Article 26(3) GDPR, data subjects have the right to bring claims against any of the joint controllers. Although the law does not require joint controllers to hold the same level of responsibility, mere contributory causation without cooperative action is not sufficient for joint responsibility.[81] Additionally, users' limited influence over data processing means users of generative models cannot effectively be held responsible to third parties, as users have no ability to grant rights of access to providers or to delete personal data from the training data.

The relationship between users and providers of generative AI therefore presents a special case that cannot be seamlessly subsumed under the categories of the GDPR. On the one hand, users are more than just data subjects, as their active inputs are required to generate and shape the model's output. On the other hand, they are neither data processors nor joint controllers, as they have no influence over the fundamental modalities of data processing. For instance, providers are able to simply deactivate models or make them subject to fees (as in the case of ChatGPT). The ECJ considers the extent to which data controllers participate in the joint data processing and the specific processing phases within which this occurs to be crucial.[82] For LLMs, users only participate in output generation, which is significantly dependent on the previous steps, such as training. The purpose and aim of the regulations concerning joint controllership is to counteract a diffusion of responsibility among multiple participants. Affected individuals should be able to clearly identify who is collecting their personal data, and for what purpose (Recital 58). Therefore, while providers may be responsible for user-generated content in the case of generative models, the inverse does not apply. This follows from the reasoning and fundamental rights protection of the GDPR provisions regarding responsibility and also corresponds to technical and economic reality.[83]

15.6 IMAGES, AUDIO, AND VIDEOS AS PERSONAL DATA

The considerations for LLMs are not always transferable to generative models that produce audio, images, and video. This is because the aim of these models is not to generate information, which may be incorrect, but to generate new audio or visual material. The primary aim of generating new content has given rise to many copyright issues arising in these cases.[84] Images and videos can also be personal data if they can be used to identify the person, something easily achieved today through image searches.

A major problem is the significant increase in deepfakes in the digital context, which now affects not only public figures but also the general population. Women in particular are often victims of deepfake pornography, where explicit images and videos are generated using their

[81] Case C-2010/16, ECLI:EU:C:2018:388, para. 45.
[82] Ibid.
[83] Similar: Case C-131/12, ECLI:EU:C:2014:317 – Google Spain, para. 33.
[84] Tim W. Dornis and Sebastian Stober, 'Urheberrecht und Training generativer KI-Modelle – Technologische und juristische Grundlagen
' (29 August 2024) <https://papers.ssrn.com/abstract=4946214> accessed 26 September 2024; Sag (n. 11).

images, without their consent.[85] This is an unlawful processing of personal data that violates the GDPR and, in many cases, national regulations. The photographed image of a person in these cases constitutes personal data if the person is still alive, regardless of whether the data is fake or not. The purpose of deepfakes is to disparage or discredit a specific individual, thus fulfilling the decisive characteristic of Article 4(1) GDPR, namely that the person is identified or identifiable. Voices may also constitute personal data, if the person is identifiable; visual or acoustic identification methods recorded using pattern recognition, such as facial or voice recognition (speaker recognition), can even be considered biometric data under Article 4(14) GDPR.[86]

The latest addition to the digital legislation cavalry, the AI Act, only imposes a labelling obligation on deepfakes (Article 50(4)), leaving considerable doubt as to whether there is an adequate level of legal protection at European level.[87]

15.7 CONCLUSION AND OUTLOOK

The popular use cases of generative AI models show that data protection law is reaching its limits when it comes to regulating data-intensive technologies. In addition to the problems highlighted here, further questions arise regarding the principle of purpose limitation for data processing for data-intensive models and their downstream applications. The use of LLMs in decision-making situations raises questions about the scope of the prohibition in Article 22 GDPR.[88]

Structural problems of almost universal concern exist between the GDPR's focus on individual protection and the volume of data processed for training purposes, and in terms of a structural enforcement deficit, particularly regarding data protection principles and data subjects' rights.

As important as the structure of data protection law is for the protection of fundamental rights, new solutions are needed for the structural challenges posed by generative AI and other data-intensive technologies. These may also lie outside data protection law. To address these challenges, it is important to recognise the structural dimension of AI as a socio-technical development.[89] As a result, there is a need for structural solutions that go beyond the enforcement of individual rights.[90] Unfortunately, these issues remain unaddressed, as the AI Act does not contribute solutions to remedy the structural and specific challenges posed by the GPDR's individual rights focus. Despite its stated goal of protecting fundamental rights including data protection, the structure of the AI Act follows product safety law parameters and as such has an approach fundamentally different from legal frameworks aimed at protecting fundamental rights, as found in the GDPR. The AI Act as part of a digital legislation framework that takes

[85] Anne Pechenik Gieseke, '"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography' (2020) 73 *Vanderbilt Law Review* 1479.

[86] Thilo Weichert, 'DS-GVO Art. 4 Nr. 14 Biometrische Daten' in J. Kühling and B. Buchner (eds.), *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO BDSG* (Munich: C. H. Beck, 4th ed., 2024).

[87] Karolina Mania, 'Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings from a Comparative Legal Study' (2024) 25 *Trauma, Violence & Abuse* 117; Lea Katharina Kumkar and Julian Philipp Rapp, 'Deepfakes: Eine Herausforderung für die Rechtsordnung' (2022) 3 *Zeitschrift für Digitalisierung und Recht* (ZfDR) 199; Don Fallis, 'The Epistemic Threat of Deepfakes' (2021) 34 *Philosophy & Technology* 623.

[88] See for example the use of Gemini for decisions about unemployment benefits: <https://gizmodo.com/googles-ai-will-help-decide-whether-unemployed-workers-get-benefits-2000496215> accessed 26 September 2024.

[89] Rainer Mühlhoff, 'Human-Aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Toward a Media Sociology of Machine Learning' (2020) 22 *New Media & Society* 1868.

[90] Mühlhoff and Ruschemeier (n. 15); Mühlhoff and Ruschemeier (n. 17).

a variety of approaches to protect from risks posed by AI does establish certain obligations for generative AI systems under Articles 51–54, like technical documentation (Article 53(1)(a)), a policy to comply with union law (Article 53(1)(b)) and a summary about the content used for training (Article 53(1)(d)), but these provisions do not address the privacy and data protection of users. Additionally, there are broad exceptions for open source models (Articles 2(12) and 53(2)). This still leaves a need for legal regulation.