Universiteit
Leiden
The Netherlands

EU privacy and data protection law applied to AI: unveiling the legal problems for individuals

Häuselmann, A.N.

Citation
Häuselmann, A. N. (2024, April 23). *EU privacy and data protection law applied to AI: unveiling the legal problems for individuals*. Retrieved from https://hdl.handle.net/1887/3747996

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3747996

Note: To cite this publication please use the final published version (if applicable).

4 Legal problems: principles

This chapter aims to answer Subquestion 3, namely, what legal problems arise or may arise when the principles enshrined in the current legal framework are applied to AI. First, this chapter introduces three types of legal problems (Section 4.1). Based on this approach, legal problems are identified for each AI discipline outlined in Chapter 2 (i.e. machine learning, computer vision, natural language processing, affective computing and automated reasoning). This chapter focusses on the *principles* enshrined in the current legal framework. Sections 4.2 to 4.8 deal with the principles enshrined in the GDPR, namely, the principles of lawfulness (Section 4.2), fairness (Section 4.3), transparency (Section 4.4), purpose limitation (Section 4.5), data minimisation (Section 4.6) and accuracy (Section 4.7), as well as the principle of enhancing protection for special categories of personal data (Section 4.8).[633] Section 4.9 elaborates on the requirements with respect to the confidentiality of communication, which is regarded as a principle in a broader sense for the purpose of this thesis. Finally, Section 4.10 concludes by providing an answer to Subquestion 3, including an overview of which AI disciplines lead to which types of legal problems. Whereas AI systems may be deployed by both governmental and private actors, I focus on the latter.

4.1 Approach

When referring to legal problems, three types of legal problems are distinguished (Table 4.1).

Type	Description
1	Legal provisions are violated
2	Legal provisions cannot be enforced
3	Legal provisions are not fit for purpose to protect the fundamental right at stake

Table 4.1 Three types of legal problems.

Let me briefly explain the need to investigate these three types of legal problems in particular. Both the right to privacy and data protection are fundamental rights in the EU.[634] Violations of fundamental rights, which constitute Type 1 legal problems, must be prevented. For example, unsupervised ML approaches process personal data for inexplicit purposes – the processing itself determines the purpose and future use of the personal data. Such processing violates the purpose limitation principle, which constitutes a Type 1 legal problem. Type 2 legal problems, namely, when legal provisions cannot be enforced, are not acceptable either because they lead to negative consequences for the de facto protection of fundamental rights. For example, the unclear substantive meaning of the fairness principle reduces legal certainty and makes it less likely that this principle will be enforced by means

[633] Admittedly, this is not a traditional data protection principle. Nonetheless, it could be regarded as a principle in a broader sense, which then also aligns with the approach taken in this chapter.
[634] Article 7 and 8 CFREU.

of private litigation and by supervisory authorities, which leads to Type 2 problems. Furthermore, Type 3 legal problems, namely, legal provisions that are not fit for purpose, point to the shortcomings of the current legal framework. Legal provisions are not fit for purpose, for instance, when they fail to achieve legislative aims, are not effective or create a gap of protection. For example, the principle that special categories of personal data receive enhanced protection and the legislator's approach to exhaustively enumerate special data cause a Type 3 legal problem. This approach does not keep up with technological developments facilitated by AI. It leads to significant gaps of protection, for example, regarding the processing of new types of sensitive personal data generated by AI, such as emotion data, neurodata and mental data. Insights about this type of legal problems are essential when considering how the legal problems should be addressed, which is the aim of Subquestion 5 (see Chapter 6).

As indicated in Section 1.4, the scope of this thesis is limited to legal problems related to the fundamental rights to privacy and data protection. Thus, this chapter identifies legal problems arising primarily from the perspective of natural persons. Obviously, violations of provisions enshrined in the current legal framework (Type 1) constitute a problem for the natural persons concerned. However, legal problems related to enforcement (Type 2) are not exclusively problematic for natural persons. They also directly concern the competent supervisory authority (SA) tasked with the regulatory enforcement of the provisions enshrined in the current legal framework.[635] When the competent SA is unable to pursue regulatory enforcement, this is not only problematic for the SA itself, but also for the natural persons concerned as they have, in the case of the GDPR, a right to lodge a complaint with a SA.[636] The SA then must handle the complaint and adopt corresponding enforcement measures. Where the complaint lodged by the natural person concerns a substantively unclear provision enshrined in the current legal framework, the SA will not be able to pursue regulatory enforcement. This is problematic for both the SA and the natural person concerned. Type 3 legal problems are discussed from the perspective of *natural persons* as the primary subject of protection envisaged by fundamental rights. These types of legal problem are identified by means of the rationales and specific aims pursued by the current legal framework relevant to natural persons. Table 4.2 lists the rationales and specific objectives[637] enshrined in the current legal framework that are relevant to natural persons. The table only mentions secondary EU law because the fundamental rights to privacy and data protection enshrined in the EU Charter of Fundamental Rights (EUCFR) are less likely to cause Type 3 legal problems due to the flexibility of these rights and the living instrument doctrine adopted by the ECtHR (see also Sections 4.10 and 5.12).

[635] For instance, Supervisory Authorities that have to enforce the GDPR as described in Article 57 GDPR.
[636] Article 77 GDPR.
[637] Expressed in the form of Recitals. For an in depth discussion see Gloria González Fuster, 'Study on the essence of the fundamental rights to privacy and to protection of personal data' (2022) <https://edps.europa.eu/system/files/2023-11/study_en.pdf> accessed 8 February 2024; Gloria González Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Springer International 2014).

GDPR
The protection of natural persons regarding their fundamental right to data protection (Recital 1)
The protection of personal data (Recital 4)
Respect for the fundamental right to privacy (Recital 4)
Processing of personal data to serve mankind (Recital 4)
Consistent and high level of protection for personal data (Recitals 6, 10)
Strong and coherent data protection framework (Recital 7)
Control for data subject over the processing of their own personal data (Recitals 7, 68)
Enhancement of legal and practical certainty for data subjects (Recital 7)
Effective protection and strengthening the rights of data subjects (Recital 11)
Same level of legally enforceable rights (Recital 13)
ePD
Full respect for the fundamental rights to privacy and data protection (Recital 2)
Guaranteeing the confidentiality of communications (Recital 3)
Protection of personal data and the privacy of the user (Recital 5)
Protection of users from risks for their personal data and privacy posed by the Internet and ECS (Recital 6)
Protection of natural persons with respect to automated storage and processing of data (Recital 7)

Table 4.2 Legislative aims pursued by EU secondary law relevant to natural persons. As indicated by Article 1 (2) GDPR, the latter's main goal is to protect the fundamental right to data protection (Article 8 EUCFR).

I do acknowledge that the rationales and specific objectives listed in Table 4.2 are to some extent arbitrary, as they solely focus on the perspective of *natural persons* as the primary subject of protection envisaged by the two fundamental rights I discuss in this thesis. However, neither the fundamental right to privacy nor the fundamental right to data protection are absolute rights. Recital 4 GDPR emphasises that the fundamental right to data protection is not an absolute right, and it must be balanced against other fundamental rights and freedoms. In its case law, also the CJEU stresses the character of this fundamental right is not absolute.[638] Interests and rights of controllers explicitly mentioned in the GDPR's recials[639] are, for instance, the freedom to conduct a business (Article 16 EUCFR), trade secrets that may be protected by the fundamental right to property (Article 17 EUCFR)[640] or intellectual property rights. Hence, Table 4.2 should not be understood as an arbitrary list. It merely contains the rationales of EU secondary law aimed at protecting natural persons in line with the focus and limitations of this thesis (see Section 1.4). Nonetheless, I do take the non-absolute nature of the fundamental right to data protection into account, which becomes particularly apparent

[638] Case C-268/21 *Norra Stockholm Bygg AB* [2023] ECR I-145 para 49; Case C-460/20, *TU* [2022] ECR I-962 para 56; Case C-136/17, *GC and Others* [2019] ECR I-773 para 57.
[639] Recitals 4, 63 GDPR.
[640] Case C-1/11 *Interseroh Scrap* [2012] ECR I-194 para 43; Case T-189/14 *Deza* [2017] para 163.

when discussing legal problems (e.g., Sections 4.2.1, 4.3.1 and 5.6.2). I also consider fundamental rights and freedoms of controllers when suggesting solutions to the legal problems identified (e.g., Sections 6.5.2 and 6.6.2).

As indicated in Sections 1.1 and 1.4, I focus on horizontal relationships. Concerning the fundamental right to data protection, I mostly elaborate on the GDPR when discussing legal problems. Article 1 (2) GDPR reveals the primary goal of this piece of EU secondary law: protecting the fundamental right to data protection according to Article 8 EUCFR. The CJEU emphasises the latter: the GDPR aims to ensure a high level of protection 'of the rights guaranteed in Article 16 TFEU and *Article 8 of the Charter*'.[641] In this sense, the GDPR 'implements'[642] this fundamental right within the realm of horizontal relationships. Whereas the primary goal of the GDPR is to guarantee the fundamental right to data protection,[643] the GDPR contains several more fine-grained objectives, as illustrated in table 4.2. For type 3 legal problems, I use these objectives to assess whether the principles contained in the GDPR are fit for purpose to protect the fundamental right to data protection[644] guaranteed by Article 8 EUCFR.

The three types of legal problems are not mutually exclusive. For example, a Type 2 legal problem may also constitute a Type 3 problem. For example, despite its role as a key tenet in EU data protection law, the substantive meaning of the fairness principle remains largely elusive, meaning it is hard to enforce (i.e. Type 2). At the same time, the fairness principle is *currently*[645] not fit for purpose (i.e. Type 3) to protect the fundamental right to data protection – a substantively unclear principle cannot ensure a high level of the protection of personal data as envisaged in EU data protection law. These three legal problems may be caused by one or more AI disciplines, as described in Chapter 2. Legal problems may be very specific to only one AI discipline or may be more general and relate to several AI disciplines. The latter applies where a provision enshrined in the current legal framework is substantively unclear (e.g., the fairness principle), which causes legal problems regardless of which discipline of AI it is applied to. Also, note that violations of the principles enshrined in the GDPR simultaneously violate the accountability principle introduced in Section 3.3.3.10. According to the

[641] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45 emphasis added by the author; see also Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[642] Article 1 (2) GDPR reveals the main objective of said regulation: to give meaning to this fundamental right. See Hielke Hijmans, Commentary of Article 1 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 56.
[643] Article 1 (2) GDPR.
[644] It could be argued that the GDPR does not protect personal data but rather natural persons. It is apparent from Article 1 (1) that the GDPR protects natural persons. This also follows clearly from the concept of personal data. Protecting personal data as intended by the GDPR (Article 1 Recitals 1, 2, 4, 6, 9, 11, 89 GDPR) indispensably protects natural persons as only information relating to a natural person constitutes personal data.
[645] It is predominantly interpreted as procedural fairness. The fairness principle might be fit for purpose when substantive fairness is added to the current interpretation. See Section 6.2.

accountability principle, controllers are i) responsible for compliance and ii) must be able to demonstrate compliance with *all the principles* mentioned in Article 5 (1) GDPR.[646]

In some cases, it can be difficult to map the legal problems one-on-one with the different AI disciplines, as well as with all provisions contained in the legal framework discussed in Chapter 3. Therefore, I focus on principles enshrined in the legal framework outlined in Chapter 3 (i.e. lawfulness and proportionality, fairness, transparency, purpose limitation, data minimisation, accuracy, the principle that special categories of personal data receive enhanced protection and the principle concerning the confidentiality of communications). Principles form the basis of the fundamental right to data protection[647] and the legislator considers the infringement of principles as more *serious* than infringements of other provisions.[648] The principle of confidentiality contained in the ePD is the key principle ensuring the confidentiality of communications as protected by the fundamental right to privacy. I do not discuss the principle of integrity and confidentiality according to Article 5 (1) lit f GDPR.[649] I also skip the principle of storage limitation according to Article 5 (1) lit e GDPR because it is not particularly relevant in the context of AI.

To determine which type of legal problem arises or may arise due to the different AI disciplines, as outlined in Chapter 2, the AI disciplines are mapped with the principles contained in the current legal framework. For each principle enshrined in the current legal framework, I assess whether the principle at hand creates Type 1, 2 or 3 legal problems. When doing so, I follow the order of the AI disciplines outlined in Chapter 2.

AI refers to adaptive machines that can autonomously execute activities and tasks that require capabilities usually associated with humans. Although AI could make its *own* decisions and perform tasks on the designer's behalf,[650] AI does not have a legal personality. Thus, AI cannot itself cause the three types of legal problems discussed in this thesis. Instead, these legal problems occur when companies use AI. Hence, when concluding that AI causes legal problems, I always refer to the deployment of AI by companies. To unveil the legal problems, I rely on Chapter 2, which explains the different AI disciplines and how they work from a technological and conceptual perspective.

[646] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 311.
[647] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.
[648] Article 29 Working Party, 'Guidelines on the application of administrative fines for the purposes of Regulation 2016/679' (WP 253, 3 October 2017) 9; European Data Protection Board, 'Guidelines on the calculation of administrative fines under the GDPR' (Guidelines 4/2022, 16 May 2022) 16.
[649] For the reasons outlined in Section 3.3.3.8.
[650] Eduardo Alonso, 'Actions and agents' in Frankish Keith and Ramsey William M. (eds) *The Cambridge Handbook of Artificial Intelligence* (2014) 235, 236.

4.2 Lawfulness

As outlined in Section 3.3.3.1, lawfulness essentially requires that processing respects all applicable legal requirements[651] and is further substantiated in Article 6 GDPR. Processing is only lawful if at least one of the lawful bases listed in Article 6 GDPR applies, for example, consent of the data subject (lit a), performance of or entering into a contract (lit b) or the legitimate interest pursued by the controller or third party (lit f).[652] In addition, the principle of lawfulness connotes proportionality in the balancing of interests of data subjects and controllers.[653] Thus, the principle of lawfulness is closely linked to the principle of proportionality, which is one of the general principles of EU law[654] and has decisive influence on the assessment of whether a violation of a person's right to data protection is justified.[655] Thus, as already outlined in Section 3.2.2, the principle of proportionality plays an important role in EU data protection law.[656] It has generally three components which involve the assessment of a measure's (i) suitability, (ii) necessity and (iii) proportionality stricto sensu.[657] When the principle of lawfulness (and proportionality) is applied to the AI disciplines introduced in Chapter 2, Type 1 legal problems may occur.

4.2.1 Legal problems: Type 1

As explained in Section 2.1, AI refers to adaptive machines that can autonomously execute activities and tasks that require capabilities usually associated with humans. However, the GDPR does not apply to AI as such because AI does not have a legal personality. Instead, the GDPR applies to controllers and processors deploying AI systems that process personal data. Due to its autonomous and adaptive characteristics, AI has the potential to decide why and how to process personal data. With this, I do not suggest that AI systems currently can act as controllers under data protection law by determining the purposes and means as well as the legal ground for processing. Instead, I refer to the possibility that the use of AI by controllers might violate the principle of lawfulness due to the current reasoning deficiencies in the AI discipline of automated reasoning. I will illustrate this through the legal ground of legitimate interest and the deployment of unsupervised machine learning.

When the processing of personal data is based on the legal ground of the legitimate interest of the controller, the latter has to perform a Legitimate Interest Assessment (LIA).[658] This LIA requires assessing the impact of processing on the fundamental rights and freedoms of the data subject by

[651] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 314.
[652] Article 6 (1) lit a) to f) GDPR.
[653] Lee A. Bygrave, *Data Privacy Law: An International Perspective* (OUP 2014) 148.
[654] Article 5 of the consolidated version of the Treaty Establishing the European Community [2006] OJ C321E/37.
[655] Charlotte Bagger Tranberg, 'Proportionality and data protection in the case law of the European Court of Justice' (2011) Vol 1 No 4 International Data Privacy Law 239-249.
[656] Ibid.
[657] Ibid.
[658] As required by Article 6 (1) lit f GDPR

considering the nature of personal data, the way in which the information is being processed, the reasonable expectations of the data subjects, and the status of the controller and the data subject.[659] It also includes the controller's obligation to consider the proportionality of processing.[660] Before implementing an AI system, the controller needs to perform a LIA[661] and determine the input and training data to be used by the AI system. However, the AI system should be able to perform a LIA if it deploys unsupervised ML. Unsupervised ML approaches process data for *in*explicit purposes – the processing *itself* determines the purpose since its goal is to detect patterns and correlations, gain knowledge, and make accurate predictions. Also, the purpose may alter given that algorithms used in AI learn and develop over time[662] (see also Section 4.5.1). The performance of an LIA is inextricably linked to the purpose of processing because it must be assessed whether the purpose serves a legitimate interest of the controller.[663] However, in the case of unsupervised ML, the specific purpose for processing is not necessarily known in advance.

Current AI systems have been called clueless[664] to understand cause and effect and devoid of common sense.[665] The lack of progress in providing general automated common sense reasoning capabilities underscores that this is a very difficult problem in the field of AI.[666] Common sense reasoning is not just the hardest problem for AI, it is also considered to be the most important problem.[667] It seems that humans are much better then machines in this context[668] and therefore, common sense reasoning still constitutes a challenge in AI,[669] and particularly in automated reasoning (see Section 2.2.5). Apparently, there is no AI system today that has a semblance of common sense or has capabilities such as human cognition. Hence, AI systems are unable to think in a manner on par with human thinking[670] and may therefore not be capable (at least not in the near future) of appropriately weighing the

[659] Art 29 Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (WP 217, 9 April 2014) at 36.
[660] Ibid at 33.
[661] If processing should occur based on the controller's legitimate interest.
[662] Norwegian Data Protection Authority, 'Artificial Intelligence and Privacy' (2018) 4 <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf> accessed 8 February 2024.
[663] Art 29 Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (WP 217, 9 April 2014) at 24.
[664] Brian Bergstein, 'What AI still can't do' MIT Technology Review (Cambridge 31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> accessed 8 February 2024.
[665] Cade Metz, 'Paul Allen Wants to Teach Machines Common Sense' *The New York Times* (New York, 28 February 2018) <https://www.nytimes.com/2018/02/28/technology/paul-allen-ai-common-sense.html> accessed 8 February 2024.
[666] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Hughe Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 405.
[667] Gary Marcus, Ernest Davis, *Rebooting AI: Building Artificial Intelligence we can trust* (Pantheon Books 2019).
[668] Davide Castelvecchi, 'AI pioneer: The dangers of abuse are very real' *Nature* (London, 4 April 2019) < https://www.nature.com/articles/d41586-019-00505-2> accessed 8 February 2024.
[669] Shoham Yoav et al, 'The AI Index 2018 Annual Report' (AI Index Steering Committee Stanford University 2018) 64 <https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf > accessed 8 February 2024.
[670] Lance Eliot, 'AI Ethics And The Quagmire Of Whether You Have A Legal Right To Know Of AI Inferences About You, Including Those Via AI-Based Self-Driving Cars' *Forbes* (New York, 25 May 2022) < https://www-forbes-com.cdn.ampproject.org/c/s/www.forbes.com/sites/lanceeliot/2022/05/25/ai-ethics-and-the-quagmire-of-whether-you-have-a-legal-right-to-know-of-ai-inferences-about-you-including-those-via-ai-based-self-driving-cars/amp/> accessed 8 February 2024.

fundamental rights and freedoms of the parties involved and implement the factors which must be considered according to the LIA.

This holds particularly true because the CJEU has been criticised for shortcomings in identifying the various elements that need to be balanced when assessing the proportionality of data processing based on a controller's legitimate interest and the data subject's right to data protection.[671] Indeed, the early practice of the CJEU concerning the interpretation of data protection law and the proportionality principle often left it up to the national laws, authorities and courts to carry out any concrete proportionality testing.[672] It can be said that the proportionality test is, cognitively, a difficult task due to the lack of clear elements that need to be considered within this assessment. Additionally, there seems to be a lack of concrete proportionality tests performed by the CJEU that could serve as training data for AI to learn and extract the logic of such balancing tests. As is the case with the purpose limitation and data minimisation principle,[673] computer scientists would need measurable definitions of the proportionality principle and concrete indications of how to practically and concretely implement its requirements.

In fact, the accountability principle, which is substantiated in Article 24 GDPR, requires controllers to 'implement appropriate and effective measures to ensure and be able to demonstrate' that personal data processing occurs in accordance with the rules set out in the GDPR.[674] Violations of the lawfulness principle simultaneously violate the accountability principle as introduced in Section 3.3.3.10 because controllers must be able to demonstrate compliance with *all the principles* mentioned in Article 5 (1) GDPR.[675]

The balancing problem (Type 1)

Due to the reasoning deficiencies in the AI discipline AR combined with the lack of computable requirements concerning the proportionality principle, AI systems that autonomously process personal data cannot appropriately balance the fundamental rights and freedoms and assess the proportionality of processing as required by Article 6 (1) lit f GDPR. Such processing violates both the lawfulness and proportionality principle.

[671] Audrey Guinchard, 'Taking proportioanality seriously: The use of contextual integrity for a more informed and transparent analysis in EU data protection law' (2018) Vol 24 Iss 6 European Law Journal 434, 435. For references to such criticism see footnote 5 and 6 in the latter publication.
[672] Charlotte Bagger Tranberg, 'Proportionality and data protection in the case law of the European Court of Justice' (2011) Vol 1 No 4 International Data Privacy Law 239, 242.
[673] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 58 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[674] Art. 24 (1), Recital 74 GDPR.
[675] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 311.

4.2.2 Legal problems: Type 2

When the lawfulness principle is applied to the AI disciplines introduced in Chapter 2, no specific Type 2 legal problems arise. This is mainly due to the reason that the lawfulness principle is substantively clear, as is further substantiated in Article 6 GDPR, which exhaustively enumerates six lawful bases that can be relied upon for the processing of personal data. Nevertheless, the CJEU has been criticised for shortcomings in identifying the various elements that need to be balanced when assessing the proportionality of processing.[676] These shortcomings could lead to Type 2 legal problems because substantively unclear principles are difficult to enforce. However, this problem arises regardless of whether the processing involves AI and thus does not relate specifically to AI. Therefore, I refrain from discussing this problem further.

4.2.3 Legal problems: Type 3

Similar to what I have outlined in Section 4.2.2, no specific Type 3 legal problems arises when the lawfulness is applied to AI, mainly because this principle is substantively clear from a legal point of view. It may be argued that the proportionality principle is not fit for purpose to protect the fundamental right to data protection due to the lack of clarity in terms of the various elements that need to be balanced. Likewise, it is questionable whether consent is a suitable concept to prevent the data subject from harm relating to the processing of personal data. However, these are general issues and therefore not specifically related to AI. Therefore, it will not be discussed further.

4.3 Fairness

The AI disciplines outlined in Section 2.2 create legal problems when applied to the fairness principle introduced in Section 3.3.3.2.[677] In academia, scholars seem to distinguish between two different types of fairness. According to Graef, Clifford and Valcke, *procedural fairness* in data protection law refers to formal or process-oriented requirements.[678] In the view of De Terwangne, procedural fairness considers whether or not the data involved have been obtained nor otherwise processed through unfair means, by deception or without the knowledge of the individual concerned.[679] Malgieri adds *substantive fairness* aiming to prevent adverse effects in concrete circumstances, in particular when

[676] Audrey Guinchard, 'Taking proportionality seriously: The use of contextual integrity for a more informed and transparent analysis in EU data protection law' (2018) Vol 24 Iss 6 European Law Journal 434, 435. For references to such criticism, see Footnotes 5 and 6 in the latter publication.
[677] Parts of Section 4.3 and Section 6.2 resulted in a publication see Andreas Häuselmann, Bart Custers, 'Substantive fairness in the GDPR: Fairness Elements for Article 5.1a GDPR' (2024) Vol 52 Computer Law & Security Review 105942.
[678] Inge Graef, Damien Clifford, Peggy Valcke, 'Fairness and enforcement: bridging competition, data protection, and consumer law' (2018) Vol 8 No 3 International Data Privacy Law 200, 203.
[679] Cecile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 314.

conflicting interests need to be balanced.[680] However, as pointed out in Section 3.3.3.2, the role and meaning of the fairness principle in data protection law remains elusive despite the fact that it is considered to be a key tenet of EU data protection law.[681] In addition, the CJEU has never defined the fairness principle nor the notion of fairness in data protection law.[682] Dictionaries define the term 'fairness' as 'impartial or just treatment or behaviour without favouritism'[683] or as 'the quality of treating people equally or in a way that is right or reasonable'.[684] Both regulatory guidance[685] and regulatory enforcement on EU level in the form binding decisions[686] adopted by the European Data Protection Board (EDPB)[687] identify key elements of the fairness principle. These key elements are: autonomy of data subjects with respect to data processing, their reasonable expectations, ensuring power balance between controllers and data subjects, avoidance of deception as well as possible adverse consequences of processing and ensuring ethical and truthful processing.[688] Despite the close and evident link[689] with the transparency and lawfulness principle, the fairness principle should be interpreted as having an independent meaning going[690] beyond transparency and lawfulness.[691]

Substantive fairness focusses on the adverse effects for data subjects caused by the processing of personal data and also considers the substantial circumstances and interests at stake: expectations of data subjects, effects on them, and the actual interests of the parties involved. Hence, it aims to mitigate unfair imbalances among interests of controllers and data subjects[692] which seems to be more

[680] Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 2, 3 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024.
[681] Damian Clifford, Jef Ausloos 'Data Protection and the Role of Fairness' (2018) Vol 37 No 1 Yearbook of European Law 130, 187.
[682] Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 6 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024.
[683] See < https://www.oxfordlearnersdictionaries.com/definition/english/fairness?q=fairness > accessed 8 February 2024.
[684] See <http://dictionary.cambridge.org/dictionary/english/fairness> accessed 8 February 2024.
[685] European Data Protection Board, 'Guidelines on Article 6(1)(b) GDPR' (Guidelines 2/2019, 8 October 2019), at 6; European Data Protection Board, 'Guidelines on Article 25 Data Protection by Design and Default' (Guidelines 4/2019, 20 October 2020), at 17 and 18.
[686] Article 65 GDPR.
[687] The EDPB consists of representatives of national EU Supervisory Authorities (SAs) responsible for data protection and the European Data Protection Supervisor (EDPS).
[688] Binding Decision 3/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Facebook service (Art. 65 GDPR), adopted on 5 December 2022 paras 103, 219-220, 222-223, 226 478; Binding Decision 4/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Instagram service (Art. 65 GDPR), adopted on 5 December 2022 paras 106, 223-224, 226-227, 445; Binding Decision 5/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its WhatsApp service (Art. 65 GDPR), adopted on 5 December 2022.
[689] Article 5 (1) lit a GDPR mentions the three different principles together.
[690] Binding Decision 3/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Facebook service (Art. 65 GDPR), adopted on 5 December 2022 paras 220, 477; Binding Decision 4/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Instagram service (Art. 65 GDPR), adopted on 5 December 2022 paras 224, 444; Binding Decision 5/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its WhatsApp service (Art. 65 GDPR), adopted on 5 December 2022.
[691] Winston J Maxwell, 'Principle-based regulation of personal data: the case of 'fair processing' (2015) Vol 5 No 3 International Data Privacy Law 205, 208.
[692] Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 10 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024; Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 78.

helpful when compared to procedural fairness focussing on fair ways of obtaining personal data. Substantive fairness also relates to the proportionality principle discussed in Sections 3.2.2 and 4.2.1 which requires controllers to balance the interests at hand and aims to limit the impact for the data subject caused by the processing of personal data. The CJEU uses fairness as an interpretative tool in order to balance the different interests at hand.[693] A fair balance requires specific consideration of the substantial circumstances and interests at issue.[694] The CJEU stresses the particular consideration of the data subject's interests: 'that balance may however depend, in specific cases, on the nature of the information in question and its sensitivity for the data subject's private life'.[695] Both regulatory guidance[696] and regulatory enforcement on EU level[697] point to substantive fairness by mentioning *reasonable expectations* of the data subjects, possible *adverse consequences* of processing and effects of *power imbalance* as some of the key elements of the fairness principle.

Admittedly, the following analysis of the fairness principle in Sections 4.3.1- 4.3.3 might appear quite pessimistic. This is mainly due to the *current* elusiveness surrounding this principle. I explicitly use 'current' because the fairness principle has significant potential to contribute to effective protection for individuals in the context of processing related to AI *if* interpreted substantively. Principles are open norms that allow judges to adjust the law to changing circumstances and to address contemporary problems. As open norms, principles are well suited to recalibrate data protection legislation to changing technological circumstances for achieving the goals set out by the fundamental right to data protection, including legislative goals pursued by the GDPR.[698] The fairness principle's broad scope and open texture[699] make it a suitable candidate to host normative parameters beyond transparency.[700] In Section 6.2, I discuss the fairness principle's potential to contribute to effective protection for individuals by focussing on substantive fairness.

[693] Case C-275/06 *Promusicae* [2008] ECR I-00271 paras 68, 70; Joined Cases C-92/09 and C-93/09, *Schecke* [2010] ECR I-662 para 88; Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 10 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024.
[694] Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 6 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024.
[695] Case C-131/12, *Google Spain* [2014] ECR I-317 para 81.
[696] European Data Protection Board, 'Guidelines on Article 6(1)(b) GDPR' (Guidelines 2/2019, 8 October 2019), at 6.
[697] Binding Decision 3/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Facebook service (Art. 65 GDPR), adopted on 5 December 2022 paras 219-220; Binding Decision 4/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Instagram service (Art. 65 GDPR), adopted on 5 December 2022 paras 223-224, 226; Binding Decision 5/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its WhatsApp service (Art. 65 GDPR), adopted on 5 December 2022.
[698] For example a consistent and high level of protection for personal data (recitals 6 and 10), a strong and coherent data protection framework (recital 7) and effective protection (recital 11) GDPR.
[699] Lee A Bygrave, 'Minding the machine v2.0: The EU General Data Protection Regulation and Autmated Decision Making' in Karen Yeung and Martin Lodge (eds) *Algorithmic Regulation* (OUP 2019) 260.
[700] Lee A Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' (2021) University of Oslo Faculty of Law Legal Studies Research Paper Series No. 202-35 at 22, 23 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3721118> accessed 8 February 2024.

4.3.1 Legal problems: Type 1

AI increasingly contributes to automated decision-making (ADM). Whereas humans have been conditioned to look for causes ('why'), AI algorithms focus on correlations and probabilities ('what').[701] Current AI systems have been called to be clueless[702] to understand cause and effect and to be devoid of common sense.[703] It seems that humans are much better then machines in this context.[704] Common sense reasoning still constitutes a challenge in AI applications.[705] Apparently, there is not one AI system today which has a semblance of common sense comparable to humans. Hence, AI is unable to think in a manner on par with human thinking[706] which is underscored by the shortcomings in automated reasoning as outlined in Section 2.2.5. The lack of progress in providing general automated common sense reasoning capabilities underscores that this is a very difficult problem in the field of AI.[707] Common sense reasoning is not just the hardest problem for AI, it is also considered to be the most important problem.[708]

As outlined in Section 2.2, the term 'learning' in the context of ML does not mean 'understanding', but is about making computers modify or adapt their actions based on experience so that these actions are more accurate.[709] One of the basic skills of ML is generalisation. Generalisation, however, does not go beyond correlation and neglects reason and drawing distinctions. The AI Index acknowledges that common sense reasoning capabilities and deep natural language understanding are still a challenge in AI applications.[710] Probabilistic predictions and generalisation in the context of ML raise concerns regarding the fairness principle. It seems questionable whether ADM and automated predictions based on ML are fair for the data subjects when the algorithms *generalise* but do not *distinguish*.

[701] Viktor Mayer-Schönberger; Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Houghton Mifflin Harcourt 2013) 14, 18.
[702] Brian Bergstein, 'What AI still can't do' MIT Technology Review (Cambridge 31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> accessed 8 February 2024.
[703] Cade Metz, 'Paul Allen Wants to Teach Machines Common Sense' *The New York Times* (New York, 28 February 2018) <https://www.nytimes.com/2018/02/28/technology/paul-allen-ai-common-sense.html> accessed 8 February 2024.
[704] Davide Castelvecchi, 'AI pioneer: The dangers of abuse are very real' *Nature* (London, 4 April 2019) < https://www.nature.com/articles/d41586-019-00505-2> accessed 8 February 2024.
[705] Shoham Yoav et al, 'The AI Index 2018 Annual Report' (AI Index Steering Committee Stanford University 2018) 64 <https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf > accessed 8 February 2024.
[706] Lance Eliot, 'AI Ethics And The Quagmire Of Whether You Have A Legal Right To Know Of AI Inferences About You, Including Those Via AI-Based Self-Driving Cars' *Forbes* (New York, 25 May 2022) < https://www-forbes-com.cdn.ampproject.org/c/s/www.forbes.com/sites/lanceeliot/2022/05/25/ai-ethics-and-the-quagmire-of-whether-you-have-a-legal-right-to-know-of-ai-inferences-about-you-including-those-via-ai-based-self-driving-cars/amp/> accessed 8 February 2024.
[707] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Hughe Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 405.
[708] Gary Marcus, Ernest Davis, *Rebooting AI: Buidling Artificial Intelligence we can trust* (Pantheon Books 2019).
[709] Steven Marsland, *Machine Learning: An Algorithmic Perspective* (2nd edn Chapman & Hall 2015) ch 1.2.1.
[710] Shoham Yoav et al, 'The AI Index 2018 Annual Report' (AI Index Steering Committee Stanford University 2018) 64 <https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf > accessed 8 February 2024.

Lack of reasoning capabilities can lead to unfair decisions, and ADM based on ML can even be discriminatory. For example, the Google AI system developed to recognise child abuse wrongfully classified a father as criminal. Because his toddler had an infection on his genitals, the father took a photo, displaying himself and the infected part of the toddler's body, as advised by a nurse who said such a photo is necessary for the doctor in order to prepare for the corresponding emergency online consultation.[711] This example clearly points to the problem that ML, which is used by Google in this particular AI system, generalises but does not distinguish. ML does not understand what it classifies as 'wrong' or 'right' and neglects the context of a given picture. In this case, this wrongful classification as child abuser had severe consequences for the individual in question. The police opened an investigation and issued search warrants served on Google and his Internet service provider. Furthermore, Google disabled the account of the father, who lost all his emails, contact information and his Google Fi account, meaning he had to obtain a new phone number with another provider.[712] Thus, the wrongful and fully automated classification as a criminal (child abuser) had adverse and detrimental effects for the data subject, leaving no doubt that such processing violates the fairness principle when interpreted as 'substantive fairness' (see Section 6.2). Computational model constructions are often based on assumptions that turn out not to be true in practice.[713] ML produces probable yet inevitably uncertain knowledge and may identify significant correlations.[714] Even if strong correlations are found in datasets, this uncertain knowledge generalises by forming groups but does not distinguish between the members of this group. Data about individuals are full of correlations, but only some of these correlations meaningfully reflect the individual's actual capacity, needs or merits.[715]

This may lead to the situation that individuals are being unfairly treated, as explained in the child abuser example. In addition, it is highly doubtful whether it is fair to act upon probabilistic predictions and correlations deployed by means of ML. Actions taken based on probabilistic predictions and correlations may have real impact on human interests[716] (e.g., to receive a loan or to get a job). This holds particularly true where such predictions or correlations are essentially considered as *facts.* When individuals are treated based on simplified models or classes, concerns regarding the accuracy principle arise. It is clear that accuracy is a distinct principle, and I will discuss this separately in

[711] Kashmir Hill, 'A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as A Criminal' *The New York Times* (New York, 21 August 2022) < https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html > accessed 8 February 2024.
[712] Kashmir Hill, 'A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as A Criminal' *The New York Times* (New York, 21 August 2022) < https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html > accessed 8 February 2024.
[713] Toon Calders, Indrė Žliobaitė, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures' in Bart Custers et al (eds) Discrimination and Privacy in the Information Society (Springer 2013) 45.
[714] Brent Daniel Mittelstadt et al, 'The ethics of algorithms: Mapping the debate' (2016) Vol 3 Iss 2 Big Data & Society 1, 4.
[715] Betsy A Williams, Catherine F Brooks, Yotam Shmargad, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) Vol 8 Journal of Information Policy 78, 82–83.
[716] Brent Daniel Mittelstadt et al, 'The ethics of algorithms: Mapping the debate' (2016) Vol 3 Iss 2 Big Data & Society 1, 5; Solon Barocas, 'Data Mining and the Discourse on Discrimination' (2014) <https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf> accessed 8 February 2024.

Section 4.7. However, even if a prediction is entirely accurate from a mathematical and statistical perspective, treating individuals based on this prediction may still be unfair. Predictions generated by ML are probabilistic and relate to future conduct that has not yet happened or may never happen at all. From this perspective, applying predictions to individuals may be unfair because predictions do not reflect reality and are thus no 'facts.'

Probabilistic predictions and correlations produced by ML may thus have adverse effects on data subjects when treated as facts and, therefore, violate the fairness principle enshrined in EU data protection law. Furthermore, it seems difficult to argue that processing complies with the fairness principle when the AI system does not understand why certain patterns or correlations exist, although these patterns or correlations build the basis of ADM. With ADM generated by means of ML, the underpinning rationale of the decision is not articulated and perhaps not even known.[717] When combined with the reasoning and common sense deficiencies relating to the AI discipline of automated reasoning (see also Sections 2.2.5, 4.4.1 and 4.7.1) processing of personal data inherent to ADM seems to have substantial potential to be detrimental, discriminatory, unexpected or misleading for the data subjects concerned.

The probability problem (Type 1)

ML generates uncertain knowledge, such as predictions and correlations that are probabilistic. This may be unfair because ML mainly generalises and does not articulate the rationale of generated outputs due to the deficiencies in AR. When such outputs are essentially considered as facts, e.g. in the context of ADM, this can have adverse and detrimental effects for data subjects (e.g., when applying for a loan). This violates the fairness principle.

Face recognition systems as described in Section 2.2.3.1 and 2.2.3.2 related to computer vision might violate the principle of fairness. This is particularly due to the opacity of such systems as they may be used without any intention of or cooperation with data subjects.[718] Both the European Data Protection Board and the European Data Protection Supervisory have called for a general ban on any use of AI for automated recognition of human features such as faces in publicly accessible spaces.[719] Covert use of face recognition systems (see Section 2.2.3.1) is not only problematic in the context of law enforcement, but also when used by private actors. [720]

[717] Sue Newell, Marco Marabelli, 'The Crowd and Sensors Era: Opportunities and Challenges for Individuals, Organizations, Society, and Researchers' (ICIS, Auckland, December 2014) 11 < https://www.researchgate.net/publication/288239046_The_crowd_and_sensors_era_Opportunities_and_challenges_for_individuals_organizations_society_and_researchers > accessed 8 February 2024.
[718] Council of Europe, Consultative Committee of Convention 108, 'Guidelines on Facial Recognition' (28 January 2021) at 11 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.
[719] European Data Protection Board and European Data Protection Supervisor, 'Joint Opinion on the Artificial Intelligence Act' (Joint Opinion 5/2021) at 32.
[720] Council of Europe, Consultative Committee of Convention 108, 'Guidelines on Facial Recognition' (28 January 2021) at 11 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.

In the Netherlands, one supermarket chain has used facial recognition technology to prevent theft. All faces of customers who entered the supermarket were registered and consequently checked against a database containing faces of individuals who had previously been banned from entering the supermarket.[721] In Spain, a similar case occurred where a supermarket relied on a facial recognition system to identify individuals who had previously committed crimes in its stores and were banned from entering.[722] It seems questionable whether such processing is 'fair' for the data subjects concerned because it might have adverse effects on the data subjects. If detected by the system, a data subject might be confronted with the police and in any case be publicly exposed to other customers of the supermarket and very likely to be suspected of having committed a crime. Substantive fairness would require striking a fair balance between the interests at hand, namely the goal of the supermarket to prevent theft and the interests of the concerned data subjects. As outlined by the CJEU, this balance also depends on the nature of the information in question and its sensitivity for the data subject's private life.[723] The consideration of the data subjects fundamental rights to privacy and data protection would arguably outweigh the interest of the supermarket to prevent theft considering the intrusive nature of face recognition systems and the corresponding sensitivity for data subjects. In addition, applying the proportionality principle (Section 3.2.2) to this case would arguably lead to the same result.

The facial recognition problem (Type 1)

When covertly applied, face recognition systems powered by the AI discipline computer vision may violate the fairness principle due the intrusive nature of such systems and the corresponding sensitivity for the data subjects concerned, e.g., to be suspected of theft by default and/or to be publicly exposed as a criminal.

In particular, processing of personal data occurring in the context of affective computing (AC) raises the question whether such processing complies with the fairness principle. As pointed out in the probability problem, it is clear that accuracy is a distinct principle that merits dedicated analysis (Section 4.7). Nonetheless, treating individuals based on inaccurate personal data can still be unfair in the context of the fairness principle.

Generally, processing of emotion data enabled by means of AC could be misleading, specifically because the accuracy of outputs generated by AC has been questioned[724] (see also Section 4.7.1). For

[721] The Dutch Data Protection Supervisory Authority has issued a formal warning against this supermarket-chain, see < https://autoriteitpersoonsgegevens.nl/en/current/dutch-dpa-issues-formal-warning-to-supermarket-for-use-of-facial-recognition-technology > accessed 8 February 2024.
[722] Summary of Spanish SA Decision PS/00120/2021 < https://gdprhub.eu/index.php?title=AEPD_(Spain)_-_PS/00120/2021 > accessed 8 February 2024.
[723] Case C-131/12, *Google Spain* [2014] ECR I-317 para 81.
[724] Kate Crawford et al, 'AI Now Report' (2018) AI Now Institute 8 <https://ainowinstitute.org/publication/ai-now-2018-report-2> accessed 8 February 2024; Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1; Sara Preto, 'Emotion-reading algorithms cannot predict intentions via facial

this thesis, emotion data are defined as information related to emotions of an individual ('emotion data'). Emotions refer to the six most-used emotion categories[725] in emotion research: anger, disgust, fear, happiness, sadness and surprise.[726] These six 'basic emotions'[727] are further described in Section 2.2.4.1. Processing of emotion data by AC could be both detrimental and unexpected for the individuals concerned. Imagine an employer that uses automated video assessments such as HireVue[728] to detect emotional states of applicants during these assessments. In particular, in these circumstances, processing of emotion data by means of AC might have adverse consequences for the data subject. Perhaps for precisely this reason, HireVue discontinued the use of the component of its services that analyses facial expressions of applicants.[729]

It has been argued that it should be prohibited to link recognition of emotions to the hiring of staff because it poses risks of great concern on both societal and individual levels.[730] Whereas prohibition seems to be a very restrictive measure, it is certainly valid to question the fairness of using information about the emotional states of individuals in an employment context. Considering the questionable accuracy of AC, the non-transparent manner of processing (candidates do not get to know which emotions the system detected), the sensitive nature of the personal processed (see Section 4.8.3) and the possible adverse effects for the applicant, it seems reasonable to conclude that such processing does not comply with the fairness principle. The asymmetrical power relations between employers and applicants also plays a role. When deciding to rely on AC-powered video assessments during the recruitment process to detect the applicants emotional state, the employer takes advantage of its stronger position. Substantive fairness aims to balance precisely these kind of power asymmetries and to prevent adverse effects in concrete circumstances.[731] Here, the adverse effects are obvious. Arguably inaccurate and rather sensitive personal data are processed to determine whether the applicant will receive a job offer. Undoubtedly, the latter decision has a considerable effect on the applicant.

expressions' *USC News* (Los Angeles, 4 September 2019) <https://news.usc.edu/160360/algorithms-emotions-facial-expressions-predict-intentions/> accessed 8 February 2024.

[725] These six emotions refer to research conducted by psychologists in the early seventies that developed the methodology of 'basic emotions'; see Paul Ekman, Wallace v Friesen, 'Constants across cultures in the face and emotion' (1971) Vol 17 (2) Journal of Personality and Social Psychology 124.

[726] Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 52.

[727] Eiman Kanjo et al, 'Emotions in context: examining pervasive affective sensing systems, applications, and analyses' (2015) Vol 19 Personal and Ubiquitous Computing 1197, 1204 <https://link.springer.com/content/pdf/10.1007/s00779-015-0842-3.pdf> accessed 8 February 2024.

[728] Nathan Mondragon, Clemens Aicholzer, Kiki Leutner, 'The Next Generation of Assessments' (HireVue 2019) <http://hrlens.org/wp-content/uploads/2019/11/The-Next-Generation-of-Assessments-HireVue-White-Paper.pdf> accessed 8 February 2024.

[729] Will Knight, 'Job Screening Service Halts Facial Analysis of Applicants' *Wired* (New York, 12 January 2021) < https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants/ > accessed 8 February 2024. However, other providers offer similar services. HumeAI provides AI-powered tools helping recruiters to assess personality traits as well as emotional states of candidates; see < https://hume.ai/products/facial-expression-model/ > and < https://gethume.com/blog5/artificial-intelligence-for-recruiting > accessed 8 February 2024.

[730] Council of Europe, Consultative Committee of Convention 108, 'Guidelines on Facial Recognition' (28 January 2021) at 3 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.

[731] Gianclaudio Malgieri, 'The concept of Fairness in the GDPR' (FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, January 2020) 2, 3 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3517264> accessed 8 February 2024.

Video assessments powered by AC are only one of many possible examples. The use of AC might also be unfair within other domains including marketing, customer service, healthcare, insurance, retail, autonomous driving, education and gaming.[732] Thus, the use of AC in important sectors is bound to increase, as will the possibility of adverse consequences for data subjects. Although AC systems are predominantly developed in the United States, they are being sold to global marketplaces. Corresponding algorithms are hardly tweaked for racial, cultural, ethnic or gender differences.[733]

The fairness principle is prone to be violated due to the questionable accuracy of emotion data and the sensitive nature of the personal data disclosed and otherwise processed in the context of AC. As outlined in the probability problem, ML generates predictions and establishes correlations that are probabilistic and thus constitute uncertain knowledge. This means that also the output generated by means of ML can violate the fairness principle and the accuracy principle (see also Sections 4.3.1 and 4.7.1). Furthermore, such processing is likely to be detrimental to the interest of the data subject because revealing such sensitive information can very well be used to manipulate a data subject. According to research in behavioural sciences, especially psychology, emotions are powerful, pervasive and predictable drivers of human decision-making.[734]

The inaccuracy problem (Type 1)

The questionable accuracy of personal data generated by the AI disciplines AC and ML violate the fairness principle as the processing of inaccurate personal data is detrimental and misleading to the data subject.

The sensitivity problem (Type 1)

AC allows for predicting and disclosing sensitive emotion data in ways that violate the fairness principle because the subsequent use of such personal data is detrimental to the data subject, particularly in situations entailing power asymmetries, and because emotion data may be used to manipulate the data subject.

4.3.2 Legal problems: Type 2

As indicated in Section 4.3, the fairness principle has thus far managed to remain elusive despite the fact that the fairness principle is considered to be a key tenet of EU data protection law. Apart from obvious examples (such as discrimination), it largely remains unclear when processing of personal

[732] Cem Dilmegani, 'Top 24 Affective Computing (Emotion AI) Use Cases in 2023' <https://research.aimultiple.com/affective-computing-applications/> accessed 8 February 2024; Deepanshu Gahlaut, 'Top Emotion AI Companies to Watch out for in 2023' <https://deepanshugahlaut.medium.com/top-emotion-ai-companies-to-watch-out-for-in-2023-db925868fd9f> accessed 8 February 2024.
[733] Peter Mantello, Ho Manh-Tung, 'Why we need to be weary of emotional AI' (2022) AI & Society <https://link.springer.com/article/10.1007/s00146-022-01576-y> accessed 8 February 2024.
[734] Jennifer S Lerner et al, 'Emotion and Decision Making' (2015) Vol 66 Annual Review of Psychology 799, 802.

data is unfair or results in unfair consequences.[735] The fairness principle enshrined in EU data protection law lacks sufficient precision due to the absence of corresponding case law. Because the CJEU did not yet rule on the substantive meaning of the fairness principle, it is difficult to enforce the fairness principle in practice. This holds true in particular for private enforcement pursued by data subjects or actors mentioned in Article 80 GDPR that represent data subjects, such as non-profit bodies or organisations. Principle-based regulation requires controllers to make a judgement what they must do to comply and to perform risk assessments.[736] When performing such risk assessments, controllers will not only take the risks for the data subject into consideration, but also focus on interpretive risk[737] and any associated risk from enforcement action in case of non-compliance. Admittedly, the fairness principle's elusive role is not a problem caused explicitly by AI. Rather, it exists due to the lack of interpretative guidance by the CJEU. However, legal problems relating to the fairness principle are AI-specific because AI leads to many fairness issues.[738]

The unclear substantive meaning of the fairness principle reduces legal certainty and makes it less likely that it will be enforced by means of litigation in front of the courts. This is proven by means of a complete lack of case law with respect to the substantive meaning of the fairness principle on the level of the CJEU. Only one request for a preliminary ruling[739] dealt with the fairness principle, in which the CJEU ruled that 'fair processing' requires a public authority to inform the data subjects of the transfer of their personal data to another public authority that would process these data for its own purposes.[740] However, this case solely underscores the close link between the fairness and transparency principle, but does not provide any guidance with regard to the substantive meaning of the fairness principle. In case of shortcomings related to the interpretation of core provisions such as the fairness principle, compliance is a matter of risk management, and non-compliance becomes an option.[741] Controllers can assess what level of non-compliance they are prepared to risk and what the potential cost of enforcement action and reputational damage may be in case of non-compliance.[742]

It may be easy to access and read the controller's privacy notice, but it is an entirely different task to verify whether the statements made in the privacy notice are in fact honoured[743] and to what extent the fairness principle is complied with, in the case of complex AI systems in particular. Even if the

[735] Damian Clifford, Jef Ausloos 'Data Protection and the Role of Fairness' (2018) Vol 37 No 1 Yearbook of European Law 130, 187.
[736] Julia Black, 'Forms and paradoxes of principles-based regulation' (2008) Capital Markets Law Journal Vol 3 No 4 425, 454.
[737] For instance, the likelihood that the interpretation of the principle will be approved by supervisory authorities or courts.
[738] For example due to reasoning AI's deficiencies, AI enabled manipulation as discussed in Section 4.3.3.
[739] Case C-201/14 *Bara and others* [2015] ECR I-638 para 34.
[740] Tim van Canneyt et al, 'Data Protection: CJEU case law review – 1995-2020' (2021) Vol 56 Computerrecht 78, 102.
[741] Julia Black, 'Forms and paradoxes of principles-based regulation' (2008) Capital Markets Law Journal Vol 3 No 4 425, 454.
[742] Ibid.
[743] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 60 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.

constraints concerning technological complexity are overcome, the legal uncertainty concerning the interpretation of the principle remains. This leaves considerable discretion to the controllers on how to interpret and apply the fairness principle. Once challenged in regulatory and private enforcement, it is likely that controllers defend such interpretation rigorously and aim to reach precedents which serve their interests.[744] This is underscored by Meta's announcement to appeal both the substance and the fines of the final decisions adopted by the Irish SA based on the EDPB's binding decisions[745] which substantively interpreted the fairness principle for the first time in regulatory enforcement.

As outlined in Section 4.3.1, AI systems may process personal data in a way which is detrimental, unexpected or misleading to the data subject, ultimately resulting in unfair processing. The elusive role and meaning of the fairness principle reduces legal certainty, although the GDPR particularly aims to enhance *legal* and *practical* certainty for data subjects (Recital 7). The elusive role makes it difficult for data subjects and supervisory authorities to challenge the fairness of processing activities enabled by AI. The fact that AI and its underlying models are likely protected by trade secrets or IP laws makes this enforcement problem even bigger (see Section 5.6.2). This Type 2 legal problem occurs regardless of which AI discipline the fairness principle is being applied to because the problem is caused by the substantively unclear meaning of the fairness principle. It is therefore a general problem and relates to all AI disciplines.

The elusiveness problem (Type 2)

AI systems are likely to process personal data in a way that would typically be considered as unfair. The elusive role and meaning of the fairness principle reduces legal certainty and makes it difficult for data subjects to challenge the fairness of processing enabled by AI systems and enforce the fairness principle accordingly.

It could be argued that the meaning of the fairness principle is substantiated by means of regulatory enforcement at the EU level in the form binding decisions.[746] The EDPB identified the following key elements: autonomy of data subjects with respect to data processing, their reasonable expectations, ensuring power balance between controllers and data subjects, avoidance of deception as well as possible adverse consequences of processing and ensuring ethical and truthful processing.[747] However, the mentioning of these key elements in the EDPB's binding decisions does not establish legal certainty. These elements reflect the view of the EU's supervisory authorities (SAs). Meta has announced

[744] This does not seem to be unrealistic considering the financial resources well-known technology companies have and the legal expertise of which they can afford to make use.
[745] See < https://about.fb.com/news/2023/01/how-meta-uses-legal-bases-for-processing-ads-in-the-eu/ > accessed 8 February 2024.
[746] Article 65 GDPR.
[747] Binding Decision 3/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Facebook service (Art. 65 GDPR), adopted on 5 December 2022 paras 103, 219-220, 222-223, 226 478; Binding Decision 4/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its Instagram service (Art. 65 GDPR), adopted on 5 December 2022 paras 106, 223-224, 226-227, 445; Binding Decision 5/2022 on the dispute submitted by the Irish SA on Meta Platforms Ireland Limited and its WhatsApp service (Art. 65 GDPR), adopted on 5 December 2022.

to appeal both the substance and the fines of the final decisions adopted by the Irish SA based on the EDPB's binding decisions.[748] As the controller, Meta has a right to an effective judicial remedy against the legally binding decision adopted by the Irish SA according to Article 78 (1) GDPR. As emphasised by the CJEU, the purpose of Article 78 GDPR is to examine the lawfulness of the decision adopted by a SA.[749] The Irish Court has full[750] and exclusive jurisdiction and needs to review the legality of the Irish SA's final decisions as well as the EDPB's binding decision.[751] Full jurisdiction in this context means the power to examine all questions of fact and law relevant to the dispute[752] and thus includes the question of law on how to interpret the fairness principle. It seems highly likely that the Irish Court will refer questions for a preliminary ruling to the CJEU regarding the contested decisions. In fact, it will be required to do so given the complete lack of judicial guidance regarding the interpretation of the fairness principle. The key elements of the principle of fairness mentioned by the EDPB have not yet been judicially tested. Thus, the Irish Court will arguably have doubts regarding this interpretation of the fairness principle and refer the matter to the CJEU. Hence, it may take several years until the CJEU rules on the matter. Consequently, the elusiveness of the fairness principle remains, which is notably detrimental to the GDPR's aim to enhance *legal* and *practical* certainty for data subjects (Recital 7).

4.3.3 Legal problems: Type 3

The conclusion reached in Section 4.3.2 that the substantive meaning of the GDPR's fairness principle[753] remains largely elusive and provides controllers with significant discretion on how to apply it in practice also leads to a Type 3 legal problem. Due to the lack of clarity concerning the scope and meaning of the fairness principle, the latter is *currently* not fit for purpose to protect the fundamental right to data protection for several reasons. However, as it becomes apparent from Section 6.2, I acknowledge the fairness principle's enormous potential for effective protection for individuals in an AI context if interpreted substantively.

A substantively elusive and unenforceable principle fails to achieve the GDPR's aim to establish a strong and coherent data protection framework[754] when considering that the principles provide the basis for the protection of personal data[755] in the GDPR. It also cannot ensure a consistent and high

[748] See < https://about.fb.com/news/2023/01/how-meta-uses-legal-bases-for-processing-ads-in-the-eu/ > accessed 8 February 2024.
[749] Case C-132/21 *Nemzeti* [2023] ECR I-2 para 35.
[750] Ibid para 41.
[751] Case T-709/21 *WhatsApp Ireland* [2022] ECR T-783 para 70.
[752] Case C-132/21 *Nemzeti* [2023] ECR I-2 para 41.
[753] I need to emphasise that I write about the GDPR's fairness principle. I do acknowledge that the concept of fairness is a constitutive element of the fundamental right to data protection according to Article 8 EUCFR ('fair processing').
[754] Recital 7 GDPR.
[755] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.

level of protection for personal data.[756] In addition, it harms the legal and practical certainty for data subjects.[757] Most importantly, the fairness principle fails to provide data subjects with *effective* protection.[758] In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect the data subject's personal data against risk of misuse.[759] In this thesis, I interpret the risk of misuse broadly, referring to any unlawful use of personal data[760] to the detriment of natural persons concerned by it. A substantively elusive principle cannot prevent misuse in the form of processing of personal data that is detrimental to the data subject's interests and be perceived as unfair.

Manipulation is a typical example of personal data being processed to the detriment of the interests of the data subject. Although it is often not defined in work on the ethics of manipulation,[761] manipulation refers to hidden acts with the aim to intentionally and covertly influence a natural person by targeting and influencing this person's decision-making vulnerabilities.[762] Typically, such influence is against this person's self-interest.[763] Put simply, it perverts the way a person reaches decisions, forms preferences or adopts goals.[764] These acts are not only used to influence what the individual decides or does, but also to influence what the individual thinks or feels, i.e. the individuals thoughts.[765] Whereas manipulation is certainly not a new phenomenon, AI and particularly the disciplines ML and AC introduce new and dedicated means to manipulate decisions, behaviour and thoughts of individuals. AI powerfully enhances the range of influence that companies have in shaping behaviour and thoughts of individuals.[766] It can modify the options and choices available to individuals to manipulate their behaviour. Options or choices available to these individuals may be amended[767] to steer behaviour towards particular goals that are not for the benefit of the individuals

[756] Recitals 6, 10 GDPR; Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[757] Recital 7 GDPR.
[758] As envisaged by Recital 11 GDPR see also Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[759] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[760] See ECtHR case law to which the CJEU refers in the rulings contained in the previous footnote: *S. and Marper v United Kingdom* App no 30562/04 and 30566/04 (ECtHR 4 December 2008) para 99; *Liberty and Others v United Kingdom* App No 58243/00 (ECtHR 1 July 2008) paras 62-63; *Rotaru v Romania*, App No 28341/95 (ECtHR 4 May 2000) paras 57 to 59.
[761] Anne Barnhill, 'What is Manipulation?' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 52.
[762] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 4.
[763] Anne Barnhill, 'What is Manipulation?' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 53.
[764] Joseph Raz, *The Morality of Freedom* (OUP 1986) 377; Cass R Sunstein, 'The Ethics of Nudging' (2015) Vol 32 Yale Journal of Regulation 413, 444.
[765] Anne Barnhill, 'What is Manipulation?' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 57.
[766] Helen Fay Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2010) 83.
[767] Ruth Faden, Tom Beachamp, Nancy King, *A History and Theory of Informed Consent* (Oxford University Press 1986) 355.

concerned, but rather for the benefit of the company which deploys the AI system.[768] There is evidence of how intelligent artificial agents can significantly control human behaviour,[769] going clearly beyond what was previously possible. Research exploring whether it is possible for machines to learn how to influence humans indicates that by means of a computational framework based on reinforcement learning (see Section 2.2.1.3) and ANN approaches (see Section 2.2.1.4), the choices of individuals in particular decision-making tasks can be shaped toward actions or goals desired by the actor exercising influence.[770] In experiments, the machine learnt from participants' responses and identified and targeted vulnerabilities in their decision-making. These vulnerabilities were then successfully used by the machine to steer the participant's decision-making towards particular actions.[771]

It is evident that user interactions with AI-powered systems whose design has been informed by behavioural science lead to behavioural change.[772] Behavioural science concerns the study of behavioural insights and establishes a reliable understanding of behaviour and how it changes. With this information, accurate predictive models can be created.[773] That AI-powered systems, developed with insights from behavioural science, cause change of preference is less evident.[774] Preferences influence behaviour, but behaviour often predates and leads to the emergence of new preferences.[775] ML systems change not only user behaviour, but also user preferences.[776] To intentionally influence preferences of individuals severely impacts their personal autonomy.[777] The essence of autonomy is indicated by the etymology of the term: autos (self) and nomos (rule or law).[778] The ruling idea of personal autonomy is 'that people should make their own lives'[779] which means facing freely both existential and every day's choices.[780] Obviously, changing preferences of individuals influences or even violates personal autonomy as preferences no longer stem from the individuals themselves.

[768] Christopher Burr, Nello Cristianini, James Lydmann, 'An Analysis of the Interaction Between Intelligent Software Agents and Human Users' (2018) Vol 28 Minds and Machines 735, 744, 769; Christopher Burr, Nello Cristianini, 'Can machines read our mind?' (2019) Vol 29 Iss 3 Minds and Machines 461, 4464.
[769] Christopher Burr, Nello Cristianini, James Lydmann, 'An Analysis of the Interaction Between Intelligent Software Agents and Human Users' (2018) Vol 28 Minds and Machines 735, 752.
[770] Amir Dezfouli, Richard Nock, Peter Dayan, 'Adversarial vulnerabilities of human decision-making' (2020) Vol 117 Iss 46 PNAS, 29221-29228.
[771] Jon Whittle, 'AI can now learn to manipulate human behaviour' *The Conversation* (London, 18 February 2021) <https://theconversation.com/ai-can-now-learn-to-manipulate-human-behaviour-155031> accessed 8 August 2021.
[772] Matija Franklin et al, 'Recognising the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of AI' (2022) <https://arxiv.org/pdf/2203.10525.pdf> 1 accessed 8 February 2024.
[773] Susan Michie, Maartje M van Stralen, Robert West, 'The behaviour change wheel: A new method for characterising and designing behaviour change interventions' (2011) Vol 6 Implementation Science 1-12 <https://implementation-science.biomedcentral.com/articles/10.1186/1748-5908-6-42> accessed 8 February 2024.
[774] Matija Franklin et al, 'Recognising the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of AI' (2022) <https://arxiv.org/pdf/2203.10525.pdf> 1 accessed 8 February 2024.
[775] Dan Ariely, Michael I Norton, 'How actions create - not just reveal – preferences' (2007) Vol 12 Iss 1 Trends in Cognitive Sciences 13-16.
[776] Matija Franklin et al, 'Recognising the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of AI' (2022) <https://arxiv.org/pdf/2203.10525.pdf> 1 accessed 8 February 2024.
[777] Matija Franklin et al, 'The EU's AI Act needs to address critical manipulation methods' *The OECD.AI Policy Observatory* (Paris, 21 March 2023) <https://oecd.ai/en/wonk/ai-act-manipulation-methods?utm_source=substack&utm_medium=email> accessed 8 February 2024.
[778] Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge University Press 1988) 12, 18.
[779] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1986) 369.
[780] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 8.

Affective computing (AC) elevates the means to manipulate individuals to an even higher level. Emotions play an important role in the elicitation of autonomous motivated behaviour.[781] According to research in behavioural sciences, especially psychology, emotions constitute powerful, pervasive and predictable drivers of decision-making.[782] Emotions can have significant effects on economic transactions and play a powerful role in decision-making, reasoning[783] and everyday economic choices.[784] Because AC provides access to emotion data of individuals, it may affect people's decisions and lives in unprecedented ways. This is particularly true with regard to manipulation that operates based on facts about the subject's psychology, such as knowledge of its emotions and desires.[785] Three field experiments that reached more than 3.5 million individuals found that their behaviour can be significantly altered, measured by clicks and purchases, when provided with psychologically tailored advertisements.[786] Thus, AI and specifically the disciplines ML and AC exhibit unprecedented means to manipulate the behaviour and thoughts of individuals. Manipulations advance the manipulator's interest at the expense of the manipulated person.[787] An individual's choices, preferences and thoughts can be manipulated[788] to the detriment of individuals, which undermines or violates their personal autonomy.[789] Information regarding the emotional state of an individual might be particularly helpful to manipulate this individual because emotions play an important role in the elicitation of autonomous motivated behaviour.[790] According to research in behavioural sciences, especially psychology, emotions constitute powerful, pervasive and predictable drivers of decision-making.[791] Emotions can therefore have significant effects on economic transactions and play a powerful role in everyday economic choices.[792] Thus, manipulation enabled by AI systems harms the personal autonomy of the individuals concerned by changing their behaviour and preferences as well as by affecting their capacity for reflective choice.[793]

[781] Leen Vandercammen et al, 'On the Role of Specific Emotions in Autonomous and Controlled Behaviour' (2014) Vol 28 Iss 5 European Journal of Personality 413, 445.
[782] Jennifer S Lerner et al, 'Emotion and Decision Making' (2015) Vol 66 Annual Review of Psychology 799, 802.
[783] Steffen Steinert, Orsolya Friedrich, 'Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces' (2020) Vol 26 Science and Engineering Ethics 351, 352.
[784] Jennifer S Lerner, Deborah A Small, George Loewenstein, 'Heart Strings and Purse Strings' (2004) Vol 15 No 5 American Psychology Society 337-340.
[785] J S Blumenthal-Barby, 'A Framework for Assessing the Moral Status of Manipulation' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 123, 127.
[786] Sandra Matz et al, 'Psychological targeting as an effective approach to digital mass persuasion' (2017) Vol 114 No 48 PNAS 12714-12719.
[787] Moti Gorin, 'Towards a Theory of Interpersonal Manipulation' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 124; James Stacey Taylor, Practical Autonomy and Bioethics (Rougledge 2009) 81.
[788] Hildebrandt Mireille, Koops Bert-Jaap, 'The challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) Vol. 73 (3) The Modern Law Review 428, 435.
[789] Newell Sue, Marabelli Marco, 'Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of datafication' (2015) Vol. 24 Iss. 1 The Journal of Strategic Information Systems 4.
[790] Leen Vandercammen et al, 'On the Role of Specific Emotions in Autonomous and Controlled Behaviour' (2014) Vol 28 Iss 5 European Journal of Personality 413, 445.
[791] Jennifer S Lerner et al, 'Emotion and Decision Making' (2015) Vol 66 Annual Review of Psychology 799, 802.
[792] Jennifer S Lerner, Deborah A Small, George Loewenstein, 'Heart Strings and Purse Strings' (2004) Vol 15 No 5 American Psychology Society 337-340.
[793] Matija Franklin et al, 'The EU's AI Act needs to address critical manipulation methods' *The OECD.AI Policy Observatory* (Paris, 21 March 2023) <https://oecd.ai/en/wonk/ai-act-manipulation-methods?utm_source=substack&utm_medium=email> accessed 8 February 2024.

Because the substantive meaning of the fairness principle remains elusive, it seems unclear whether processing personal data enabled by AI systems that deploy ML and AC approaches to manipulate the behaviour of individuals would, in fact, be considered as violating the fairness principle. The latter is *currently*[794] not fit for purpose to effectively protect[795] data subjects, which is detrimental to their interests and thus unfair. In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect the data subject's personal data against the risk of misuse.[796] A substantively elusive and unenforceable principle leads to legal uncertainty. The principle hardly prevents misuses such as manipulations enabled by the AI disciplines AC and ML because this legal uncertainty is likely to be exploited by controllers. Thus, the elusiveness of the fairness principle fails to protect[797] data subjects from such practices effectively. It also fails to achieve other legislative aims of the GDPR, namely, to ensure a consistent and high level of protection for personal data,[798] a strong and coherent data protection framework[799] and legal and practical certainty for data subjects.[800] The substantively elusive fairness principle is also not fit for purpose to ensure that processing of personal data is designed to serve mankind[801] because it does not prevent the manipulation of data subjects and similar practices.

As the introduction (Section 4.3) indicates, this section's analysis and conclusions might appear rather negative. However, I acknowledge the fairness principle's considerable potential[802] to protect individuals from risks caused by AI effectively. I will discuss this thoroughly in Section 6.2.

Arguably, other areas of law, consumer protection law in particular, might be better equipped to prevent manipulation. Whereas this is generally a rightful observation, it should be noted that the processing of personal data enabling, for instance, the detection of emotional states of individuals is primarily governed by the GDPR. However, the fairness principle as it currently stands does not

[794] Because it is predominantly interpreted as procedural fairness and as a mere proxy for transparency see Section 6.2.
[795] Recital 11 GDPR; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[796] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[797] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[798] Recitals 6, 10 GDPR; Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[799] Recital 7 GDPR.
[800] Recital 7 GDPR.
[801] Recital 4 GDPR.
[802] As also pointed out by Bygrave see Lee A Bygrave, 'Minding the machine v2.0: The EU General Data Protection Regulation and Autmated Decision Making' in Karen Yeung and Martin Lodge (eds) *Algorithmic Regulation* (OUP 2019) 260; Lee A Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' (2021) University of Oslo Faculty of Law Legal Studies Research Paper Series No. 202-35 at 22, 23 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3721118> accessed 8 February 2024.

effectively protect[803] data subjects from such processing and has its limitations due to the elusive meaning of this principle. The subsequent use of personal data generated by means of AC and ML might, in some cases, fall under the scope of consumer law, for instance, if the use of such information would be considered an unfair commercial practice. However, it is questionable whether current EU consumer law is, in fact, capable of dealing with such practices. This is indicated by the fitness check on EU consumer law launched by the European Commission in May 2022 which focusses on digital fairness.[804] Irrespective of the outcome of this fitness check, it is important that the fairness principle, as an overarching principle, ensures that personal data are not processed to the detriment of the data subjects concerned.

The manipulation problem (Type 3)

The AI disciplines AC and ML enable controllers to manipulate data subjects by intentionally and covertly exploiting their behaviour, preferences, thoughts and decision-making vulnerabilities, which can be perceived as unfair. Due to the unclear substantive meaning of the fairness principle, it remains unclear whether such processing actually violates the fairness principle. Therefore, the fairness principle is not fit for purpose to effectively protect the fundamental right to data protection and prevent misuses such as manipulations.

In addition, the unclear substantive meaning of the fairness principle also is at odds with the accountability principle which aims to strengthen the responsibility of controllers when they process personal data.[805] The accountability principle enshrined in Article 5 (2) GDPR states that the controller shall be i) responsible for compliance and ii) able to demonstrate compliance with all the principles mentioned in Article 5 (1) GDPR.[806] Shortcomings with regard to the substantive meaning of the fairness principle makes it primarily difficult for controllers to ensure compliance with it. This also affects the data subjects. Requiring controllers to demonstrate compliance with substantively unclear provisions not only fails to effectively protect[807] data subjects. It also fails to establish the responsibility and liability of controllers[808] by imposing legally enforceable obligations on controllers.[809] The accountability principle and related Article 24 GDPR demand controllers to comply with the fairness principle whose actual substantive meaning remains largely unclear. Consequently, controllers cannot, as intended, be held accountable and responsible for complying with it. This holds true regardless of which

[803] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73; Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[804] See < https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en > accessed 8 February 2024.
[805] Recital 74 GDPR.
[806] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 311.
[807] Recital 11 GDPR; Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[808] Recital 74 GDPR.
[809] Recital 13 GDPR.

AI discipline the fairness principle is being applied to. Because controllers cannot be held responsible for failing to comply with obligations that are substantively unclear, the accountability principle misses its aim.[810] This negatively affects the envisaged high level of protection[811] as well as the strong data protection framework[812] intended by the GDPR. Thus, the elusiveness of the fairness principle sabotages the accountability principle. Requiring controllers to demonstrate compliance with a substantively unclear principle is not fit for purpose to effectively protect the fundamental right to data protection. This constitutes a Type 3 legal problem. It is a reoccurring problem because the accountability principle requires compliance with all principles enlisted in Article 5 (1) GDPR. Admittedly, the sabotage problem as described here is not a problem of AI in particular, but one created by the principles contained in the GDPR.

The sabotage problem (Type 3)

Since the substantive meaning of the fairness principle remains largely unclear, it sabotages the accountability principle. Because the accountability principle demands controllers to comply with a substantively unclear principle, it is not fit for purpose to protect the fundamental right to data protection. A principle demanding compliance with substantively unclear provisions cannot hold controllers responsible, nor can it effectively protect data subjects and ensure a high level of data protection.

4.4 Transparency

As outlined in Section 3.3.3.3, the transparency principle enshrined in Article 5 (1) GDPR requires that personal data be processed in a 'transparent manner'. Recital 39 GDPR clarifies that it must be 'transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed.'[813] Articles 13 and 14 GDPR implement the transparency principle and impose specific information duties on the controller. In view of the EDPB, these provisions are the concretisation of the transparency principle, and violations of these provisions may also amount to the violation of the transparency principle itself.[814] As will be discussed in this section, the AI disciplines introduced in Chapter 2 create legal problems concerning the transparency principle itself as well as the specific information duties imposed on controllers.

[810] To hold controllers responsible see Recital 74 GDPR.
[811] Recitals 6, 10 GDPR; Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[812] Recital 7 GDPR.
[813] Recital 39 GDPR.
[814] EDPB, 'Binding Decision 1/2021 on the dispute arisen on the draft decision of the Irish Supervisory Authority regarding WhatsApp Ireland under Article 65 (1) lit a GDPR' (2021) paras 191, 193.

4.4.1 Legal problems: Type 1

AI systems may be rather ubiquitous[815] and all AI disciplines[816] introduced in Chapter 2 may potentially clash with the transparency principle. Research has shown that users of smart homes are unaware of the possibility that machine learning (ML) algorithms may infer highly sensitive information, including sleep patterns and home occupancy.[817] Natural language processing (NLP) embedded in virtual assistants[818] such as Alexa or Siri facilitate the interception, recording and analysis of private communications without the users being aware of it, as unveiled by the press.[819] Computer vision (CV) applications allow identification of individuals from a distance and in a covert manner by means of face detection or gait analysis, without the knowledge of the individuals concerned. Regarding affective computing (AC) applications, transparent processing would presuppose that an individual is able to see what emotion the machine recognised, a requirement that also has been propagated by the pioneer in the field of AC.[820] However, in practice, this does not seem to be the case. The automated video assessment system provided by HireVue[821] aims to detect emotional states of applicants during job assessments. Similarly, the automated border control system called IBORDERCTRL 'analyses the micro-gestures of travellers to figure out if the interviewee is lying'.[822] Both systems do not communicate the detected emotions or detected 'lies' to the individuals concerned.

In addition, the dynamic nature of AI contradicts the static nature of the transparency principle because AI systems are continuously updated and changed whereas transparency disclosure only concerns algorithms used at a given moment.[823] All the examples mentioned illustrate that applications of AI potentially violate the transparency principle because personal data are not processed in a transparent manner as required by Article 5 (1) lit a GDPR. The following example regarding ML makes

[815] Finale Doshi-Velez et al, 'Accountability of AI Under the Law: The Role of Explanation' (2017) Berkman Klein Center Working Group on Explanation and the Law Working Paper 1 <https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf> accessed 8 February 2024; Jenna Burrel, 'How the machine 'thinks': understanding opacity in machine learning algorithms' (2016) Vol 3 Iss 1 Big Data Society 1-12 <https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512> accessed 8 February 2024.
[816] With the sole exception of AR, which is not problematic in this context.
[817] Zheng Serena, 'User Perceptions of Smart Home IoT Privacy' (2018) Vol. 2 Proceedings of the ACM on Human-Computer Interaction 3.
[818] See Section 4.2.6 below for an explanation of how virtual assistants work.
[819] See for example <https://www.forbes.com/sites/blakemorgan/2018/02/05/are-digital-assistants-always-listening/>, <https://www.theverge.com/2019/7/11/20690020/google-assistant-home-human-contractors-listening-recordings-vrt-nws>, < https://www.nytimes.com/2018/05/25/business/amazon-alexa-conversation-shared-echo.html> accessed 8 February 2024; see also Silvia de Conca, 'The enchanted house' Doctoral Thesis, Tilburg University 2021) <https://pure.uvt.nl/ws/portalfiles/portal/50798678/De_Conca_The_Enchanted_23_06_2021_emb_tot_23_06_2022.pdf > accessed 8 February 2024.
[820] Rosalind W Picard, *Affective Computing* (MIT Press 1997) 122.
[821] Nathan Mondragon, Clemens Aicholzer, Kiki Leutner, 'The Next Generation of Assessments' (HireVue 2019) <http://hrlens.org/wp-content/uploads/2019/11/The-Next-Generation-of-Assessments-HireVue-White-Paper.pdf> accessed 8 February 2024. HireVue halted the use of this component, but other providers offer similar services; see < https://hume.ai/products/facial-expression-model/ > and < https://gethume.com/blog5/artificial-intelligence-for-recruiting > accessed 8 February 2024.
[822] European Commission, 'Smart lie-detection system to tighten EU's busy borders' (24 October 2018) < https://ec.europa.eu/research-and-innovation/en/projects/success-stories/all/smart-lie-detection-system-tighten-eus-busy-borders > accessed 8 February 2024.
[823] Council of Europe, Committee of Convention 108, 'Guidelines on Artificial Intelligence and Data Protection' (25 January 2021) at 3 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.

this more concrete. Litera c of Article 13 (1) and 14 (1) GDPR impose the obligation on controllers to inform data subjects about the purposes of the processing for which the personal data are intended. Where personal data are directly collected from data subjects, this information must be provided at the time when personal data are obtained, and in all other cases within one month, at the latest, after obtaining the personal data or at the time of the first communication with the data subjects. Unsupervised ML approaches process data for *un*specified and *in*explicit purposes – the processing *itself* determines the purpose of the future use of the data – since its goal is to detect patterns and correlations, gain knowledge and make accurate predictions. There is no transparency issue if the controller processes personal data within the AI system for training purposes. However, this is different when the controller intends to detect correlations, patterns, and commercially valuable insights in data by deploying unsupervised ML. In such a case, the controller will determine the *specific* purpose of processing based on the processing activity's results, i.e., after the processing. Consequently, the transparency principle as further substantiated in Articles 13 (1) and 14 (1) GDPR cannot be complied with because the purpose of processing is not known at the time of data collection or when obtained from sources other than the data subject. Indeed, regulatory guidance demands that one 'always specify the purposes of the processing at the time of collection.'[824]

Take, for example, inferred personal data defined as 'the product of probability-based processes' that are used to create predictions of behaviour deployed to categorise individuals.[825] Where the purpose of processing consists of the creation of inferred personal data as is the case with ML, regulatory guidance requires one to communicate, at the time of collection or prior to further processing, 'the *intended purpose* of creating and further processing such inferred personal data, as well as the *categories* of the inferred data processed'.[826]

ML aims to create inferred personal data by detecting patterns and correlations, gaining knowledge and making accurate predictions. Therefore, controllers will not be able to inform data subjects about the specific purposes for which personal data are further processed because this information is completely unknown at the time of data collection or prior to further processing. Controllers could certainly inform data subjects about the intended purpose of processing in rather general terms such as 'We may use your personal data for detecting patterns/correlations and make accurate predictions about you.' However, such information will ultimately not meet the level of transparency required for the purpose specification, as foreseen by regulatory guidance, which states that the phrase 'We may use your personal data to develop new services' is not sufficiently clear about the purpose of

[824] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 14.
[825] OECD Working Party on Security and Privacy in the Digital Economy JT03357584 (2014) <https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg(2014)3&doclanguage=en> accessed 8 February 2024.
[826] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 14 emphasis added by the author.

processing.[827] Likewise, controllers are unable to inform data subjects about the categories of inferred data processed as required by regulatory guidance. The categories of inferred personal data are unknown prior to further processing. First, an AI system needs to generate the inferred personal data before the controller can inform data subjects about the categories thereof.

The opacity problem (Type 1)

Unsupervised ML approaches process data for inexplicit purposes – the processing itself determines the purpose of the future use of the data. Controllers cannot inform data subjects about the purpose of processing nor the categories of inferred personal data because this information is not known at the time of data collection or prior to further processing. This violates the transparency principle.

Transparency regarding automated decision-making (ADM) constitutes a particular issue when applied to AI. Articles 13 (2) lit f and 14 (2) lit g GDPR require controllers to provide 'meaningful information about the logic' involved in ADM. Wachter, Mittelstadt and Floridi take the view that meaningful information according to Articles 13 (2) lit f and 14 (2) lit g GDPR can logically only address system functionality, namely, information about the logic, significance, envisaged consequences and general functionality of an ADM system, but not the rationale of *specific* ADM as the latter cannot be known before the decision is made.[828] Their reasoning suggests that information according to Articles 13 (2) lit f and 14 (2) lit g GDPR can only be provided *ex-ante* because notification occurs before ADM takes place, namely, at the point when personal data are collected for processing.[829] Malgieri and Comandé argue that meaningful information about the logic involved must adhere to the standard of legibility, which requires that the information to be provided is both transparent and comprehensible, and that such information must go 'beyond the mere mathematical functionality of an algorithm' and consider contextual use, expected and actual impact, rationales and purposes.[830] More generally, there is a vivid debate in scholarship whether or not the GDPR provides a right to explanation of specific ADM.[831]

Irrespective of this debate which will be discussed in the context of the right of access (Section 5.6.2), the notion of 'meaningful information' remains elusive. It is not yet clear what 'meaningful information' precisely means in practice. This notion also appears in Article 15 (1) lit h GDPR. Custers

[827] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 12.
[828] Sandra Wachter, Brent Mittelstadt, Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) Vol 7 Iss 2 IDPL 76, 78.
[829]Ibid 76, 82.
[830] Giancludio Malgieri, Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Makig Exists in the General Data Protection Regulation' (2017) Vol 7 Iss 4 IDPL 243, 245, 257, 258.
[831] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 75-101; Sandra Wachter, Brent Mittelstadt, Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) Vol 7 Iss 2 IDPL 76-99; Giancludio Malgieri, Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Makig Exists in the General Data Protection Regulation' (2017) Vol 7 Iss 4 IDPL 243-265.

and Heijne performed research related to this notion enshrined in the right of access. They suggest interpreting it as information that is useful and/or has practical value for data subjects.[832] This interpretation has a contextual component and arguably means useful and practical for data subjects to (i) become aware of processing relating to ADM, (ii) enforce their data subject rights and thus (iii) exercise control over the processing of their personal data. For this section, I interpret meaningful information as useful and/or having practical value for data subjects. This is also in line with the CJEU's focus on intelligibility with respect to Article 12 (1) GDPR, which ensures that the data subject fully understands the information sent to it.[833]

It seems clear that controllers must understand the functionality of an ADM system to be able to provide data subjects with information that is useful and/or of practical value (meaningful information). Such information can only be provided if the trained model used for the ADM system can be articulated and understood by a human.[834] Giving information about the type of input data and the expected output, explaining the variables and their weight, or shining light on the analytics architecture are various forms of transparency concerning the logic of AI algorithms.[835] However, providing such information constitutes a two-sided problem: some information might effortlessly be provided by humans, but not by AI systems and vice versa.[836] The reasons for this are as follows.

First, AI lacks common sense reasoning capabilities due to deficiencies in automated reasoning as outlined in Sections 2.2.5 and 4.2.1. Systems based on ML do not *know* why specific input should receive some label, they solely know that certain input correlate with such a label. For example, an ML model trained with a dataset in which all basketballs are orange might classify all future input that is orange as basketballs.[837] For humans, it would be common sense not to do so. Due to these reasoning deficiencies, it seems reasonable to argue that AI itself is currently not capable of displaying the logic involved in ADM systems and the rationale behind or the criteria relied on to make the automated decision. Consequently, controllers cannot provide data subjects with information that is useful or of practical value for them.

[832] Bart Custers, Anne-Sophie Heijne, 'The right of access in automated decision-making: The scope of article 15(1)(h) GDPR in theory and practice' (2022) Vol 46 Computer Law & Security Review 1, 14.
[833] Case C-487/21, *F.F.* [2022] ECR I-1000 paras 37-38; in addition, Opinion of AG Pitruzella paras 55-56.
[834] Bryce Goodman, Seth Flaxman, 'European Union regulations on algorithmic decision-making and a right to explanation' (2017) Vol 38 No 3 AI Magazine 50, 55.
[835] Council of Europe, Committee of Convention 108, 'Guidelines on Artificial Intelligence and Data Protection' (25 January 2021) at 31 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.
[836] Finale Doshi-Velez et al, 'Accountability of AI Under the Law: The Role of Explanation' (2017) Berkman Klein Center Working Group on Explanation and the Law Working Paper 1 <https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf> accessed 8 February 2024.
[837] Zachary C Lipton, 'The Mythos of Model Interpretability' (2018) Vol 16 Iss 3 ACMQueue 3 <https://dl.acm.org/doi/pdf/10.1145/3236386.3241340?download=true> accessed 8 February 2024.

Second, the complexity of many AI systems makes it impossible to present the casual factors which have led to a decision in a manner which is understandable for data subjects.[838] In particular, the complexity of adopted ML models represents a major challenge for human cognition.[839] With some algorithms used in AI systems, it is practically impossible to retroactively connect specific input to specific output and vice versa.[840] The difficulty to establish a nexus between specific input and output and thus to derive the logic involved in ADM differs considerably between the techniques used for ML. ML algorithms deploying sparse linear models such as regression introduced in Section 2.2.1.1 tend to generate interpretable models, allowing to identify the role of each model component (e.g., weight of a feature in a linear regression model) within the whole computing process, which ultimately leads to traceability and transparency in ADM.[841] However, this is different in case of deep learning (DL) and artificial neural networks (ANN). When an ANN is used for pattern recognition in CV or NLP, an ex-post analysis of a specific ADM will likely not establish a linear causal connection which is easily comprehensible for human minds.[842] Complex processes applied in deep learning (DL) are challenging for human cognition, both in terms of explaining the logic of the algorithms and the specific ADM. Non-deterministic systems make it hard to provide detailed information about the logic involved in the processing of personal.[843] With regard to explainability seen as the identification of factors that have caused a decision,[844] ANN and DL pose perhaps the biggest challenge.[845] Most current DL models lack reasoning and explanatory capabilities, which makes them vulnerable to produce unexplainable outcomes. DL methods based on ANN generally lack interpretability.[846] It seems neither possible to understand which artificial neuron contributed to a distinct part of the output nor to understand what happened in the intermediate (hidden) layers of the ANN.[847] Therefore, humans will hardly be able to extract any underlying rules which may be used to determine the logic involved in ADM: the many numeric values of the weights produced by the model do not have a meaning to the supervisor.[848] Consequently, controllers cannot provide data subjects with meaningful information, namely, information that is useful and/or has practical value. However, the field of

[838] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 89.
[839] Zachary C Lipton, 'The Mythos of Model Interpretability' (2018) Vol 16 Iss 3 ACMQueue 18 <https://dl.acm.org/doi/pdf/10.1145/3236386.3241340?download=true> accessed 8 February 2024.
[840] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 81.
[841] Apostolos Vorras, Lilian Mitrou, 'Unboxing the Black Box of Artificial Intelligence: Algorithmic Transparency and/or a Right to Functional Explainability' in Titania-Eleni Synodinou et al (eds) *EU Inernet Law in the Digital Single Market* (Springer Nature 2021) 256.
[842] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 81.
[843] Council of Europe, Committee of Convention 108, 'Guidelines on Artificial Intelligence and Data Protection' (25 January 2021) at 3 <https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3> accessed 8 February 2024.
[844] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 88.
[845] Bryce Goodman, Seth Flaxman, 'European Union regulations on algorithmic decision-making and a right to explanation' (2017) Vol 38 No 3 AI Magazine 50, 55.
[846] Deng Li and Liu Yang, 'A Joint Introduction to Natural Language Processing and Deep Learning' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 11, 12.
[847] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 155.
[848] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence* (2nd edn, Springer 2008) 12, 25, 35, 44.

Explainable AI ('xAI') has made significant progress in recent years. xAI aims to develop explainable techniques that empower end users to comprehend, trust, and efficiently manage AI systems.[849] Nonetheless, causal explanations, which are crucial for ADM, are still a challenge and are anticipated to be the next frontier of ML.[850]

Regulatory guidance acknowledges the challenge for humans to understand how ADM processes work in the context of ML. Nevertheless, the guidance also states that complexity is no excuse and controllers should find 'simple ways to tell the data subject about the rationale behind, or the criteria relied on reaching the decision'.[851] However, considering current deficiencies in terms of interpretability in the context of ML and ANNs and deficiencies in automated reasoning, it seems that the ideal of transparency with respect to meaningful information about the logic involved in ADM is technologically not possible (yet). Due to interpretability and reasoning deficiencies, controllers are unable to provide data subjects with meaningful information, namely, information that is useful and/or has practical value. This leads to a Type 1 legal problem, because Articles 13 (2) lit f and 14 (2) lit g GDPR are violated.

The interpretability problem (Type 1)
Due to the deficiencies in AR, AI systems cannot themselves display the logic involved in ADM systems and the rationale behind or the criteria relied on reaching the automated decision. AI systems deploying DL and ANN approaches from ML are likely to produce non-interpretable outputs. When used in the context of ADM, controllers cannot provide data subjects with meaningful information about the logic involved in ADM and thus violate the transparency principle.

4.4.2 Legal problems: Type 2

The interpretability problem (Type 2)
The interpretability problem outlined in Section 4.4.1 also leads to a Type 2 legal problem. Due to the deficiencies in terms of interpretability in the context of DL and ANN as well as deficiencies in AR, it is technologically not possible for controllers to induce meaningful information about the logic involved in ADM. Therefore, data subjects and regulators cannot enforce the transparency principle and obtain the corresponding information.

[849] Waddah Saeed, Christian Omlin, 'Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities' (2023) Vol 263 Knowledge-Based Systems 1-22.
[850] Ibid 9.
[851] Art 29 Working Party 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679', (WP251rev.01, 6 February 2018) at 25.

Making a methodological remark here regarding type 1 and 2 problems seems appropriate. The interpretability problem (type 1) should not automatically lead to a type 2 problem because the controller is required to cease processing after becoming aware that processing is unlawful.

Under the GDPR, a processing activity whose complexity makes it impossible for the controller to respect data protection principles should not occur.[852] This follows from the accountability principle[853] and other obligations, such as performing a Data Protection Impact Assessment (DPIA). When deploying ADM systems described in Section 4.4.1, the controller must perform a DPIA according to Article 35 GDPR. DPIAs are required if the envisaged processing is likely to result in a high risk to the rights and freedoms of data subjects. The performance of a DPIA is mandatory when the controller uses 'new technologies'[854] for processing, including AI systems. In such cases, controllers should consult the competent Supervisory Authority (SA) if the high risks cannot be mitigated.[855] If compliance with the GDPR principles is impossible, the controller should stop the processing. In addition, the competent SA may also ban such processing based on Article 58 (2) GDPR. Hence, a type 1 problem should not lead to a type 2 problem, as the processing should simply not occur. However, the possibility remains that controllers perform a cost-risk analysis and continue with such processing even after warnings or fines from SAs. The latter is not only a theoretical possibility. For instance, OpenAI continued to provide its services after bans and warnings imposed by the Italian SA.[856] Also, Clearview AI continued with its processing activities even after receiving clear signs from the EDPB concerning the lawfulness of the processing.[857] Thus, although non-compliance with data protection principles should result in ceased processing activities, this might be ignored in practice (e.g., by powerful tech companies). Ultimately, the principles are being violated and cannot be enforced simultaneously, leading to a type 2 problem.

4.4.3 Legal problems: Type 3

The GDPR contains provisions requiring controllers to inform data subjects if their personal data will be processed for a different purpose which is compatible with the one for which personal data were initially collected. Articles 13 (3) and 14 (4) GDPR specifically relate to the purpose limitation principle enshrined in Article 5 (1) lit b GDPR[858] which states that further processing for scientific research purposes or statistical purposes shall not be incompatible with the initial purpose (i.e. privileged purposes). As will be outlined in Section 4.5.3, there are reasons to argue that ML serves

[852] This also applies to the verification problem discussed in Section 4.6.2.
[853] Article 5 (2) GDPR.
[854] Article 35 (1) GDPR.
[855] Article 36 GDPR.
[856] See < https://iapp.org/news/a/garante-issues-notice-to-openai-over-alleged-gdpr-violations/ > accessed 8 February 2024.
[857] See <https://edpb.europa.eu/sites/default/files/files/file1/edpb_letter_out_2020-0052_facialrecognition.pdf> accessed 8 February 2024.
[858] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 23.

research or statistical purposes, which might enable controllers to rely on these privileged purposes to generate inferred personal data. Inferred data are 'the product of probability-based processes' and are used, for instance, to create predictions of behaviour.[859]

Controllers may also further process personal data for compatible purposes other than privileged ones based on an assessment that takes into account the factors mentioned in Article 6 (4) GDPR.[860] Article 6 (4) GDPR stipulates a series of criteria to determine whether further processing for a purpose other than the one for which personal data have been initially collected is 'compatible' with this initial purpose.[861] According to the CJEU, these criteria reflect the need for a concrete, coherent and sufficiently close link between the purpose of collection and further processing of data. These criteria make it possible to determine that further processing does not detract from the data subject's legitimate expectations as to the further use of their personal data.[862] Where controllers can establish such a link, they may further process personal data in order to detect an individual's emotional state by means of AC. Provided that the purposes for further processing are compatible, either privileged[863] or otherwise compatible,[864] controllers solely need to notify data subjects in advance about these compatible purposes and with any relevant further information as referred to in paragraph *2* of Article 13 and 14 GDPR. Both provisions *do not* include information about the nature of inferred personal data or categories of personal data.

Controllers do not need to outline which personal data or categories of personal data are processed if the initial personal data are directly collected from data subjects.[865] Where the initial personal data are not directly collected from the data subject, information about the categories of personal data as received by the controller must be provided according to Article 14 (1) lit e GDPR. However, because this requirement is enshrined in paragraph 1 and *not* 2 of Article 14 to which Article 14 (4) GDPR refers, controllers do not need to inform data subjects about the actual category of the personal data inferred by means of ML or AC. In other words, controllers must indicate the categories of personal data they have received from another controller, but not the ones inferred from such data. With regard to inferred personal data, regulatory guidance on transparency requires that 'the *intended purpose* of creating and further processing such inferred personal data, as well as the *categories of the inferred*

[859] OECD Working Party on Security and Privacy in the Digital Economy JT03357584 (2014) <https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg(2014)3&doclanguage=en> accessed 8 February 2024.
[860] Ibid, e.g. the link between the initial and envisaged purposes, the context of collection, the nature of the personal data (e.g., special categories) and the possible consequences for data subjects.
[861] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 315, 316.
[862] Case C-77/21 *Digi* [2022] ECR I-805 para 36; Case C-77/21 *Digi* [2022] ECR I-805 Opinion of AG Pikamäe paras 28, 59, 60.
[863] Research or statistical purposes in the case of ML.
[864] Article 6 (4) GDPR
[865] See Article 13 (1) GDPR and regulatory guidance which confirms that controllers must not provide individuals about the categories of personal data processed. See Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 36.

data processed, must always be communicated to the data subject *at the time of collection* or *prior to the further processing* for a new purpose in compliance with Article 13.3 or Article 14.4'.[866] Regulatory guidance derives this requirement *not* from the transparency principle and the related obligations contained in the GDPR, but from the fairness and purpose limitation principles.[867] By relying on the purpose limitation and fairness principle, regulatory guidance confirms, at least implicitly, the interpretation that controllers are not obliged to inform data subjects about the actual category of the inferred personal data or the detected emotional state (AC) based on Articles 13 and 14 GDPR. This is contradictory to the objectives of the transparency principle, which aim to enable data subjects to (i) become aware of processing[868] and (ii) enforce their rights.[869] It also prevents data subjects from exercising control over the processing of their personal data[870] (see also the profiling problem).

The interpretation that controllers are not obliged to inform data subjects about the actual category of the inferred personal data or the detected emotional state based on Articles 13 and 14 GDPR leads to opacity rather than transparency. Data subjects will not be informed about the inferred personal data generated by ML because there is no specific legal obligation for controllers to do so.[871] Imagine, for example, an insurance company which deploys unsupervised ML techniques to detect patterns and correlations in rather simple personal data such as sex and place of residence of their clients. The AI system detects correlations between sex and place of residence, in particular that women living in certain areas tend to live longer. Based on this correlation, the AI system automatically predicts the life expectancy of these clients and stores this information within the insurance customer relationship management system. Life expectancy constitutes inferred personal data generated by means of unsupervised ML techniques and is based on personal data directly collected from the data subjects. Therefore, the insurance company is not required under the transparency obligations enshrined in the GDPR to inform the data subjects concerning the knowledge gained, namely, the detected correlation and the predicted life expectancy. Controllers are not obliged to inform data subjects about the categories of such inferred personal data. The insurance company must inform the data subject only about the purpose of further processing and any other information mentioned in Article 13 (2) GDPR, but not about the actual personal data generated by the AI systems or at least the categories thereof. Article 13 GDPR, which is applicable in this case,[872] does not contain such a requirement, contrary to Article 14 (1) lit d GDPR.

[866] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) in Footnote 30 at page 14, emphasis added by the author.
[867] Ibid at page 14.
[868] Recital 39 GDPR.
[869] Articles 15-22 GDPR as well as remedies contained in Articles 77-80 GDPR.
[870] Recital 7 GDPR.
[871] Provided that the privileged 'statistical purpose' or 'research purpose' apply to processing by means of ML.
[872] Because the personal data used by the system as input data was collected from the data subject.

Arguably, inferred data constitute 'new' personal data not collected from the data subject, triggering the transparency obligations contained in Article 14 GDPR. However, this view is not convincing for several reasons. First, Article 14 GDPR clearly covers situations where personal data was collected from third-party sources.[873] The latter is emphasised by the wording contained in Article 14 (2) lit f which requires controllers to disclose 'from which source the personal data originate'. Recital 61 GDPR refers to the situation where the personal data do not originate from the data subject but are 'obtained from another source.' In the example at hand, the inferred personal data originate from the data subjects but not from another source (e.g., a third-party controller). Second, this also makes sense when applying a systematic interpretation. Generating inferred personal data constitutes 'further processing' mentioned in Articles 13 (3) and 14 (4) GDPR. Article 13 (3) GDPR would be obsolete if Article 14 (4) GDPRP would govern the insurance company's further processing. Third, data subjects may enforce their right of access to obtain information about the personal data generated by the AI system. The CJEU has clarified that the scope of a copy under Article 15 (3) GDPR includes personal data *generated* by the *controller*[874] and thus inferred personal data.

The outcome will be the same when personal data are inferred by means of affective computing. Controllers are not required to inform the data subject about the specific detected emotional states or about the category of inferred personal data. Regulatory guidance that suggests otherwise, namely, that controllers need to inform data subjects about the *categories of the inferred data* processed, based on the purpose limitation and fairness principle,[875] may be easily refuted. First, Articles 13 and 14 GDPR implement the transparency principle and impose specific information duties on the controller. These provisions do not include an obligation to inform about the categories of inferred personal data, as suggested by regulatory guidance. Second, controllers already comply with the principles of transparency and purpose limitation by informing data subjects about the purpose for further processing, provided that the latter is compatible with the initial purpose. Third, as outlined in Section 4.3.2, the substantive meaning of the fairness principle is elusive.[876] This makes it easy to challenge the interpretation that controllers must inform data subjects about the categories of inferred personal data, in particular because the transparency obligations enshrined in the GDPR do not entail such a specific obligation. The conclusion that the GDPR does not require controllers to inform data subjects about inferred personal data constitutes a Type 3 legal problem. Articles 13 and 14 GDPR are not fit for purpose to protect the fundamental right to data protection. These provisions fail to achieve the objectives of the transparency principle, namely, enabling data subjects to (i) become aware of

[873] Gabriela Zanfir-Fortuna, Commentary of Article 14 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 436, 445, 446.
[874] Case C-487/21, *F.F.* [2022] ECR I-1000 para 21; see also the opinion of AG Pitruzella paras 45, 70
[875] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) in footnote 30 at page 14 emphasis added by the author.
[876] Damian Clifford, Jef Ausloos 'Data Protection and the Role of Fairness' (2018) Vol 37 No 1 Yearbook of European Law 130, 187.

processing[877] (ii) enforce their rights[878] and (iii) exercise control over the processing of their personal data[879]. Consequently, these provisions fail to effectively protect data subjects.[880] Data subjects will not be aware of the actual personal data inferred by means of ML or AC, such as the predicted life expectancy or the emotional state detected by the AI system. Therefore, data subjects cannot exercise their rights as a data subject because they are simply not aware of the inferred personal data.

The inference problem (Type 3)

ML and AC enable controllers to infer personal data such as predictions or emotion data based on personal data provided by data subjects or obtained otherwise. Transparency obligations contained in the GDPR do not require controllers to inform data subjects about personal data inferred by means of AI if the data are processed for compatible purposes. Consequently, data subjects do not become aware of such data and cannot exercise their rights. Therefore, Articles 13 and 14 GDPR are not fit for purpose to protect the fundamental right to data protection.

Controllers may predict preferences, behaviour and attitudes of data subjects using ML techniques such as regression, classification (see Section 2.2.1.1) or clustering (Section 2.2.1.2), amounting to profiling as defined in the GDPR. AC empowers controllers to predict an individual's personal state and thus to evaluate particular personal aspects related to that individual. Article 4 (4) GDPR defines profiling as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person'. In academia, profiling is referred to as 'the process of (i) inferring a set of characteristics about an individual or group of persons (i.e. the process of creating a profile) and/or (ii) treating that person or group (or other persons/groups) in light of these characteristics (i.e. the process of applying a profile)'.[881] AI, particularly ML and AC, can be used for both steps contained in the process of profiling, namely, first to infer a profile by means of unsupervised or supervised ML or predict an individual's emotional state (AC) and subsequently treat the individual accordingly. Controllers may rely on dark patterns to collect personal data required for profiling purposes.[882] Dark patterns are design practices which undermine a user's autonomy by coercing, misleading or manipulating their decision-making and behaviour.[883]

[877] Recital 39 GDPR.
[878] Articles 15-22 GDPR as well as remedies contained in Articles 77-80 GDPR.
[879] Recital 7 GDPR.
[880] Recital 11 GDPR; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73; Joined cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; joined cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[881] Isak Mendoza, Lee A Bygrave, 'The Right Not to be Subject to Automated Decisions Based on Profiling' in Tatiana-Eleni Synodinou et al (eds) *EU Internet Law: Regulation and Enforcement* (Springer International 2017) 77.
[882] Arunesh Mathur, Jonathan Mayer, Mihir Kshirsagar, 'What Makes a Dark Pattern... Dark?' (CHI Conference on Human Factors in Computing Systems, Yokohama, May 2021) 16; Tim Kollmer, Andreas Eckhardt, 'Dark Patterns' (2022) Vol 64 Iss 6 Business & Information Systems Engineering 1.
[883] Sanju Ahuja, Jyoti Kumar, 'Conceptualizations of user autonomy within the normative evaluation of dark patterns' (2022) Vol 24 Iss 4 Ethics and Information Technology 1.

Profiling can have a highly predictive nature[884] and may generate stereotypes by assuming that certain behaviour of an individual, such as receiving good grades at a renowned university, is an indicator for a corresponding outcome, for example, securing a well-paid job.[885] Such predictive profiling may be used to predict an individual's behaviour, character, risk (e.g. score values) and to treat the individual accordingly.[886] For example, an individual's phone-charging habit is currently used as a relevant factor for determining this individual's creditworthiness. AI systems powered by ML in particular assess data points such as phone-charging habits that would commonly not be considered when determining someone's creditworthiness. Smart Finance disclosed that customers who regularly let their phone batteries drop below 12% are not considered good prospects. Another FinTech company called Lenddo states the opposite and considers hyper well-maintained smartphone batteries as a red flag because such a phone-charging habit seems to be robotic or not human enough.[887] In fact, research suggests that behaviour revealed in mobile phone usage accurately predicts the likelihood of credit repayment. By means of ML, the likelihood of repayment was predicted using behavioural features derived from mobile phone usage.[888]

The predictive nature of profiling is also emphasised by Recital 24 GDPR, which states that profiling may be used for analysing or predicting the personal preferences, behaviour and attitudes of data subjects. ML as introduced in Section 2.2.1 is the favoured way of deriving profiles[889] particularly because profiles are patterns resulting from probabilistic processing of data.[890] Apart from obvious examples such as discrimination, risks of profiling relate to the one-sided supply of information (information asymmetry) and the negative influence on the data subject's personal autonomy.[891] Profiling exacerbates the power inequality and information asymmetry between those that profile (controllers) and those that are being profiled (the data subjects).[892] It also threatens personal autonomy by surreptitiously influencing, formatting and customising individual behaviour.[893] The essence of

[884] Helena U Vrabec, 'Uncontrollable: Data Subject Rights and the Data-driven Economy' (Dissertation, Leiden University 2019) 220.
[885] Frederick F Schauer, Profiles, Probabilities, and Stereotypes (Harvard University Press 2006) 6.
[886] Helena U Vrabec, 'Uncontrollable: Data Subject Rights and the Data-driven Economy' (Dissertation, Leiden University 2019) 220; Hans Lammerant, Paul de Hert, 'Predictive profiling and its legal limits: Effectiveness gone forever' In Bart van der Sloot et al (eds) *Exploring the boundaries of big data* (2016 Amsterdam University Press/WRR) 145-173.
[887] Tanya Goodin, 'The battery life of your phone could affect your loan application' (2022) <https://tanya-goodin.com/2022/08/credit-rating-algorithmic-transparency/> accessed 8 February 2024.
[888] Daniel Björkegren, Darrell Grissen, 'Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment' (2020) Vol 34 Iss 3 The World Bank Economic Review 618, 623.
[889] Lilian Edwards, Michael Veale, 'Slave to the Algorithm: Why a 'Right to Explanation' is Probably not the Remedy You are Looking for' (2017) Vol 16 Iss 1 Duke Law & Technology Review 19, 46.
[890] Helena U Vrabec, 'Uncontrollable: Data Subject Rights and the Data-driven Economy' (Dissertation, Leiden University 2019) 220.
[891] Bart Custers, 'Data Dilemmas in the Information Society' in Bart Custers et al (eds), *Discrimination and Privacy in the Information Society* (Springer 2013) 1.
[892] Mireille Hildebrandt, Bert-Jaap Koops, 'The challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) Vol. 73 (3) The Modern Law Review 428, 435; Serge Gutwirth, Mireille Hildebrandt, 'Some Caveats on Profiling' in Serge Gutwirth et al (eds), *Data Protection in a Profiled World* (Springer Nature 2010) 34.
[893] Serge Gutwirth, Mireille Hildebrandt, 'Some Caveats on Profiling' in Serge Gutwirth et al (eds), *Data Protection in a Profiled World* (Springer Nature 2010) 34; Serge Gutwirth, *Privacy and the information age* (Lanham:Rowman & Littlefield Publishers 2002).

autonomy is indicated by the etymology of the term: autos (self) and nomos (rule or law).[894] Put simply, autonomy refers to a person's ability to make rational and uncoerced choices and decisions[895] or, in other words, to 'make their own lives'[896] and face freely both existential and every day's choices.[897] As noted by AG Pikamäe, profiling may reinforce existing stereotypes, increase the social divide, restrict the data subject's freedom of choice regarding certain products or services and result in the denial of services.[898] Profiling deprives data subjects not only of the means to reflect on the choices the environment makes for them, but may proactively impact the choices they make. This is called 'the autonomy trap'.[899] I now outline why the GDPR fails to address the information asymmetry concerning profiling and the subsequent negative impact on the data subject's personal autonomy.

Profiling defined in Article 4 (4) GDPR refers to any form of automated processing ('regular profiling') and also covers profiling with subsequent human involvement, as opposed to profiling used for ADM ('ADM profiling') which must be fully automated and satisfy the two other cumulative requirements of Article 22 (1) GDPR.[900] According to regulatory guidance, ADM has a different scope than regular profiling but may partially overlap with or result from profiling (see also Section 3.3.4.6). Decisions which are not solely automated according to Article 22 GDPR might also include profiling.[901] Regulatory guidance dealing with the transparency principle stresses the importance of informing data subjects about the consequences of processing, also with regard to regular profiling and not only ADM profiling which is captured by Article 22 GDPR.[902] This information duty is derived from Recital 60 GDPR stating that data subjects 'should be informed of the existence of profiling and the consequences of such profiling'. Interestingly, regulatory guidance with respect to ADM adopted *prior* to the transparency guidelines stresses that if ADM and profiling '*does not* meet the Article 22 (1) definition it is *nevertheless good practice* to provide' the information according to Article 13 (2) lit f and 14 (2) lit g.[903] These two provisions oblige controllers to inform data subjects about 'the existence of automated decision-making, *including profiling*, referred to in Article 22(1) and (4) and, *at least in those cases*, meaningful information about the logic involved, as well as the *significance and the envisaged consequences* of such processing for the data subject'.[904] Because these provisions contain the wording '*at least in those cases*', controllers are *not* legally required to inform data

[894] Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge University Press 1988) 12, 18.
[895] Maurits Clemens Kapitein, 'Personalized Persuasion in Ambient Intelligence' (Doctoral Thesis, TU/e Eindhoven 2012) 179 < https://pure.tue.nl/ws/files/3470131/729200.pdf> accessed 8 February 2024.
[896] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1986) 369.
[897] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 8.
[898] Case C-634/21, *SCHUFA Holding AG* [2023] ECR I-220, Opinion AG Pikamäe Footnote 6 in para 19.
[899] Tal Z. Zarsky, 'Mine your own business!' (2003) 5 Yale Journal of Law and Technology 35.
[900] ADM profiling must involve a decision and has to produce legal or similarly significant effects see Case C-634/21, *SCHUFA Holding AG* [2023] ECR I-957 para 43.
[901] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 251rev.01, 6 February 2018) at 7 and 8.
[902] Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260 rev.01, 11 April 2018) at 41.
[903] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 251rev.01, 6 February 2018) at 25.
[904] Emphasis added by the author.

subjects about the significance and envisaged consequences of 'regular' profiling as this obligation solely applies to ADM profiling meeting the three cumulative conditions of Article 22 GDPR.[905] Even regulatory guidance confirms this reading: It is 'good practice' to provide this information also regarding regular profiling.[906]

Furthermore, the preparatory documents of the GDPR confirm this interpretation. During the GDPR negotiation process, Poland suggested to use the wording 'where applicable, information about the existence of profiling referred to in Article 4 (12a) *and/or* about automated decision-making'[907] instead the final wording of Article 13 (2) lit f and 14 (2) lit g GDPR. Therefore, it is likely that if the intent of the legislator was to oblige controllers to provide data subjects with information on the importance and implications envisaged of regular profiling, the final language of Articles 13 (2) lit f and 14 (2) lit g GDPR would contain a specific reference to the definition of profiling. The objection that controllers are in fact obliged to disclose such information regarding regular profiling based on Recital 60 GDPR is not very strong. Recitals may cast light on the interpretation to be given to a rule, but cannot in itself constitute such a rule.[908] In addition to that, recitals are legally not binding.[909] The results of regular profiling might constitute 'new' inferred personal data. However, as discussed in the inference problem, Article 14 GDPR does not apply to inferred personal data originating from data provided by the data subject. Instead, Article 14 GDPR applies where personal data have been obtained from a source other than the data subject (third party).

Thus, the transparency principle as implemented in Articles 13 (2) lit f and 14 (2) lit g GDPR does not require controllers to inform data subjects about the significance and consequences of regular profiling as defined in Article 4 (4) GDPR. These provisions fail to achieve the objectives of the transparency principle, namely, enabling data subjects to (i) become aware of processing[910] and (ii) enforce their rights.[911] The fact that data subjects will not be informed about the significance and consequences of regular profiling also sharpens the power inequality and information asymmetry

[905] The CJEU confirmed that three cumulative conditions must be met to render Article 22 GDPR applicable see Case C-634/21, *SCHUFA Holding AG* [2023] ECR I-957 para 43.
[906] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 251rev.01, 6 February 2018) at 25.
[907] Emphasis added by the author. Also, note that Article 4 (12a) refers to the definition of profiling as finally enshrined in Article 4 (4) GDPR. Council of the European Union, General Data Protection Regulation Interinstitutional File: 2012/0011 (COD) (2015) at 117 < https://data.consilium.europa.eu/doc/document/ST-9281-2015-INIT/en/pdf > accessed 8 February 2024.
[908] Case T-709/21, WhatsApp Ireland Ltd [2022] ECR I-783 para 71.
[909] Case C-162/97, *Nilsson* [1998] ECR I-7477, para. 54.
[910] Recital 39 GDPR.
[911] Articles 15-22 GDPR as well as remedies contained in Articles 77-80 GDPR.

between controllers and data subjects instead of mitigating them.[912] The provisions also neglect the corresponding adverse effect on the data subject's personal autonomy.[913]

Profiling can be used to predict an individual's behaviour, character, risk (e.g., score values), evaluate an individual's personal aspects (e.g., emotional states) and to treat the individual accordingly.[914] The latter may involve (i) limiting the choices available to an individual[915] or (ii) proactively pushing the individual to make a certain decision. In terms of (i), AG Pikamäe notes that profiling may restrict the data subject's freedom of choice regarding certain products or services and result in the denial of services.[916] For example, a negative score value based on profiling limits the choices available for individuals to obtain a loan or even mobile subscriptions.[917] The limited choice undermines the individual's autonomy to 'make their own lives'[918] and face freely both existential and every day's choices.[919] In terms of (ii), AI-powered profiling enables controllers to push a person towards choices it may have resisted if being aware of what is known about him or her.[920] AI entails the characteristics of a persuasive technology, which is an 'interactive computing system designed to change people's attitudes and behaviours'.[921] This holds particularly true where companies use AI to influence consumers by tailoring their products and services to their needs, interests, personality or other factors relevant for them.[922] Companies analyse any kind of customer behaviour for profiling purposes and the gained knowledge is then used to proactively change the behaviour and decisions of these customers, which is called 'actuation'.[923] Persuasion is seen as an 'attempt to change attitudes or behaviour or both' without making use of practices such as coercion or deception.[924] Behaviour also includes decisions taken by individuals.

[912] Mireille Hildebrandt, Bert-Jaap Koops, 'The challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) Vol. 73 (3) The Modern Law Review 428, 435; Serge Gutwirth, Mireille Hildebrandt, 'Some Caveats on Profiling' in Serge Gutwirth et al (eds), *Data Protection in a Profiled World* (Springer Nature 2010) 34.
[913] Bart Custers, 'Data Dilemmas in the Information Society' in Bart Custers et al (eds), *Discrimination and Privacy in the Information Society* (Springer 2013) 1; Tal Z. Zarsky, 'Mine your own business!' (2003) 5 Yale Journal of Law and Technology 35.
[914] Helena U Vrabec, 'Uncontrollable: Data Subject Rights and the Data-driven Economy' (Dissertation, Leiden University 2019) 220; Hans Lammerant, Paul de Hert, 'Predictive profiling and its legal limits: Effectiveness gone forever' In Bart van der Sloot et al (eds) *Exploring the boundaries of big data* (2016 Amsterdam University Press/WRR) 145-173.
[915] Tal Z. Zarsky, 'Mine your own business!' (2003) 5 Yale Journal of Law and Technology 35
[916] Case C-634/21, *SCHUFA Holding AG* [2023] ECR I-220, Opinion AG Pikamäe Footnote 6 in para 19.
[917] Case C-203/22 *Dun & Bradstreet Austria* p 2 <https://www.ris.bka.gv.at/Dokumente/Lvwg/LVWGT_WI_20220211_VGW_101_042_791_2020_44_00/LVWGT_WI_20220211_VGW_101_042_791_2020_44_00.pdf> accessed 8 February 2024.
[918] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1986) 369.
[919] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 8.
[920] Hildebrandt Mireille, Koops Bert-Jaap, 'The challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) Vol. 73 (3) The Modern Law Review 428, 436.
[921] Brian Jeffrey Fogg, *Persuasive Technology*: *Using Computers to Change What We Think and Do* (EBSCO Publishing 2003), 1.
[922] Ibid 38.
[923] Shoshana Zuboff, *The age of surveillance capitalism* (Public Affairs 2019) 204, 293.
[924] Brian Jeffrey Fogg, *Persuasive Technology*: *Using Computers to Change What We Think and Do* (EBSCO Publishing 2003) 16.

Machines may predict the emotional states of individuals which, in turn, can easily be linked with other information.[925] Advanced behavioural inference systems as discussed in Section 2.2.4.3 powered by AI (particularly AC, DL and CV) allow fine-grained tracking of shoppers' behaviour, enabling retailers to analyse the collected data for profiling purposes and place personalised offers based on nuanced insights of individuals' behaviour and profiles.[926] For example, Facebook developed CV and AC-powered systems that feed staff in a retail store with information on their customers based on customers user profiles. The information can include detected emotions of the customers and enable retailers to target them with specific products informed by their Facebook activity and detected emotional states.[927] Because emotions play an important role in the elicitation of autonomous motivated behaviour,[928] AC may be used against the interests of the person concerned, namely, to persuade or manipulate this individual.[929] Understanding emotions increases the scope to influence decision-making of individuals and makes practices such as manipulation more effective.[930] Applications of AC may affect people's decisions and lives in ways that undermine their autonomy because emotions can influence decision-making powerfully, predictably and pervasively.[931] To be autonomous presupposes that a person has the capacity of self-reflection and rationality. A person must also enjoy 'procedural independence', meaning not to be under the influence of factors that comprise her capacities for self-reflection and rationality.[932] Information about a person's emotional state has implications for procedural independence: if it becomes available, it can restrict options in ways that a person would not choose herself.[933] The capacity for emotion to influence decision-making, combined with the ability to detect emotion by means of AC, strongly impacts an individual's personal autonomy.[934]

With the help of AI, manipulation and persuasion can be automated. Research suggests that intelligent software agents can significantly influence human behaviour.[935] Automated manipulation or

[925] Holger Baumann, Sabine Dörig, 'Emotion-Oriented Systems and the Autonomy of Persons' in Paolo Petta, Catherine Pelachaud, Roddie Cowie (eds) *Emotion-Oriented Systems* (Springer 2011) 745.
[926] Vasilios Mavroudis, Michael Veale 'Eavesdropping Whilst You're Shopping: Balancing Personalisation and Privacy in Connected Retail Spaces' (Living in the Internet of Things Conference, London, March 2018)1, 2 <https://ieeexplore.ieee.org/document/8379705> accessed 8 February 2024.
[927] Katie Gibbons, 'Facebook develops facial recognition cameras that feed shop staff their customers' profile details' *The Times* (London, 01 December 2017) <https://www.thetimes.co.uk/edition/news/facebook-develops-facial-recognition-cameras-that-feed-shop-staff-their-customers-profile-details-58lx0jckt> accessed 8 February 2024.
[928] Leen Vandercammen et al, 'On the Role of Specific Emotions in Autonomous and Controlled Behaviour' (2014) Vol 28 Iss 5 European Journal of Personality 413, 445.
[929] Rosalind W Picard, *Affective Computing* (MIT Press 1997) 136.
[930] Andrew McStay, Lachlan Urquhart 'This time with feeling? Assessing EU data governance implications of out of home appraisal based emotional AI' (2019) Vol 24 No 10 First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/9457/8146> accessed 8 February 2024.
[931] Jennifer S Lerner et al, 'Emotion and Decision Making' (2015) Vol 66 Annual Review of Psychology 799, 802.
[932] Holger Baumann, Sabine Dörig, 'Emotion-Oriented Systems and the Autonomy of Persons' in Paolo Petta, Catherine Pelachaud, Roddie Cowie (eds) *Emotion-Oriented Systems* (Springer 2011) 735, 736, 739.
[933] Roddy Cowie, 'Ethical Issues in Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 340.
[934] Damian Clifford, 'Citizen-consumers in a personalised Galaxy: Emotion influenced decision-making, a true path to the dark side?' (2017) CiTiP Working Paper 31/2017, 13 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037425> accessed 8 February 2024.
[935] Christopher Burr, Nello Cristianini, James Lydmann, 'An Analysis of the Interaction Between Intelligent Software Agents and Human Users' (2018) Vol 28 Minds and Machines 735, 752.

persuasion may lead to so-called 'impulse buying', where individuals make their decisions spontaneously and are dominated by emotions. Impulsive buying occurs when individuals experience an urge to buy a product, without thoughtful consideration why one needs a specific product.[936] In 2013, Amazon was granted a US patent called 'method and system for anticipatory package shipping'.[937] Based on AI-powered profiling applications that analyse a customer's historical shopping patterns by means of ML, Amazon predicts whether a customer is interested in a certain product. Then, Amazon sends that product to the customer, even if there has not been an order placed beforehand. In some situations, e.g. if the customer is 'particularly valued (e.g., *according to past ordering history, appealing demographic profile, etc.*), delivering the package to the given customer as a promotional gift may be used to build goodwill.'[938] Arguably less 'valued' customers can be provided with a discount in order to convert the potential interest in an order.[939] This is a prime example of how AI-powered profiling may be used to proactively push an individual to make a certain decision. Such profiling predicts the individual's interests and is then used to intentionally and covertly influence the person's decision-making, i.e. pushing to buy a certain product. AI-powered profiling undermines the sense of autonomy that consumers seek in their decision-making. The autonomy in choice is akin to exercising free will, and self-determination is a state of exercising one's autonomy.[940] Aggregation and analysis of data by means of profiling powerfully enhance the range of influence that marketers can have in shaping people's choices and actions.[941]

The examples in the previous paragraphs outline that AI-powered profiling may influence individuals in ways that adversely affect their autonomy and capacity to understand and author their own lives.[942] Treating individuals based on information gained from AI-powered profiling may (i) limit the choices available to an individual[943] or (ii) proactively push an individual towards a certain decision. This impacts the individual's ability to make rational and uncoerced choices and decisions.[944] It deprives data subjects not only of the means to reflect on the choices the environment makes for them, but

[936] Verhagen Tilbert, van Dolen Willemijn 'The influence of online store beliefs on consumer online impulse buying: A model and empirical application' (2011) Vol. 48 Information & Management 320.
[937] Spiegel Joel et al., 'Method and System for anticipatory Package Shipping' US Patent US 8615473B2 (Assignee: Amazon Technologies, Inc.) December 2013 <https://patentimages.storage.googleapis.com/8a/67/ff/299703230243b5/US8615473.pdf>, accessed 8 February 2024.
[938] Spiegel Joel et al., 'Method and System for anticipatory Package Shipping' US Patent US 8615473B2 (Assignee: Amazon Technologies, Inc.) December 2013 <https://patentimages.storage.googleapis.com/8a/67/ff/299703230243b5/US8615473.pdf>, accessed 8 February 2024.
[939] Ibid.
[940] André Quentin et al, 'Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data' (2018) Vol 5 Customer Needs and Solutions 28, 29.
[941] Helen Fay Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2010) 83.
[942] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 13.
[943] Tal Z. Zarsky, 'Mine your own business!' (2003) 5 Yale Journal of Law and Technology 35
[944] Maurits Clemens Kapitein,'Personalized Persuasion in Ambient Intelligence' (Doctoral Thesis, TU/e Eindhoven 2012) 179 < https://pure.tue.nl/ws/files/3470131/729200.pdf> accessed 8 February 2024.

proactively impact the choices they make.[945] An individual is autonomous when its decisions and actions are its own and thus self-determined.[946] In the examples mentioned, the individual is no longer autonomous. Individuals no longer act themselves; instead, they are acted upon.[947]

By not requiring controllers to inform data subjects about the significance and consequences of regular profiling, Articles 13 (2) lit f and 14 (2) lit g GDPR fail to achieve the objectives of the transparency principle.[948] These provisions also neglect possible harms of regular profiling, in particular the sharpening of power and information asymmetries and the adverse effects on the data subject's personal autonomy. Ultimately, the concept of control is a common denominator of transparency, power symmetry and autonomy.[949] The concept of control is not defined in the GDPR, although it was one of the main reasons for the data protection reform[950] and constitutes one of the GDPR's legislative aims, namely, that 'natural persons should have control of their own personal data'.[951] The GDPR does not contain an enforceable right specifically dedicated to the concept of control. Control seems to emerge from the concept of informational self-determination. It was interpreted as individual informational control or empowerment, i.e. the ability of a natural person to control the terms under which their personal information is acquired and used.[952] Control in this sense is subsequently often presented as the hallmark of data protection law[953] and is attributed with the role of a normative anchor for personal data protection as a fundamental right.[954]

Control-related provisions in data protection law can be classified in two mechanisms: consent and data subject rights.[955] In fact, control in the context of the fundamental right to data protection, and particularly as implemented in the GDPR, grants data subjects the possibility to act,[956] i.e. to invoke their data subject rights enshrined in Articles 15-22 GDPR or enforce their rights to lodge a complaint with a SA or their right to an effective judicial remedy against the controller (Articles 77 and 79

[945] Mireille Hildebrandt, Bert-Jaap Koops, 'The challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) Vol. 73 (3) The Modern Law Review 428, 435.
[946] Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge University Press 1988) 13.
[947] See Berlin, which explains the concept of autonomy under the heading positive liberty: 'Isaiah Berlin, *Liberty* (Hendry Hardy ed Oxford University Press 1969) 185; Marijn Sax, *Between Empowerment and Manipulation* (Kluwer Law International B.V. 2021) 131.
[948] Namely enabling data subjects to (i) become aware of processing according to Recital 39 GDPR, (ii) enforce their rights according to Articles 15-22 GDPR as well as remedies contained in Articles 77-80 GDPR and (iii) exercise control over the processing of their personal data Recital 7 GDPR.
[949] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 48.
[950] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona paras 69, 70.
[951] Recital 7 GDPR.
[952] Mary J Culnan, 'Protecting Privacy Online: Is Self-Regulation Working?' (2000) Vol 19 Iss 1 Journal of Public Policy & Marketing 20-26.
[953] Antoinette Rouvroy, Yves Poullet, 'The Right to Informational Self-determination and the Value of Self-development: Reassessing the Importance of Privacy for Democracy' in Serge Gutwirth et al (eds), *Reinventing Data Protection?* (Springer 2009) 68; Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 55.
[954] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 54; Stefano Rodotà, 'Data Protection as a Fundamental Right' in Serge Gutwirth et al (eds), *Reinventing Data Protection?* (Springer 2009) 79; Orla Lynskey, *The Foundations of EU Data Protection Law* (OUP 2015).
[955] Christophe Lazaro, Daniel Le Métayer, 'The Control over Personal Data: True Remedy or Fairy Tale?' (2015) Vol 12 Iss 1 SCRIPT-ed 1, 16–17; Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 59.
[956] Julie E Cohen, 'Affording Fundamental Rights' (2017) Volume 4 Iss 1 Critical Analysis of Law 78, 81

GDPR). Data subjects need to invoke their rights to exercise control over the processing of their personal data. Therefore, control in the sense of the GDPR seems to be rather limited from a conceptual point of view. In a preliminary ruling, even the AG stated that 'the scope for individual action is limited' and 'confined to the exercise of those rights in specified circumstances'.[957] The AG interprets the concept of control under the GDPR as 'rights of supervision and intervention in operations carried out by others on their data, as one tool […] for the protection of those data'.[958] Also, consent, the other mechanism for data subjects to exercise control over processing, is rather limited. Consent is just one of the legal bases in the GDPR and simply empowers the data subject to accept or reject the processing of personal data suggested by a controller. It does not otherwise empower them to intervene or influence how controllers process their personal data.[959] In my view, enforceable data subject rights are the main, though limited, mechanism for data subjects to exercise control over the processing of their personal data under the GDPR.

Articles 13 and 14 GDPR fail to achieve the GDPR's objective for data subjects to be able to exercise control over the processing of their personal data.[960] Transparency is a necessary precondition for control,[961] and without being informed about the significance and possible consequences of profiling, data subjects cannot exercise control over processing by enforcing their rights (e.g., object to profiling or lodging a complaint with an SA). It could be argued that controllers need to inform data subjects about the significance and possible consequences of profiling based on Article 22 (3) GDPR. According to this provision, controllers need to 'implement suitable measures to safeguard the data subject's right and freedoms', enabling them to obtain human intervention, to express their point of view and to contest automated decision-making. However, this obligation is only triggered if profiling involves automated decision making in the sense of Article 22 GDPR. 'Regular profiling' as defined in Article 4 (4) GDPR does not trigger the obligation contained in Article 22 (3) GDPR (see also Section 5.11). This is problematic because the concept of control is a common denominator of transparency, power symmetry and autonomy.[962] With transparent data processing and effective individual control over processing of personal data, data subject's risks to autonomy generally and manipulation particularly could be reduced.[963] Therefore, these provisions are not fit for purpose to effectively protect[964] the fundamental right to data protection. The CJEU has repeatedly stressed that EU data protection law

[957] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona para 72.
[958] Ibid para 71.
[959] Ibid para 73.
[960] Recital 7 GDPR.
[961] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 48.
[962] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 48.
[963] Frederik Zuiderveen Borgesius, 'Improving Privacy Protection in the area of Behavioural Targeting' (Doctoral thesis, Universiteit van Amsterdam 2015) 127 <https://dare.uva.nl/search?identifier=c74bdba6-616c-4cd9-925e-33a5858935e5> accessed 8 February 2024.
[964] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.

aims to effectively protect the data subject's personal data against risk of misuse.[965] Because data subjects are not informed about the significance and consequences of regular profiling, they cannot exercise control[966] over such processing. Therefore, misuse of personal data with adverse effects on personal autonomy cannot be prevented. Articles 13 and 14 fail to ensure a high level of protection[967] and provide data subjects with control over the processing of their personal data.

The profiling problem (Type 3)

ML and AC facilitate profiling as defined in the GDPR. Articles 13 (2) lit f and 14 (2) lit g GDPR do not require controllers to inform data subjects about the significance and consequences of profiling not involving ADM. These provisions sharpen the information asymmetries between controllers and data subjects instead of mitigating them, which may lead to adverse effects on the data subject's personal autonomy. The transparency principle embodied in Articles 13 & 14 GDPR is not fit for purpose to effectively protect the fundamental right to data protection.

4.5 Purpose limitation

The purpose limitation principle as introduced in Section 3.3.3.4 demands data to be collected for specified, explicit and legitimate purposes. In addition, personal data shall not be further processed in a manner which is *incompatible* with those legitimate purposes.[968]

4.5.1 Legal problems: Type 1

Generally, all AI disciplines are at odds with the purpose limitation principle. This conflict has not remained unnoticed in academia.[969] Natural language processing (NLP) relies on the processing of text or speech originating from conversations in various contexts; AC relies on video footage recorded during job interviews to detect emotional states; CV uses CCTV footage initially recorded for security purposes to identify individuals based on their gait. AR is devoted to answering questions from diverse data without human intervention, including decision-making.

The tension with the purpose limitation principle particularly applies to ML, which extracts models and properties from training data and recursively derives more data. Thus, data often goes through a

[965] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[966] Recital 7 GDPR.
[967] Recitals 6, 10 GDPR; Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[968] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 315.
[969] For an overview, see Footnote 27 in Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 4 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.

series of computations arguably for varying purposes.[970] The requirement stemming from the purpose limitation principle that personal data shall be processed for predefined *explicit* purposes is difficult if not impossible to comply with in the context of ML. The explicitness requirement demands controllers to clearly reveal, explain and expose the processing purpose to ensure an unambiguous understanding of the purpose. Notably, the purpose must be explicit and determined at the time of data collection i.e. ex-ante.[971] Unsupervised ML processes data for *in*explicit purposes – the processing *itself* determines the purpose since its goal is to detect patterns and correlations, gain knowledge and make accurate predictions. This makes it impossible to comply with the explicitness requirement ex-ante, i.e. at the time of data collection.

Purpose *specification* is particularly challenging to reconcile with unsupervised ML because it is often used without very specific objectives.[972] Thus, the challenges of defining a purpose for processing and only using the corresponding personal data for that purpose are exacerbated.[973] As indicated in regulatory guidance, it may be impossible to predict what the algorithm will learn, and the purpose may alter given that algorithms used in AI learn and develop over time.[974] Unsupervised ML seems to be at odds with the very core of the purpose limitation principle because it aims to identify associations and patterns among a set of input data. This would be the case if a bank uses unsupervised ML in order to identify associations and patterns in Facebook activities of its potential customers that could be useful for the bank.[975] In general, unpredictability of outcomes in the context of ML processing is considered one of the characteristic features of ML analytics.[976] ML leads to the discovery of patterns that were unimageable previously.[977] Unsupervised ML processes data for *un*specified and *in*explicit purposes – the processing *itself* predicts the purpose of the future use of the data since its goal is to detect patterns and correlations, gain knowledge and make accurate predictions. However, processing personal data for unspecified purposes as in the case of unsupervised ML is unlawful because the

[970] Yinzhi Cao, Junfeng Yang, 'Towards Making Systems Forget with Machine Unlearning' (IEEE Symposium on Security and Privacy, San Jose, May 2015) <https://www.ieee-security.org/TC/SP2015/papers-archived/6949a463.pdf> accessed 8 February 2024.
[971] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 68, 70 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[972] Lee A Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' (2021) University of Oslo Faculty of Law Legal Studies Research Paper Series No. 202-35 at 22 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3721118> accessed 8 February 2024.
[973] Christopher Kuner et al, 'Expanding the artificial intelligence-data protection debate' (2018) Vol 8 No 4 International Data Privacy Law 289, 290.
[974] Norwegian Data Protection Authority, 'Artificial Intelligence and Privacy' (2018) 4 <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf> accessed 8 February 2024.
[975] Norwegian Data Protection Authority, 'Artificial intelligence and privacy' (2018) 17 <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf> accessed 8 February 2024.
[976] Nadezha Purtova, 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) Vol 10 No 1 Law, Innovation and Technology 40, 56.
[977] Whereas Zarsky draws this conclusion in the context of Big Data, it is also valid with regard to ML, because its aim is to detect and extrapolate patterns; Tal Z Zarsky 'Incompatible: The GDPR in the Age of Big Data' (2017) Vol 47 Iss 4 Seton Hall Law Review 996, 1006.

scope of processing is not sufficiently delineated.[978] The purpose limitation principle prohibits unspecified processing and the explicitness requirement demands controllers to ensure an unambiguous understanding of the processing purpose at the time of data collection.[979] This violates the purpose limitation principle and leads to a Type 1 legal problem.

The inexplicitness problem (Type 1)

All AI disciplines process personal data originating from various sources for a plethora of other purposes. Also, ML processes personal data for unspecific and inexplicit purposes – the processing itself determines the purpose and future use of the personal data. Such processing violates the purpose limitation principle.

AI is prone to cause function creep and secondary use. Function creep refers to situations where 'previously authorised arrangements…now being applied to purposes and targets beyond those envisaged at the time of installation.'[980] In the context of data protection law, function creep occurs when personal data initially collected for a specific purpose are subsequently used beyond what was originally understood and considered socially, ethically and legally acceptable.[981] Secondary use, i.e. using data for purposes other than the initial collection purpose, could be seen as a violation of the purpose limitation principle according to data protection law.[982] Function creep is prohibited when such secondary use goes beyond the purposes specified in advance,[983] if the purpose for further processing is not compatible with the initial purpose (see Section 4.5.2). As already outlined in Chapter 2, data needed for the development and deployment of AI are enormous. AI relies on data from different sources initially collected for different purposes.[984] In addition, ML extracts models and properties from training data and recursively derives more data. Thus, data often goes through a series of computations arguably for different purposes.[985] However, the purpose limitation principle demands data to be collected for specified, explicit and legitimate purposes and not further processed in a manner which is incompatible with those purposes.[986]

[978] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 315.
[979] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 58, 68, 70 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[980] Richard Fox, 'Someone to watch over us: Back to the panopticon?' (2001) Vol 1 Iss 3 Criminal Justice 251, 261.
[981] Johanne Yttri Dahl, Ann Rudinow Sætnan, 'It all happened so slowly – On controlling function creep in forensic DNA databases' (2009) Vol 37 International Journal of Law, Crime and Justice 83, 84.
[982] Frederik Zuiderveen Borgesius, 'Improving Privacy Protection in the area of Behavioural Targeting' (Doctoral thesis, Universiteit van Amsterdam 2015) 117 <https://dare.uva.nl/search?identifier=c74bdba6-616c-4cd9-925e-33a5858935e5> accessed 8 February 2024.
[983] Bart Custers, Helena Ursic, 'Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection' (2016) Vol 6 Iss 1 International Data Privacy Law 1, 6.
[984] CIPL, 'Artificial Intelligence and Data Protection in Tension' (2018) 13 <https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ai_first_report_-_artificial_intelligence_and_data_protection_in_te....pdf> accessed 8 February 2024.
[985] Yinzhi Cao, Junfeng Yang, 'Towards Making Systems Forget with Machine Unlearning' (IEEE Symposium on Security and Privacy, San Jose, May 2015) <https://www.ieee-security.org/TC/SP2015/papers-archived/6949a463.pdf> accessed 8 February 2024.
[986] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 315.

AI seems to pose inherent risks of function creep. Smart home technologies based on AI such as Google's 'Nest thermostat' collect data about residents' behaviour and gather data from other connected products or devices such as cars, ovens, lights, TVs, game consoles, kettles, fitness trackers, beds and Google's digital assistant.[987] Collected data can be shared with Google's patented 'Privacy-aware personalised content for the smart home' AI system, which enables secondary use of collected data by companies to draw inferences from the generated home data (e.g. when residents are at home, when they shower, when they cook, when they watch TV and when they sleep). The patent states that 'the answers to these questions may help third parties benefit consumers by providing them with interesting information, products and services as well as with providing them with targeted advertisements'.[988]

Secondary use of data is also likely to occur in the context of virtual assistants that deploy NLP and speech recognition techniques based on RL and approaches from the specific kind of ML called deep learning (DL). Amazon's US patent 'Keyword Determinations from Voice Data'[989] indicates such secondary use of data. The patent describes a system that can capture voice content when a user speaks into or near the device (e.g., Alexa), notably without activating the virtual assistant by mentioning the 'wake word' (e.g., 'hey Alexa'). Sniffer algorithms attempt to identify trigger words that indicate statements of preference (such as like or love) and translate them into keywords. The identified keywords can subsequently be transmitted to a location accessible to advertisers, who can use the keywords to select content that is likely relevant to the user.[990] Amazon has denied that it uses voice recordings for advertising and mentioned that the patent might never actually come to the market.[991] Nevertheless, incidents unveiled in the press imply that such secondary use already takes place. For example, a journalist in the US reported that she was discussing a specific kitchen gadget with her husband and some neighbours within the reach of Alexa and received ads on Amazon for the kitchen gadget the next day.[992] A marketing team within media giant Cox Media Group claims it can listen to ambient conversations of consumers through embedded microphones in smartphones, smart TVs, and

[987] Shoshana Zuboff, *The age of surveillance capitalism* (PublicAffairs 2019) 7.
[988] Zomet Asaf, Urbach Shlomo Reuben, 'Privacy-Aware Personalised Content for the Smart Home' US Patent Number US 10'453'098 (Assignee: Google LLC) October 2019 <US20160260135A1 - Privacy-aware personalized content for the smart home - Google Patents> accessed 8 February 2024.
[989] Edara Kiran, 'Key Word Determinations From Voice Data' US Patent Number US 8798995B1 (Assignee: Amazon Technologies, Inc.) August 2014 <https://patentimages.storage.googleapis.com/bd/ed/2b/c4c67cc5a9f1ab/US8798995.pdf> accessed 8 February 2024.
[990] Edara Kiran, 'Key Word Determinations From Voice Data' US Patent Number US 8798995B1 (Assignee: Amazon Technologies, Inc.) August 2014 <https://patentimages.storage.googleapis.com/bd/ed/2b/c4c67cc5a9f1ab/US8798995.pdf> accessed 8 February 2024.
[991] Griffin Andrew, 'Amazon files for Alexa patent to let it listen to people all the time and work out what they want' *The Independent* (London, 11 April 2018) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/amazon-alexa-patent-listening-to-me-facebook-phone-talking-ads-a8300246.html> accessed 8 February 2024.
[992] Morgan Blake, 'Are Digital Assistants Always Listening?' *Forbes* (New York, 5 February 2018) <https://www.forbes.com/sites/blakemorgan/2018/02/05/are-digital-assistants-always-listening/#2f000e1a4eeb> accessed 8 February 2024.

other devices to gather data and use it to serve targeted ads.[993] In addition, it is not a secret that companies such as Google, Amazon, Meta and Apple maintain and improve their voice recognition devices and software by means of assessing various audio snippets recorded by such devices.[994] For example, Amazon has publicly confirmed to manually review Alexa requests to confirm that Alexa understood and responded correctly.[995]

The function creep problem (Type 1)

Particularly the AI disciplines ML and NLP significantly contribute to function creep and secondary use of personal data, which violates the purpose limitation principle.

4.5.2 Legal problems: Type 2

When the purpose limitation principle is applied to the AI disciplines introduced in Chapter 2, no specific Type 2 legal problems arise. Judicial guidance is scarce with respect to the criteria to be applied on the precision of the purpose. Research suggests that ECtHR case law makes the precision of the purpose dependent on the extent to which the data subject is affected by the processing.[996] Although the requirement to specify the purpose is a 'key element in the implementation of the European regime for the protection of personal data',[997] it does not seem to play a prominent role in CJEU case law. Cases dealing with purpose limitation do not specifically deal with the specification of purposes.[998] This is problematic when considering that the purpose specification requirement plays a central role in data protection law as all data protection principles depend on it.[999] In addition, the EU legal framework itself does not provide explicit criteria in order to determine how precisely the purposes should be specified.[1000] According to regulatory guidance, purposes which are too vague or general do not meet the criteria of being specific. For example, the guidance refers to 'elastic purposes' sometimes used by controllers such as 'future research', 'product innovation' and 'improving user experience'.[1001]

[993] Joseph Cox, 'Marketing Company Claims That It Actually Is Listening to Your Phone and Smart Speakers to Target Ads' *404 Media* (United States, 14 December 2023) <Marketing Company Claims That It Actually Is Listening to Your Phone and Smart Speakers to Target Ads (404media.co)> accessed 8 February 2024.
[994] Tine Munk, 'Does Online Privacy Exist in the GDPR Era? The Google Voice Assistant Case' in Tatiana-Eleni Synodiou et al (eds) *EU Internet Law in the Digital Single Market* (Springer Nature 2021) 480.
[995] Dorian Lynskey, 'Alexa, are you invading my privacy? the dark side of our voice assistants' *The Guardian* (London, 9 October 2019) <https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants> accessed 8 February 2024.
[996] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 66, 162, 167 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[997] Case C-77/21 *Digi* [2022] ECR I-805 Opinion of AG Pikamäe para 40.
[998] Case C-77/21 *Digi* [2022] ECR I-805 para 27; Case C-175/20 *'SS' SIA* [2022] ECR I-124 para 64.
[999] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 102 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[1000] Maximilian von Grafenstein, *The Principle of Purpose Limitation in Data Protection Laws* (Nomos 2017) 232, 233, 244.
[1001] Art 29 Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203, 2 April 2013) at 16; Art 29 Working Party, 'Opinion 02/2013 on apps on smart devices,' (WP 202, 27 February 2013) at 23.

Whereas this regulatory guidance is certainly needed and welcome, it is legally not binding[1002] and is not judicially tested. Shortcomings in terms of purpose specification may lead to Type 2 legal problems because substantively unclear principles are difficult to enforce. Nevertheless, this problem arises regardless of whether the processing involves AI and thus does not relate specifically to AI. Therefore, I refrain from discussing this problem in further detail.

4.5.3 Legal problems: Type 3

The basic idea of the purpose limitation principle is to restrict the processing of personal data. In the words of AG Pikamäe, the purpose of this principle is to 'delimit as clearly as possible the use of personal data'.[1003] However, interdisciplinary research on the application of the purpose limitation principle in personalisation and profiling systems has revealed that purpose specification hardly restricts the ways in which personal data can be processed.[1004] Where controllers do their best to define purposes with enough specificity and can demonstrate that such purposes are legitimate,[1005] any purpose is a valid purpose under the GDPR. Thus, purpose limitation does not seem to be an appropriate legal tool to ensure data processing is restricted in data-driven systems. Instead, it is a procedural criterion that at least requires controllers to consider the need and implications of processing from the beginning.[1006] This Type 3 legal problem occurs regardless of which AI discipline the purpose limitation is applied to because the principle itself is not suitable to restrict the collection and further processing of personal data. Therefore, it is a general problem and relates to all AI disciplines as introduced in Chapter 2.

> ***The restriction problem (Type 3)***
> *The purpose limitation principle does not, as intended, restrict the collection and further processing of personal data. It thus fails to achieve its aim to limit data processing and is therefore not fit for purpose to effectively protect the fundamental right to data protection.*

The purpose limitation principle enshrines two requirements: (i) personal data must be collected for specified, explicit and legitimate purposes, and (ii) personal data must not be further processed for incompatible purposes.[1007] Apart from specifically privileged purposes, any processing taking place after collection constitutes 'further processing' and must comply with the principle of compatible

[1002] Footnote 40 refers to an opinion issued by Article 29 Working Party; see Joined Cases C-141/12 & C-372/12, *YS, M* [2014] ECR I-2081, Opinion of AG Sharpston para 49.
[1003] Case C-77/21 *Digi* [2022] ECR I-805 Opinion of AG Pikamäe para 27.
[1004] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 49 and 55 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1005] Which does not appear to be difficult.
[1006] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 49 and 55 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1007] Case C-77/21 *Digi* [2022] ECR I-805 Opinion of AG Pikamäe para 28; Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 58 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.

use.[1008] Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes are privileged purposes. They are a priori considered a lawful processing operation which is a compatible purpose[1009] provided that such processing is subject to appropriate safeguards.[1010]

Processing for compatible purposes does not require an additional legal basis[1011] and prevails over the interests of the data subject when she objects to such processing if it serves a public interest.[1012] Recital 159 GDPR envisages a broad interpretation of scientific research, including technological development, demonstration, fundamental and applied research and privately-funded research. Not only academic institutions but also profit-seeking companies can carry out scientific research based on this exception.[1013] Regulatory guidance requires that scientific research performed under this exception occurs in accordance with relevant sector-related methodological and ethical standards and in conformity with good practice.[1014] Whereas it is clear that publicly funded and externally published work at academic institutes fall under the research exception, [1015] this is less obvious for research performed at private companies. However, given the broad interpretation of scientific research derived from Recital 159 GDPR and relevant regulatory guidance, companies can argue that processing of personal data in the context of AI falls under the research exception.

Statistical purposes refer to the elaboration of statistical surveys or the production of statistical, aggregated results.[1016] Because ML is strongly based on statistics, it could be argued that further processing by means of ML constitutes processing for statistical purposes and is thus allowed without the need for an additional legal basis. Statistical purposes can be construed broadly, covering uses by companies for commercial gain and permitting to use this exception for big data applications and purposes.[1017] It seems that computer scientists do not come to terms whether ML is different from statistics. Some argue that ML is different from statistics, and others argue that statistics and ML are complementary.[1018] Also in the legal domain, the scope of the statistical purpose exception is not

[1008] Case C-77/21 *Digi* [2022] ECR I-805 Opinion of AG Pikamäe para 28.
[1009] Ibid, Footnote 14.
[1010] Article 89 GDPR.
[1011] Recital 50 GDPR; Waltraut Kotschy, Commentary of Article 6 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 341.
[1012] Article 21 (6) GDPR.
[1013] European Data Protection Supervisor, 'A Preliminary Opinion on data protection and scientific research' (2020) 11 <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf> accessed 8 February 2024.
[1014] Article 29 Working Party, 'Guidelines on consent under Regulation 2016/679' (WP 259rev.01, 10 April 2018) at 28.
[1015] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 51 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1016] Recital 162; Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 317.
[1017] Viktor Mayer-Schönberger, Yann Padova, 'Regime change? Enabling Big Data through Europe's new Data Protection Regulation' (2016) Vol 17 No 2 Science and Technology Law Review 315, 325-326.
[1018] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 52 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.

entirely clear. Some argue that it facilitates inferential analytics by means of ML approaches[1019] and the construction of ML models based on personal data.

Regulatory guidance states that the statistics exception applies in commercial settings and to 'analytical tools of websites or big data applications aimed at market research'.[1020] Personal data may be used to draw inferences and lead to a model, which then can be applied to other individuals, for example to take decisions.[1021] Whereas Recital 162 GDPR indicates that the result of processing for statistical purposes must be aggregated data, and this result must not be used to support measures or decisions regarding any particular person, its effect remains unclear. The question is what qualifies as a decision or measure in the latter sense. Both concepts require some binding effect, distinguishing them from mere recommendations.[1022] At least some forms of ML output could qualify to fall under the scope of the statistics exception, such as the prediction of customers ceasing their relationship with a company (customer churn). Whether the prediction of specific customer churn and subsequent action taken to avoid this also fall under the statistics exception is less clear[1023] since this might be considered 'a measure or decision regarding any particular person'.[1024] Targeted advertisement is another illustrative example. Displaying ads to individuals online based on their interests inferred by ML does not necessarily constitute a decision or measure regarding the individuals concerned. Arguably, such targeted ads are mere recommendations to purchase a product or subscribe to a service, lacking the binding effect of a measure or decision. In addition, the different processing stages of the ML pipeline seem to be relevant as ML produces aggregate and individual results at different processing stages.[1025] Furthermore, ML models are likely to fall under trade secrets protection and controllers could refrain from providing meaningful information (see Sections 5.6 and 5.6.2 below).

In addition, due to the opening clause contained in Article 89 GDPR, the scope of the statistical purpose exception might vary across EU Member States. Recital 162 GDPR demands the latter to 'determine statistical content, control of access, specifications for the processing of personal data for statistical purposes' within the limits of the GDPR. This opening clause and the corresponding implementation in the Member States lead to additional legal uncertainty besides the already considerable uncertainties regarding this exception.[1026]

[1019] Sandra Wachter, Brent Mittelstadt 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) No 2 Columbia Business Law Review 1, 550-551; see also 549, 592.
[1020] Art 29 Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203, 2 April 2013) at 29.
[1021] Sandra Wachter, Brent Mittelstadt 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) No 2 Columbia Business Law Review 1, 550-551; see also 549, 592.
[1022] For the notion of a decision in the sense of Article 22 GDPR see Case C-634/21, *SCHUFA Holding AG* [2023] ECR I-957 paras 44-46; see also corresponding Opinion AG Pikamäe para 37.
[1023] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 52 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1024] Recital 162 GDPR.
[1025] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 52 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1026] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 55 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.

Arguably, both the research and statistical purposes exception for further processing undermine the GDPR's aim that data subjects should have control over their own personal data.[1027] Scholars place the purpose limitation principle in light of the concept of control, as well as informational self-determination and autonomy.[1028] The concept of control is not defined in the GDPR, although it was one of the main reasons for the data protection reform[1029] and constitutes one of the GDPR's legislative aims.[1030] As outlined in Section 4.4.3, the main mechanism for data subjects to exercise control over the processing of their personal data under the GDPR are data subject rights. However, this mechanism is rather limited. AG Campos Sánchez-Bordona correctly notes that 'the scope for individual action is limited' and 'confined to the exercise of those rights in specified circumstances'.[1031] The principle of compatible use, and the privileged purposes concerning research and statistics in particular, hinders data subjects to enforce their rights and thus to exercise control over the processing of their data. Article 17 (3) GDPR states that the right to erasure does not apply if erasure of personal data is likely to render the achievement of the objectives of processing for research or statistical purposes impossible or seriously impair these objectives. In addition, processing of personal data for scientific and statistical purposes for the performance of a task carried out for reasons of public interests prevails over the data subjects' right to object to such processing.[1032] Therefore, the concept of compatible use undermines the individual's control over the processing of personal data because it allows one to further process personal data by means of ML. This is detrimental to the aim of GDPR to provide data subjects with control over their data[1033] and ultimately leads to a problem of Type 3, that is, the concept of compatible use is not fit for purpose to protect the fundamental right to data protection.

However, It could be argued that neither the GDPR nor the EUCFR contains a 'right of control' that transforms it into an illusory objective pursued by the GDPR and the EU's data protection reform. This criticism has its merits, but the concept of compatible use still leads to a type 3 legal problem. It undermines the GDPR's objective to protect natural persons from risks related to the processing of personal data. There are considerable uncertainties regarding the interpretations of the research and statistical purposes exception.[1034] Creative controllers will utilise these considerable uncertainties surrounding the concept of compatible use. This is detrimental to the GDPR's aim to effectively protect

[1027] Recital 7 GDPR.
[1028] For an overview, see Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 72 <https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y> accessed 8 February 2024.
[1029] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona paras 69, 70.
[1030] Recital 7 GDPR.
[1031] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona para 72.
[1032] Article 21 (6) GDPR.
[1033] Recital 7 GDPR.
[1034] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 55 <https://arxiv.org/abs/2101.06203>accessed 8 February 2024.

the fundamental right to data protection, which the CJEU emphasises.[1035] Neither do these uncertainties contribute to a high level of protection as envisaged by the GDPR.[1036]

The compatible use problem (Type 3)

Processing in the context of ML might fall under the concept of compatible use because it relates to the privileged statistical and/or research purposes. This undermines the data subject's control over the processing of personal data, which is detrimental to the GDPR's aim. The concept of compatible use is therefore not fit for purpose to protect the fundamental right to data protection.

4.6 Data minimisation

The data minimisation principle as introduced in Section 3.3.3.5 requires that personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.[1037] This wording indicates that the data minimisation principle is a manifestation of the proportionality principle as introduced in Section 3.2.3. In the CJEU's words, the data minimisation principle 'gives expression to the principle of proportionality'.[1038]

4.6.1 Legal problems: Type 1

Quite contradictory to the data minimisation principle introduced in Section 3.3.3.5, AI needs substantial amounts of data in order to operate effectively, particularly in the training phase.[1039] AI has an 'insatiable appetite' for data and contradicts the data minimisation principle.[1040] Advanced AI applications employing complex models such as deep learning (DL) and natural language processing (NLP) need to learn many parameters and require *enough* data to function properly.[1041] As outlined in Section 2.2.1, *accurate* predictions are the main goal of data processing in ML. The underlying algorithm is decisive in terms of the required amount of data. DL, a particular kind of ML, requires large-scale training data.[1042] DL applications using the supervised training method in NLP for speech

[1035] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73; Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1036] Recitals 6, 10 GDPR; Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[1037] Article 5 (1) lit c GDPR.
[1038] Case C-439/19 *B* [2021] ECR I-504 para 98; Case C-175/20 *'SS' SIA* [2022] ECR I-124 para 83.
[1039] CIPL, 'Artificial Intelligence and Data Protection How the GDPR Regulates AI' (2020) 13 <https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf> accessed 8 February 2024.
[1040] Christopher Kuner et al, 'Expanding the artificial intelligence-data protection debate' (2018) Vol 8 No 4 International Data Privacy Law 289-292.
[1041] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 57 and 58 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1042] Zhou Zhi-Hua, Feng Ji, 'Deep Forest: Towards an Alternative to Deep Neural Networks' (IJCAI Conference, Melbourne, August 2017) 1 <https://www.ijcai.org/proceedings/2017/0497.pdf > accessed 8 February 2024.

recognition require large amounts of training data with labels.[1043] Computer vision (CV) heavily relies on the processing of photographs, in particular for facial recognition and automated face analysis systems used in the AI discipline affective computing (AC). Moreover, data analytics in the context of ML does not only require vast amounts of data, but also causes more data processing and therefore creates a closed circle: with more data, more accurate models can be trained, which generates more services and users of those services, which leads to more data being processed.[1044] Ultimately, AI violates the data minimisation principle, which constitutes a Type 1 legal problem.

The data appetite problem (Type 1)

AI has an insatiable appetite for data. Contrary to the data minimisation principle, AI and particularly DL requires substantial amounts of data to function well and generate accurate output. This violates the data minimisation principle.

However, the data appetite problem does not suggest that AI and the data minimisation principle are per se incompatible. Instead, applying data minimisation to complex AI systems is difficult. This is to a significant extent due to the current incomputability of data protection principles[1045] (Section 4.7.3). It is challenging to determine which data are necessary when personal data are processed in the context of AI and thus to limit such data accordingly. The problem with data minimisation and AI lies at the core of this principle, namely, how to exactly define what should be considered necessary for processing activities based on AI applications.[1046] Recital 39 relating to the data minimisation principle simply states that personal data 'should be processed only if the purpose of the processing could not reasonably be fulfilled by other means.' It is difficult for computer scientists to determine in a given computation of individual pieces which personal data are adequate, relevant and necessary. Consequently, computer scientists apply various, often inconsistent, approaches to the data minimisation principle.[1047] This becomes most apparent in the case of unsupervised ML that processes data for *un*specified and *in*explicit purposes. With unsupervised ML, the processing *itself* determines the purpose and future use of the data since its goal is to detect patterns, correlations, gain knowledge and make accurate predictions. Thus, in the context of unsupervised ML, the purpose of processing

[1043] Deng Li and Liu Yang, 'Epilogue: Frontiers of NLP in the Deep Learning Era' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 1.
[1044] Zhao Jianxin et al., 'Privacy-preserving Machine Learning Based Data Analytics on Edge Devices' (AIES Conference, New Orleans, January 2018) 1 <http://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_161.pdf > accessed 8 February 2024.
[1045] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 58 and 60 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1046] Ronald Leenes, Silvia De Conca, 'Artificial intelligence and privacy – AI enters the house through the Cloud' in Woodrow Barfield, Ugo Pagallo (eds) *Research handbook on the law of artificial intelligence* (Edward Elgar Publishing Inc. 2018) 299, See also Mireille Hildebrandt, 'Primitives of legal protection in the era of data-driven platforms' (2018) 13 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3140594> accessed 8 February 2024.
[1047] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 59 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.

is not yet known and there is no supervisor who directs the machine on the purpose for processing.[1048] It cannot be determined what pieces of personal data are necessary for yet a unknown purpose.

Even if the purpose is known and defined as 'development of AI systems', limiting the use of personal data necessary to achieve this purpose seems illusory due to the insatiable appetite for data of AI (see the data appetite problem). In addition, the purpose 'development of AI systems' arguably does not meet the criteria of being 'specific'. This purpose appears to 'elastic' as controllers use phrases such as 'future research', 'product innovation', 'improving user experience', which regulators are likely to consider as too vague or general.[1049] In addition, such an elastic purpose is not suitable for proportionality decisions as required by the data minimisation principle, namely, to limit the processing of personal data to what is necessary in relation to that purpose because the purpose specification requirement is a precondition for that proportionality assessment.[1050] As a consequence, the data minimisation principle is violated. This constitutes a Type 1 legal problem.

The necessity problem (Type 1)

In the case of unsupervised ML, it is impossible to determine whether a given computation of specific pieces of personal data is necessary, and to limit the personal data processed in accordance with the proportionality principle. Such processing violates the data minimisation principle.

4.6.2 Legal problems: Type 2

Verifying whether a controller complies with the data minimisation principle is technically difficult, if not impossible. The complexity of models adopted by AI represents a major challenge for human cognition.[1051] AI equipped systems are becoming highly opaque black boxes and individuals are unable to follow the steps these machines are taking to reach whatever conclusions they reach.[1052] DL methods based on artificial neural networks (ANN) generally lack interpretability[1053] and are particularly challenging due to their hierarchical and nonlinear structure and the central concept in DL called connectionism. In connectionism, a large number of simple computational units (artificial neurons) achieve intelligent behaviour when networked together[1054] (see Section 2.2.1.4). It seems neither possible to understand which artificial neuron contributed to a distinct part of the output nor to understand

[1048] Similarly, see Christopher Kuner et al, 'Expanding the artificial intelligence-data protection debate' (2018) Vol 8 No 4 International Data Privacy Law 289, 290.
[1049] Art 29 Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203, 2 April 2013) at 16; Art 29 Working Party, 'Opinion 02/2013 on apps on smart devices,' (WP 202, 27 February 2013) at 23.
[1050] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 68, 108 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[1051] Zachary C Lipton, 'The Mythos of Model Interpretability' (2018) Vol 16 Iss 3 ACMQueue 18 <https://dl.acm.org/doi/pdf/10.1145/3236386.3241340?download=true> accessed 8 February 2024.
[1052] Amitai Etzioni and Oren Etzioni, 'Keeping AI Legal' (2016) 19 Vand. J. Ent. & Tech. L. 133, 137.
[1053] Deng Li and Liu Yang, 'A Joint Introduction to Natural Language Processing and Deep Learning' in Deng Li and Liu Yang (eds) *Deep learning in natural language processing* (Springer 2018) 11, 12.
[1054] Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning* (MIT Press 2016) 16 <www.deeplearningbook.org> accessed 8 February 2024.

what happened in the intermediate (hidden) layers of the ANN.[1055] When an ANN is used for pattern recognition in CV or NLP, an ex-post analysis of the model used will likely not establish linear causal connections which are comprehensible for human minds.[1056] Where the model used is not interpretable, it is difficult or impossible to verify whether the processing of individual pieces of personal data are adequate, relevant and necessary for a specific purpose according to the data minimisation principle. This cannot be mediated by the AI discipline of automated reasoning. As outlined in Sections 2.2.5, 4.3.1, 4.4.1 and 4.7.1, AI systems do not have a semblance of common sense or capabilities such as human cognition and are therefore unable to think in a manner on par with human thinking.[1057] Therefore, AI systems do not deploy arguments that may be used to determine which factors, for example, personal data, are necessary or relevant for generating certain output. Therefore, the data minimisation principle cannot be enforced, whether by means of private enforcement initiated by data subjects or in the form of regulatory enforcement pursued by SAs.

The verification problem (Type 2)

The reasoning deficiencies in AR and the complexity of models adopted by AI, particularly approaches from ML (specifically DL) as well as CV and NLP, render it difficult or impossible to verify whether the processing of personal data complies with the data minimisation principle. Consequently, the data minimisation principle cannot be enforced.

4.6.3 Legal problems: Type 3

When consequently applied, the data minimisation principle might negatively affect the accuracy principle as introduced in Section 3.3.3.6. In the context of a prediction system powered by ML, deciding that a certain piece of personal data should not be used might reasonably lead to inaccurate predictions,[1058] which violates the accuracy principle. However, it could be argued that both principles are not in conflict when the purpose is defined as 'processing *all data* necessary to make accurate predictions'. The purpose specification requirement plays a central role, also regarding the data minimisation principle. In my view, this purpose is not specific enough to effectively implement the data

[1055] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 155.
[1056] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer Nature 2020) 81.
[1057] Lance Eliot, 'AI Ethics And The Quagmire Of Whether You Have A Legal Right To Know Of AI Inferences About You, Including Those Via AI-Based Self-Driving Cars' *Forbes* (New York, 25 May 2022) < https://www-forbes-com.cdn.ampproject.org/c/s/www.forbes.com/sites/lanceeliot/2022/05/25/ai-ethics-and-the-quagmire-of-whether-you-have-a-legal-right-to-know-of-ai-inferences-about-you-including-those-via-ai-based-self-driving-cars/amp/> accessed 8 February 2024.
[1058] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 57 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.

minimisation principle.[1059] An insufficiently defined purpose leads to excessive processing of personal data and violates the data minimisation principle.[1060]

Similarly, there might be trade-offs between data minimisation and the fairness principle. To ensure fairness, it might be required to process more personal data than strictly necessary according to the data minimisation principle to guard against bias and error,[1061] for example, to avoid discrimination. An empirical study suggests that the decision to not collect data on gender or other protected attributes could make it challenging or impossible to identify discrimination against those groups once the ML algorithm has been deployed.[1062] Thus, minimisation of sensitive features such as gender may diminish the ability to detect unfairness,[1063] which is detrimental to the fairness principle. To figure out means that overcome such trade-offs requires creativity and reasoning skills. However, AI currently lacks reasoning capabilities that would allow to solve the difficult task of overcoming trade-offs between data protection principles. The trade-offs between principles combined with the reasoning deficiencies of AI lead to a Type 3 legal problem. Principles leading to trade-offs are not fit for purpose to effectively[1064] protect the fundamental right to data protection, to ensure a high level of the protection of personal data[1065] and to achieve the GDPR's aim to establish a strong and coherent data protection framework.[1066] This holds particularly true when considering that principles provide the basis for the protection of personal data.[1067] This Type 3 legal problem occurs regardless of which AI discipline the data minimisation, fairness and accuracy principles are applied to because they themselves create the trade-offs between each other. Therefore, it is a general problem and relates to all AI disciplines as introduced in Chapter 2.

[1059] Merel Elize Koning, 'The purpose and limitations of purpose limitation' (Doctoral thesis, Radboud University Nijmegen 2020) 102 < https://repository.ubn.ru.nl/bitstream/handle/2066/221665/221665.pdf?sequence=1&isAllowed=y > accessed 8 February 2024.
[1060] Art 29 Working Party, 'Opinion 01/2014 on the application of necessity and proportionality concepts and data protection within the law enforcement sector' (WP 211, 27 February 2014), at 18; Art 29 Working Party, 'Opinion 05/2013 on Smart Borders' (WP 206, 6 June 2013) at 10.
[1061] Christopher Kuner et al, 'Expanding the artificial intelligence-data protection debate' (2018) Vol 8 No 4 International Data Privacy Law 289, 290.
[1062] Gemma Galdon Cavell et al, 'Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization' (Proceedings of the AAAI/ACM Conference on AI, Ethics and Society, New York 2020) 266 <https://dl.acm.org/doi/pdf/10.1145/3375627.3375852> accessed 8 February 2024.
[1063] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 59 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024; Gemma Galdon Clavell et al, 'Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization', (2020) Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (ACM) <https://dl.acm.org/doi/10.1145/3375627.3375852> accessed 8 February 2024.
[1064] Recital 11 GDPR; Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1065] Recitals 6, 10 as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[1066] Recital 7 GDPR.
[1067] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.

The trade-off problem (Type 3)

When consequently applied in the context of AI, the data minimisation principle might lead to trade-offs with the accuracy and fairness principles. Due to the shortcomings in AR, AI currently lacks reasoning capabilities that may overcome these trade-offs, and may fail to adequately protect the fundamental right to data protection.

4.7 Accuracy

As outlined in Section 3.3.3.6, the GDPR states that the processing of personal data must be accurate.[1068] The accuracy principle intends to protect the individual concerned from being irrationally or unfairly treated based on wrong and inaccurate representations.[1069] According to regulatory guidance, accurate means 'accurate as to a matter of fact'.[1070] The need for personal data to mirror the reality with respect to the data subject concerned[1071] is also stressed in academia: personal data shall, at any given time, reflect reality.[1072] Case law[1073] of the CJEU indicates that the level of accuracy of personal data is determined by the purpose of the processing:[1074] the assessment whether personal data are accurate and complete depends on the *purpose* for which data were collected.[1075] Nevertheless, the precise substantive requirements of the accuracy principle remain an underexplored topic in academia, which is problematic when considering the developments in AI and its significance with regard to the right to rectification[1076] (see also Section 5.7). However, to apply the accuracy principle to the AI disciplines introduced in Section 2.2, I distinguish between two distinct types of accuracy. These are *absolute accuracy* referring to 'accurate as a matter of fact'[1077] aiming to reflect reality[1078] (e.g. date of birth) and *relative accuracy* which is more nuanced and determines accuracy based on the purpose of processing[1079] (e.g. data 'measured' by means of a percentage).

[1068] Art. 5 (1) lit d GDPR.
[1069] Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 9.
[1070] Art 29 Working Party, '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' (WP 225, 26 November 2014), at 15 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=667236> accessed 8 February 2024.
[1071] Ibid 15; Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 4.
[1072] Paul Voigt, Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer 2017) 91.
[1073] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.
[1074] Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 4.
[1075] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.
[1076] Diana Dimitrova, 'The rise of the personal data quality principle: is it legal and does it have an impact on the right to rectification?' (2021) Vol 12 No 3 European Journal of Law and Technology 2.
[1077] Art 29 Working Party, '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' (WP 225, 26 November 2014), at 15 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=667236> accessed 8 February 2024.
[1078] Paul Voigt, Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer 2017) 91.
[1079] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.

4.7.1 Legal problems: Type 1

ML is particularly problematic in the context of the accuracy principle. Companies increasingly offer services and products with embedded ML components which aim to predict behaviour of individuals or to detect correlations, for example with regard to credit risk and health status.[1080] Such services and products involve probabilistic predictions and detected correlations.[1081] Predictions can be defined as 'the output emitted by a model of a data generating process in response to specific configurations of input'.[1082] ML is deployed to draw inferences about the behaviours, preferences and private lives of individuals, information that can subsequently be used to nudge or manipulate individuals or to take decisions on them.[1083] Put simply, inference may be described as the process whereby a conclusion is drawn without complete certainty, but with some degree of probability.[1084] Any inferential method is built on assumptions[1085] which may be correct or not. Inference enables decision-making under conditions of uncertainty.[1086] Prediction and inference are inextricably linked to each other because inference involves the systematic comparison of predictions. Both industry and academic literature focus on predictions, in particular in the AI discipline ML.[1087] The very nature of inferences, predictions and correlations increases the risk of inaccuracy[1088] because of its probabilistic nature.[1089] To be clear, the output generated by ML does not necessarily equal inaccurate data. Suppose processing aims, as a purpose, to infer a chance of something happening in the future. In that case, the probabilistic nature of such a prediction does not automatically lead to a violation of the accuracy principle. Instead, the problem in terms of accuracy emerges when predictions are treated as facts, which is context-dependent. If such predictions or correlations are essentially considered as *facts* this might lead to detrimental effects for data subjects (e.g., when applying for a job or a loan). There is experimental evidence that humans closely follow algorithmic output and cannot correctly *assess* its quality. In this online experiment, 1,263 participants received algorithmic advice and were free to choose whether to incorporate this advice in their own response. Most of the participants followed the algorithmic

[1080] Pedreschi Dino et. al., 'Open the Black Box: Data-Driven Explanation of Black Box Decision Systems' (2018) 1 <https://arxiv.org/pdf/1806.09936.pdf > accessed 8 February 2024.
[1081] Viktor Mayer-Schönberger; Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Houghton Mifflin Harcourt 2013) 52-55.
[1082] Nathan Sanders, 'A Balanced Perspective on Prediction and Inference for Data Science in Industry' (2019) Iss 1.1 Harvard Data Science Review 1, 15 <https://assets.pubpub.org/zmmen09c/644ef4a4-5a71-43f8-9bcd-f2f6cb92ea65.pdf> accessed 8 February 2024.
[1083] Sandra Wachter, Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) Issue 2 Columbia Business Law Review 494, 497, 548, 549.
[1084] Michael P Cohen, 'Inference' in Paul J Lavrakas (eds) *Encyclopedia of Survey Research Methods* (Sage Publication, Inc 2008) 334.
[1085] Michael Betancourt, 'A Unified Treatment of Predictive Model Comparison' (2015) 1 <https://arxiv.org/pdf/1506.02273.pdf > accessed 8 February 2024.
[1086] Lawrence Hazelrigg, 'Inference' in Melissa Hardy, Alan Bryman (eds) *Handbook of Data Analysis* (Sage Publications 2004) 14.
[1087] Nathan Sanders, 'A Balanced Perspective on Prediction and Inference for Data Science in Industry' (2019) Iss 1.1 Harvard Data Science Review 1, 7, 21 <https://assets.pubpub.org/zmmen09c/644ef4a4-5a71-43f8-9bcd-f2f6cb92ea65.pdf> accessed 8 February 2024.
[1088] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 251rev.01, 6 February 2018), at 17.
[1089] Christopher Burr, Nello Cristianini, 'Can machines read our mind?' (2019) Vol 29 Iss 3 Minds and Machines 461, 483.

recommendation closely and never realised that the algorithm was purposely biased. The setup of the experiment enabled the participants to compare their prediction with the algorithm prediction, allowing them to realise that the algorithm is biased.[1090]

ML systems that aim to predict future behaviour of individuals cannot be designed with absolute accuracy due to their predictive nature and the lack of a truth as a baseline for comparison.[1091] Predictions generated by ML relate to future conduct that has *not yet happened*. Predictive accuracy will vary for each situation and this is not necessarily obvious for the ones who deploy the system or are subject to the system.[1092] What requires scrutiny is not the input data but rather the accuracy of the inferences drawn from input data,[1093] i.e. the output of the AI system. Finding correlations in data and acting on them is often considered to be good enough.[1094] Correlations based on a sufficient volume of data are increasingly seen as sufficiently credible to direct action without first establishing causality. Even if strong correlations or causal knowledge are found, this knowledge may only concern groups, whereas actions are directed towards individuals. This may lead to situations in which individuals are inaccurately described via simplified models or classes.[1095] Inferences or predictions can never be absolutely certain and are poorly verifiable or not verifiable at all (e.g. the individual is a 'high credit risk' or 'likely to buy a house in two years').[1096] Inference 'is always an invasion of the unknown, a leap from the known'.[1097] Admittedly, it might be argued that this also applies to inferences drawn by humans. However, human inferences are based on *human reasoning* and are usually not considered facts. Machine-generated inferences are more problematic because they are *not* based on human reasoning and are often treated as facts,[1098] although they are simply probabilistic and relate to future conduct that has not yet happened. Consider the following example which occurred in a case referred to the CJEU. Due to a poor credit score value allocated to a data subject, the mobile network

[1090] Jan Biermann, John Horton, Johannes Walter, 'Algorithmic Advice as a Credence Good' (2022) Centre for European Economic Research Discussion Paper No 22-071 at 2, 14 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4326911 > accessed 8 February 2024.
[1091] Diana Dimitrova, 'The rise of the personal data quality principle: is it legal and does it have an impact on the right to rectification?' (2021) Vol 12 No 3 European Journal of Law and Technology 21.
[1092] Mireille Hildebrandt, 'Primitives of legal protection in the era of data-driven platforms' (2018) 15 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3140594> accessed 8 February 2024.
[1093] Omer Tene, Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11 Northwestern Journal of Technology and Intellectual Property 239, 270.
[1094] Viktor Mayer-Schönberger Viktor, Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Houghton Mifflin Harcourt 2013) 42, 48, 49.
[1095] Brent Daniel Mittelstadt et al, 'The ethics of algorithms: Mapping the debate' (2016) Vol 3 Iss 2 Big Data & Society 1, 5; Solon Barocas, 'Data Mining and the Discourse on Discrimination' (2014) <https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf> accessed 8 February 2024.
[1096] Sandra Wachter, Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) Issue 2 Columbia Business Law Review 494, 510.
[1097] John Dewey, *The Middle Works of John Dewey, Volume 9, 1899-1924* (Carbondale Southern Illinois University Press 1980) 165.
[1098] Jan Biermann, John Horton, Johannes Walter, 'Algorithmic Advice as a Credence Good' (2022) Centre for European Economic Research Discussion Paper No 22-071 at 2, 14 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4326911 > accessed 8 February 2024.

operator denied to extend a mobile contract subscription with a rather low monthly fee of 10 €.[1099] This score value was used as a fact, although it was merely a prediction about future behaviour that had not yet materialised and may never do. Inferences generated by machines are highly scalable and less likely to be correct due to current reasoning deficiencies in AI (see also Sections 2.2.5, 4.3.1, 4.4.1 and 4.7.1).

Additionally, overfitting negatively affects the accuracy of predictions generated by ML and DL models. Overfitting is a common side effect of training in ML and occurs when models reach a higher accuracy on the training data than for new input data. It is inherent to training ANNs.[1100] With overfitting, the model learns how to represent well the training data, but performs poorly on new information as input.[1101] In fact, several factors determine a model's ability to generalise well, namely, the model architecture, regularisation techniques and the dataset design.[1102] Overfitting may proactively be addressed by means of lowering the number of weights an ANN has.[1103] To tune the parameters of a given model in a way that they perform well not only on training data but also on new information is a general problem in ML. Regularisation techniques are a vital tool to prevent overfitting and aim to reduce errors in predicting data that do not form part of the training set. Regularisation algorithms for ANNs may be divided into three main categories: i) data augmentation algorithms changing the input of the ANN, ii) internal algorithms changing values and inner structures of the ANN and iii) label algorithms performing their changes over the desired output.[1104] However, overfitting remains a problem despite the technical means to mitigate it. The problem of avoiding overfitting is subject to ongoing research, with regard to ANNs in particular. Overfitting mysteries in ANNs are not yet fully understood, partly due to the 'black-box' characteristics of ANNs.[1105] In any case, because overfitting occurs during the training process of an ANN, it results in high accuracy in terms of training data, but a poor prediction performance with regard to new input.[1106] Therefore, the common problem of

[1099] Case C-203/22 *Dun & Bradstreet Austria* p 2 <https://www.ris.bka.gv.at/Dokumente/Lvwg/LVWGT_WI_20220211_VGW_101_042_791_2020_44_00/LVWGT_WI_20220211_VGW_101_042_791_2020_44_00.pdf> accessed 8 February 2024.
[1100] Nicholas Carlin, 'Evaluating and Testing Unintended Memorization in Neural Networks' (*Berkeley Artificial Intelligence Research Blog*, 13 August 2019) <https://bair.berkeley.edu/blog/2019/08/13/memorization/> accessed 8 February 2024.
[1101] Claudio Filipi Gonçalves dos Santos, João Paulo Papa 'Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks' (2022) Vol 54 No Iss 10s ACM Computing Surveys 1-25 < https://dl.acm.org/doi/pdf/10.1145/3510413 > accessed 8 February 2024.
[1102] Chenlei Fang et al, 'The Overfitting Iceberg' (Machine Learning Carnegie Mellon University 31 August 2020) < https://blog.ml.cmu.edu/2020/08/31/4-overfitting/ > accessed 8 February 2024.
[1103] Joren Verspeurt, 'Applying the GDPR to AI – a practitioner's perspective on some of the main challenges' (AI Summer School Blog KU Leuven 10 January 2023) < https://www.law.kuleuven.be/ai-summer-school/AI-GDPR > accessed 8 February 2024.
[1104] Claudio Filipi Gonçalves dos Santos, João Paulo Papa 'Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks' (2022) Vol 54 No Iss 10s ACM Computing Surveys at 3, 5 < https://dl.acm.org/doi/pdf/10.1145/3510413 > accessed 8 February 2024.
[1105] Chenlei Fang et al, 'The Overfitting Iceberg' (Machine Learning Carnegie Mellon University 31 August 2020) < https://blog.ml.cmu.edu/2020/08/31/4-overfitting/ > accessed 8 February 2024.
[1106] Jianchun Chu et al, 'A novel method overcoming overfitting of artificial neural network for accurate prediction: Application on thermophysical property of natural gas'(2021) Vol 28 Case Studies in Thermal Engineering 1-13 < https://www.sciencedirect.com/science/article/pii/S2214157X21005694 > accessed 8 February 2024.

overfitting is likely to negatively affect the accuracy of predictions and is therefore detrimental to the accuracy principle.

ML produces probable, yet inevitably uncertain knowledge and may identify significant correlations. However, these correlations are rarely sufficient to posit the existence of a causal connection and to motivate action based on such uncertain knowledge[1107] (e.g., to grant or not to grant a loan). In other words, probabilistic data does not merit to be considered and treated as facts. Thus, output generated by ML can violate the accuracy principle because it is probabilistic, uncertain and likely inaccurate.[1108] This is amplified by the phenomenon called overfitting and it does not play a role whether 'absolute accuracy' or 'relative accuracy' is considered. Other aspects of ML, such as the risk of biased training data, could lead to inaccurate or wrong representations of data subjects. Output generated by biased training data typically violates the accuracy principle.[1109] Thus, ML can violate the accuracy principle, which constitutes a Type 1 legal problem. When controllers cannot prove the accuracy of the personal data processed, they simultaneously violate the accountability principle. It follows from the accountability principle itself and CJEU case law that the burden of proof concerning compliance with the principles enshrined in Article 5 (1) GDPR lies with the controller.[1110] However, in the case of output generated by means of ML, controllers are unable to prove the accuracy of the personal data processed. This violates the accountability principle.

The inaccuracy problem (Type 1)

As indicated in Section 4.3.1, ML generates output that constitutes uncertain knowledge because it is probabilistic. Overfitting amplifies this problem. Therefore, such output is likely to be inaccurate and can violate the accuracy principle. When controllers cannot prove the accuracy of such personal data, they simultaneously violate the accountability principle.

Affective computing (AC) and the underlying processing of emotion data is in direct contrast with the accuracy principle. Different studies have rebutted the idea that a person's emotional state can be accurately inferred from his facial movements[1111] as suggested by automated face analysis (AFA) systems that deploy AC approaches (see Section 2.2.4.1) to detect emotional states. Research suggests that facial movements are not diagnostic displays that reliably and specifically signal particular emotional states regardless of context, person and culture. It is not possible to confidently infer happiness

[1107] Brent Daniel Mittelstadt et al, 'The ethics of algorithms: Mapping the debate' (2016) Vol 3 Iss 2 Big Data & Society 1, 4.
[1108] Lance Eliot, 'AI Ethics And The Quagmire Of Whether You Have A Legal Right To Know Of AI Inferences About You, Including Those Via AI-Based Self-Driving Cars' *Forbes* (New York, 25 May 2022) < https://www-forbes-com.cdn.ampproject.org/c/s/www.forbes.com/sites/lanceeliot/2022/05/25/ai-ethics-and-the-quagmire-of-whether-you-have-a-legal-right-to-know-of-ai-inferences-about-you-including-those-via-ai-based-self-driving-cars/amp/> accessed 8 February 2024.
[1109] Philipp Hacker, 'Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law' (2018) Vol 55 Iss 4 Common Market Law Review 1143, 1172.
[1110] Case C-175/20 *'SS' SIA* [2022] ECR I-124 paras 77, 81.
[1111] Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 46.

from a smile, anger from a scowl or sadness from a frown because these emotion categories are more variable in their facial expressions.[1112] Another study revealed that the accuracy levels of eight commercial automatic classifiers used for facial affect recognition were consistently lower when applied to spontaneous affective behaviours compared to 'posed' affective behaviours. Validation accuracy rates of the tested classifiers varied from 48% to 62%.[1113] When absolute accuracy[1114] is considered, it is obvious that such accuracy rates do not meet this level of accuracy. The same holds true about relative accuracy when AC is applied in the context of hiring procedures. The level of accuracy required for relative accuracy depends on the purpose for processing.[1115] Processing of emotion data for the purpose of recruitment[1116] by means of AC demands a particularly high level of accuracy due to the possible impact on the data subject concerned. Thus, it can be said that the accuracy for such processing essentially must reflect reality and thus ultimately achieve absolute accuracy.

In addition, other means to detect emotions, for example based on speech (see Section 2.2.4.2) and physiological data (see Section 2.2.4.3), have been called into question due to a lack of scientific consensus whether such methods can ensure accurate or even valid results.[1117] While humans can efficiently recognise emotional aspects of speech, it is still an ongoing subject of research to automate this. Research in this context has been restricted to laboratory conditions with full-bandwidth, uncompressed and noise-free audio recordings. However, recent studies indicate that speech compression, filtering, band reduction and the addition of noise reduce accuracy significantly.[1118] Despite this, speech emotion recognition (SER) is already being applied 'in the wild'. Real-world applications of AC aiming to derive emotional states from speech are Amazon's wearable 'Halo', which analyses voice tones to detect user emotions[1119] or Spotify's patented voice assistant[1120] which, based on

[1112] Lisa Feldman Barrett et al. 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 46.
[1113] Damian Dupré et al, 'A performance comparison of eight commercially available automatic classifiers for facial affect recognition' (2020) 15 (4) PLoS ONE 1, 10 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231968> accessed 8 February 2024.
[1114] Art 29 Working Party, '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' (WP 225, 26 November 2014), at 15 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=667236> accessed 8 February 2024; Paul Voigt, Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer 2017) 91.
[1115] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.
[1116] For instance, HireVue; see Nathan Mondragon, Clemens Aicholzer, Kiki Leutner, 'The Next Generation of Assessments' (HireVue 2019) <http://hrlens.org/wp-content/uploads/2019/11/The-Next-Generation-of-Assessments-HireVue-White-Paper.pdf> accessed 8 February 2024.
[1117] Kate Crawford et al, 'AI Now Report' (2019) AI Now Institute 12 <https://ainowinstitute.org/publication/ai-now-2019-report-2> accessed 8 February 2024.
[1118] Margaret Lech et al, 'Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Computing' (2020) Vol 2 Frontiers in Computer Science 1, 3 <https://www.frontiersin.org/articles/10.3389/fcomp.2020.00014/full> accessed 8 February 2024.
[1119] Alex Hern, 'Amazon's Halo wristband: the fitness tracker that listens to your mood' *The Guardian* (London, 28 August 2020) < https://www.theguardian.com/technology/2020/aug/28/amazons-halo-wristband-the-fitness-tracker-that-listens-to-your-mood > accessed 8 February 2024; Austin Carr, 'Amazon's New Wearable Will Know If I'm Angry. Is That Weird?' *Bloomberg* (New York, 31 August 2020) < https://www.bloomberg.com/news/newsletters/2020-08-31/amazon-s-halo-wearable-can-read-emotions-is-that-too-weird > accessed 8 February 2024.
[1120] Daniel Bromand et al, 'Systems and Methods for Enhancing Responsiveness to Utterances Having Detectable Emotion' US Patent Number US 10566010 B2 (Assignee: Spotify AB) February 2020 11 < https://patentimages.storage.googleapis.com/2a/9d/2d/926b58a2bd956f/US10566010.pdf >, accessed 8 February 2024.

commands or other utterances (e.g., 'ugh'), recognises when a user sounds sad and then offers encouragement by 'cheering' the user up.[1121] Emotions are inferred from speech recorded or streamed in daily life environments, sometimes with significantly low accuracy rates. The Hungarian supervisory authority sanctioned a bank for unlawfully processing personal data (voice recordings) based on an SER-powered AI system which promised emotion detection and measurement for customers who called the bank, resulting from voice recordings.[1122] The AI Now Institute at New York University stated AC to be based on 'debunked pseudoscience'[1123] and recommended that 'regulators should ban the use of affect recognition in important decisions that impact people's lives and access to opportunities'.[1124]

In conclusion, it is obvious that processing personal data by AC described in this section violates the accuracy principle, which constitutes a Type 1 legal problem. This holds true when absolute accuracy[1125] is considered, but arguably also in the case of relative accuracy, which is more nuanced and depends on the purpose of processing. I take the view that validation accuracy rates between 48% and 62%[1126] are not acceptable even if the purpose of processing is not particularly sensitive for the data subject concerned. Admittedly, emotions dected by humans can also be inaccurate. However, AI systems function on a much larger scale, and could therefore cause more harm. Because controllers cannot prove the accuracy of the personal data processed, they simultaneously violate the accountability principle. It follows from the accountability principle itself as well as CJEU case law that the burden of proof regarding compliance with principles enshrined in Article 5 (1) GDPR lies with the controller.[1127] However, in the case of output generated by means of AC, controllers are unable to prove the accuracy of the personal data processed. This also violates the accountability principle.

[1121] Josh Mandell, 'Spotify Patents A Voice Assistant That Can Read Your Emotions' *Forbes* (New York, 12 March 2020) < https://www.forbes.com/sites/joshmandell/2020/03/12/spotify-patents-a-voice-assistant--that-can-read-your-emotions/ > accessed 8 February 2024.
[1122] Sebastião Barros Vale, Gabriela Zanfir-Fortuna, 'Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities' (2022) 48 <https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf> accessed 8 February 2024.
[1123] Kate Crawford et al, 'AI Now Report' (2018) AI Now Institute 8 <https://ainowinstitute.org/publication/ai-now-2018-report-2> accessed 8 February 2024.
[1124] Kate Crawford et al, 'AI Now Report' (2019) AI Now Institute 6 <https://ainowinstitute.org/publication/ai-now-2019-report-2> accessed 8 February 2024.
[1125] Art 29 Working Party, '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' (WP 225, 26 November 2014), at 15 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=667236> accessed 8 February 2024; Paul Voigt, Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide* (Springer 2017) 91.
[1126] Damian Dupré et al, 'A performance comparison of eight commercially available automatic classifiers for facial affect recognition' (2020) 15 (4) PLoS ONE 1, 10 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231968> accessed 8 February 2024.
[1127] Case C-175/20 *'SS' SIA* [2022] ECR I-124 paras 77, 81.

The rebuttal problem (Type 1)

Research rebutted that a person's emotional state can accurately be inferred from facial movements as suggested by AFA systems powered by AC. There is also a lack of scientific consensus whether other methods used in AC generate accurate results. Output generated by AC systems is likely inaccurate and violates the accuracy principle and simultaneously the accountability principle as controllers cannot prove the accuracy of such personal data.

AI currently lacks reasoning capabilities due to deficiencies in the AI discipline of automated reasoning as outlined in Sections 2.2.5, 4.3.1, 4.4.1 and 4.7.1. Common sense reasoning constitutes a central part of human behaviour and is a precondition for human intelligence. However, common sense reasoning capabilities are still a challenge in AI applications[1128] and AI has been called 'devoid of common sense'.[1129] Apparently, there is not one AI system today which has a semblance of common sense or has capabilities such as human cognition or can think in a manner on par with human thinking.[1130] The lack of progress in providing general automated common sense reasoning capabilities underscores that this is a very difficult problem in the field of AI.[1131] It is not just the hardest problem for AI, it is also considered to be the most important problem.[1132]

Due to these limited reasoning capabilities, AI systems may generate output that is potentially inaccurate and sometimes even discriminatory. Because AI systems lack reasoning capabilities and do not *know why* a specific input should receive some label, they only detect that the particular input correlates with a given label. For example, as outlined in Section 4.3.1, Google's AI system developed for recognising child abuse inaccurately classified a father as criminal[1133] which clearly points to the problem that AI generalises but does not distinguish. The system does not understand what it classifies as 'wrong' or 'right' and neglects the context. An AI system trained with a dataset in which only basketballs were orange is a harmless example. This system might classify all future inputs that are orange as basketballs,[1134] which obviously leads to inaccurate outcomes. Meanwhile, though,

[1128] Shoham Yoav et al, 'The AI Index 2018 Annual Report' (AI Index Steering Committee Stanford University 2018) 64 <https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf > accessed 8 February 2024.
[1129] Cade Metz, 'Paul Allen Wants to Teach Machines Common Sense' *The New York Times* (New York, 28 February 2018) < https://www.nytimes.com/2018/02/28/technology/paul-allen-ai-common-sense.html > accessed 8 February 2024.
[1130] Lance Eliot, 'AI Ethics And The Quagmire Of Whether You Have A Legal Right To Know Of AI Inferences About You, Including Those Via AI-Based Self-Driving Cars' *Forbes* (New York, 25 May 2022) < https://www-forbes-com.cdn.ampproject.org/c/s/www.forbes.com/sites/lanceeliot/2022/05/25/ai-ethics-and-the-quagmire-of-whether-you-have-a-legal-right-to-know-of-ai-inferences-about-you-including-those-via-ai-based-self-driving-cars/amp/> accessed 8 February 2024.
[1131] Brandon Bennet, Anthony G Cohn, 'Automated Common-sense Spatial Reasoning: Still a Hughe Challenge' in Stephen Muggleton, Nicholas Chater (eds) Human-Like Machine Intelligence (Oxford University Press 2021) 405.
[1132] Gary Marcus, Ernest Davis, *Rebooting AI: Buidling Artificial Intelligence we can trust* (Pantheon Books 2019).
[1133] Kashmir Hill, 'A Dad Took Photos of His Naked Toddler for the Doctor. Goolge Flagged Him as A Criminal' *The New York Times* (New York, 21 August 2022) < https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html > accessed 8 February 2024.
[1134] Zachary C Lipton, 'The Mythos of Model Interpretability' (2018) Vol 16 Iss 3 ACMQueue 3 <https://dl.acm.org/doi/pdf/10.1145/3236386.3241340?download=true> accessed 8 February 2024.

Google's photo app automatically classified images of black people as gorillas.[1135] In some circumstances, neglection of context and lack of common sense can have catastrophic consequences. The case of Molly Russel is a tragic example thereof.[1136] A recommendation system showed Molly Russel, a depressed and lonely teenage girl, 20,000 images promoting depression, suicide and self-harm – including a page of images titled 'Depression content you may like'. This system was programmed to fulfil the objectives Instagram and Pinterest gave them. It is common sense that a teenage girl looking at posts about depression does not want to be made more depressed. Ultimately, Molly Russel committed suicide.[1137] In New Zealand, a man of Asian descent had his passport application rejected because the software that approves photos claimed his eyes were closed.[1138] These examples outline that AI might generate completely inaccurate output and sometimes also discriminatory and defamatory outputs. Therefore, AI reasoning deficiencies are prone to violate the accuracy principle, regardless of whether 'absolute accuracy' or 'relative accuracy' is considered. This leads to a Type 1 legal problem.

The problem of common sense (Type 1)

AI systems can generate inaccurate data due to the reasoning deficiencies in the AI discipline of automated reasoning. AI is devoid of common sense, which may lead to completely inaccurate output. Also, controllers cannot prove the accuracy of such personal data. This violates the accuracy and accountability principles.

4.7.2 Legal problems: Type 2

The accuracy principle does not outline specific levels of accuracy that personal data processed in the context of AI must reach, and there is also no one-size-fits all approach[1139] considering that the level of accuracy depends on the purpose of processing when interpreted as relative accuracy as suggested by relevant case law.[1140] In addition, regulators so far neglected the accuracy principle by not providing substantive guidance on the matter.

When looking for more specific approaches that are helpful to interpret the accuracy principle in the context of AI, it is not possible to simply refer to the concept of accuracy or information quality in

[1135] Crawford Kate, 'Artificial Intelligence's White Guy Problem' *The New York Times* (New York, 25 June 2016) < https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> accessed 8 February 2024.
[1136] Angus Crawford, Bethan Bell, 'Molly Russell inquest: Father makes social media plea' BBC (London, 30 September 2022) <https://www.bbc.com/news/uk-england-london-63073489> accessed 8 February 2024.
[1137] Matija Franklin et al, 'The EU's AI Act needs to address critical manipulation methods' *The OECD.AI Policy Observatory* (Paris, 21 March 2023) <https://oecd.ai/en/wonk/ai-act-manipulation-methods?utm_source=substack&utm_medium=email> accessed 8 February 2024.
[1138] Titcomb James, 'Robot passport checker reject Asian man's photo for having his eyes closed' The Telegraph (London, 7 December 2016) < https://www.telegraph.co.uk/technology/2016/12/07/robot-passport-checker-rejects-asian-mans-photo-having-eyes/> accessed 8 February 2024.
[1139] Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 4.
[1140] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.

the field of computer science.[1141] The latter goes far beyond the principle of accuracy as enshrined in EU data protection law.[1142] Information quality in computer science is a multidimensional concept[1143] covering at least four dimensions, namely, intrinsic, contextual, representational and accessibility information quality. What exactly falls under the scope of these four dimensions seems to vary from the perspectives of academics and practitioners[1144] and further clarification and formalisation of these dimensions is required.[1145] Nevertheless, intrinsic information quality is particularly interesting in the context of the accuracy principle[1146] because accuracy is often considered an intrinsic information quality dimension.[1147] Literature discussing the intrinsic information quality dimension explicitly refers to the terms accuracy and correctness.[1148]

In computer science,[1149] definitions of accuracy vary. Accuracy has been defined as 'the closeness between a value v and a value v′, considered the correct representation of the real-life phenomenon that v aims to represent'.[1150] Another definition states that accuracy 'measures the degree of correctness of a given collection of data'.[1151] Furthermore, two distinct kinds of accuracy exist: syntactic and semantic accuracy. The former is defined as the closeness of a value v to the elements of the corresponding definition domain D and is measured by means of comparison functions.[1152] It is expressed by means of a *numeric* value called edit distance. Take, for example, the incorrect value 'computer viion' that is included in a database that describes the AI disciplines. The edit distance between 'computer viion' and the correct term 'computer vision' is equal to one because it corresponds to the

[1141] Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 4.
[1142] Diana Dimitrova, 'The rise of the personal data quality principle: is it legal and does it have an impact on the right to rectification?' (2021) Vol 12 No 3 European Journal of Law and Technology 9-10; Luciano Floridi, Phyllis Illari, 'Information Quality, Data and Philosophy' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 6.
[1143] Luciano Floridi, Phyllis Illari, 'Information Quality, Data and Philosophy' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 6; Leo Pipino et al, 'Developing Measurement Scales for Data Quality Dimensions' in Richard Y Wang et al (eds) *Information Quality* (Routledge 2005 1 edn) 37.
[1144] Yang W Lee et al, 'AIMQ: a methodology for information quality assessment' (2002) Vol 40 Iss 2 Information & Management 133, 134, 136; Luciano Floridi, Phyllis Illari, 'Information Quality, Data and Philosophy' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 6.
[1145] Carlo Batini, Matteo Palmonari, Gianluigi Viscusi, 'Opening the Closed World: A Survey of Information Quality Research in the Wild' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 44.
[1146] Also, contextual information quality is at least partially relevant for the accuracy principle as it often refers to the term 'completeness'. However, it also contains other less relevant aspects such as timeliness; see also Yang W Lee et al, 'AIMQ: a methodology for information quality assessment' (2002) Vol 40 Iss 2 Information & Management 133, 134, 136.
[1147] Carlo Batini, Matteo Palmonari, Gianluigi Viscusi, 'Opening the Closed World: A Survey of Information Quality Research in the Wild' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 60.
[1148] Yang W Lee et al, 'AIMQ: a methodology for information quality assessment' (2002) Vol 40 Iss 2 Information & Management 133, 134, 136; Carlo Batini, Monica Scannapieco, *Data Quality* (Springer 2006) 20-27; Yang W Lee et al, 'AIMQ: a methodology for information quality assessment' (2002) Vol 40 Iss 2 Information & Management 133, 134, 136; Luciano Floridi, Phyllis Illari, 'Information Quality, Data and Philosophy' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 7.
[1149] In the domain of Information Quality.
[1150] Carlo Batini, Monica Scannapieco, *Data Quality* (Springer 2006) 20.
[1151] Thomas C Redman, 'Measuring Data Accuracy: A Framework and Review' in Richard Y Wang et al (eds) *Information Quality* (Routledge 2005 1 edn) 24.
[1152] Carlo Batini, Monica Scannapieco, *Data Quality* (Springer 2006) 20.

insertion of the letter ‘s’ in the value ‘computer viion’. Therefore, the syntactic accuracy is 1.[1153] Semantic accuracy is more difficult to measure.[1154] Semantic accuracy coincides with the concept correctness and is measured with yes/no or correct/incorrect. For measuring the semantic accuracy of a certain value v, the true corresponding value must be known, or it must at least be possible with additional knowledge to deduce whether the value v is or is not the true value.[1155] Semantic accuracy seems to be quite similar to absolute accuracy in the legal sense as is measured with ‘correct/incorrect’. Syntactic accuracy is more nuanced and allows for development of more flexible approaches, for example, by means of defining accuracy ranges that are considered still accurate (e.g., syntactic accuracies between 1 and 10 are considered accurate enough) which could prove to be helpful regarding relative accuracy.

In addition, the interpretation of the term ‘completeness’ varies in computer science and might relate to absolute or relative accuracy in the legal sense. For example, completeness is described as ‘the extent to which data are of sufficient breadth, depth, and scope for the *task at hand*’[1156] which seems to be similar to the notion of relative accuracy in the legal sense. Another interpretation of completeness in computer science seems to be comparable to absolute accuracy in the legal sense. A data unit consisting of one or more components (such as number, file, record), is complete if each data item constituting the data unit has been assigned a value in accordance with the data definition for the data item. If the latter is not fulfilled, the data unit is incomplete.[1157]

The concepts of accuracy and completeness in computer science will not be the ultimate solution to applying the accuracy principle. With semantic accuracy, the problem is that the correct value might *not* be known, for example, in the case of predictions or inferences produced by ML which are solely probabilistic (see Section 4.7.1). Syntactic accuracy might be too imprecise because it only allows one to calculate the closeness of a value but does not indicate that a value is inaccurate or incorrect. More generally, there is no single way to measure the accuracy of the data under all circumstances. Measuring the accuracy of the data is particularly difficult due to the nature of data. Determining data accuracy must necessarily make reference to human knowledge, other data or the real world.[1158] Another issue with respect to accuracy in computer science is a phenomenon called concept drift: Even if an AI system might initially be accurate, accuracy might change over time when it is applied in

[1153] Carlo Batini, Monica Scannapieco, *Data Quality* (Springer 2006) 21.
[1154] Carlo Batini, Matteo Palmonari, Gianluigi Viscusi, ‘Opening the Closed World: A Survey of Information Quality Research in the Wild’ in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 54.
[1155] Carlo Batini, Monica Scannapieco, *Data Quality* (Springer 2006) 20.
[1156] Richard Y Wang, Diane M Strong, ‘Beyond Accuracy: What Data Quality Means to Data Consumers’ (1996) Vol 12 No 4 Journal of Management Information Systems, 5, 32 (emphasis added).
[1157] Leo Pipino et al, ‘Developing Measurement Scales for Data Quality Dimensions’ in Richard Y Wang et al (eds) *Information Quality* (Routledge 2005 1 edn) 44.
[1158] Thomas C Redman, ‘Measuring Data Accuracy: A Framework and Review’ in Richard Y Wang et al (eds) *Information Quality* (Routledge 2005 1 edn) 23.

practice or 'real world', in particular when the behaviour of individuals that the system seeks to evaluate changes. In this case, an AI system is likely to inaccurately evaluate these individuals.[1159] Validation accuracy which tests ML models on data unseen during training to estimate how well the model is expected to perform later in real life seems to be an interesting instrument for applying the accuracy principle in practice,[1160] particularly regarding relative accuracy. Validation accuracy rates (e.g., 80, 90 or 100%) could be helpful when applying the accuracy principle in practice because the degree of accuracy to be achieved always depends on the purpose of processing.[1161]

There has been no exchange of ideas between computer science and law on the matter of information quality and accuracy.[1162] Corresponding interdisciplinary research is a relatively recent development.[1163] This is unfortunate because such interdisciplinary research could be helpful when applying the accuracy principle to AI. Nevertheless, within this section I have outlined that the concepts of information quality, accuracy, completeness and validation accuracy from research in the field of computer science might be helpful to interpret the accuracy principle in the context of AI. More interdisciplinary research is needed to develop an interpretation of the accuracy principle which is valid and practical both from a legal and computational perspective.

Consequently, when assessing the accuracy of personal data generated by means of AI, the model upon which inferred personal data are based also must be considered to ensure a comprehensive assessment. The quality of such information, i.e. the personal data generated by means of AI, is affected by the quality of the AI system used to generate it.[1164] Regulators so far neglected the accuracy principle by not providing substantive and practice-oriented guidance on the matter, which reduces legal certainty. This makes it difficult if not impossible to enforce the accuracy principle in the context of AI, both in regulatory enforcement (by SAs)[1165] and in private enforcement pursued by data subjects and their representatives. This leads to a Type 2 legal problem and is caused by the accuracy principle itself and may arise *regardless* of which AI discipline it is applied to. Nonetheless, this problem is most apparent regarding predictions, inferences and other probabilistic output generated by means of ML and AC (see also Section 4.3.1).

[1159] Diana Dimitrova, 'The rise of the personal data quality principle: is it legal and does it have an impact on the right to rectification?' (2021) Vol 12 No 3 European Journal of Law and Technology 21.
[1160] Ethem Alpaydin, *Machine Learning: The New AI* (3rd edn MIT Press 2016) 181.
[1161] Case C-434/16, *Nowak* [2017] ECR I-994 para 53.
[1162] Dara Hallinan, Frederik Zuiderveen Borgesius, 'Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle' (2020) Vol 10 No 1 IDPL 1, 4.
[1163] Burkhard Schäfer, 'Information Quality and Evidence Law: A New Role for Social Media, Digital Publishing and Copyright Law?' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 217.
[1164] Burkhard Schäfer, 'Information Quality and Evidence Law: A New Role for Social Media, Digital Publishing and Copyright Law?' in Luciano Floridi, Phyllis Illari (eds) *The Philosophy of Information Quality* (Springer Nature 2014) 217.
[1165] See Articles 51 to 58 GDPR.

The guidance problem (Type 2)
The lack of guidance concerning the accuracy principle and the absence of interdisciplinary research in the fields of computer science and law leads to legal uncertainty and makes it difficult if not impossible to enforce in the context of AI.

4.7.3 Legal problems: Type 3

The guidance problem explained in Section 4.7.2 automatically leads to a Type 3 legal problem. The accuracy principle is not fit for purpose to effectively protect[1166] data subjects from being inaccurately represented in the form of output generated by AI. In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect the data subject's personal data against risk of misuse.[1167] A principle that lacks substantive detail cannot prevent misuse in the form of inaccurate representations of data subjects. Likewise, it cannot ensure a high level of data protection.[1168] Due to the accuracy principle's lack of detail caused by absent guidance and respective interdisciplinary research, it fails to achieve the GDPR's aim to establish a strong and coherent data protection framework[1169] when considering that principles provide the basis for the protection of personal data[1170] in the GDPR.

The fairness principle as well as the accuracy principle as discussed in Sections 4.3.2 and 4.7.2 respectively have in common that they lack sufficient guidance when applied to AI. This leads to legal uncertainty and ultimately to a Type 3 legal problem. The lack of regulatory guidance and the absence of interdisciplinary research make these principles 'incomputable'. As it is the case with the purpose limitation and data minimisation principles, [1171] measurable definitions of the accuracy and fairness principles and concrete indications on how to practically implement them are needed to make them 'computable'. To replicate and apply legal reasoning, AI requires the translation of the linguistic categories used by law into mathematical functions. This is not a straightforward task, because there is an element of flexibility and contestability in natural language used to express juridical forms that cannot be completely captured by mathematical algorithms.[1172] Whereas this points more generally to

[1166] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[1167] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1168] Recitals 6, 10 as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[1169] Recital 7 GDPR.
[1170] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.
[1171] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 58 and 61 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1172] Christopher Markou, Simon Deakin, 'Ex Machina Lex: Exploring the Limits of Legal Computability' in Simon Deakin, Christopher Markou (eds) *Is Law Computable?: Critical Perspectives on Law and Artificial Intelligence* (Hart Publishing 2020) 66.

the limits of legal computability, it holds especially true in case of principles, which are by nature less concrete and provide a great deal of flexibility when applied in practice. This is even more true where the substantive meaning of principles remains largely unclear, as is the case with the fairness and accuracy principle.

Computability of principles is an essential requirement to develop AI systems which implement data protection principles and thus comply with the concept of data protection by design and default according to Article 25 GDPR. Although the latter, as introduced in Section 3.3.3.9, does not appear under the principles for processing named in Article 5 of the GDPR, it is closely intertwined with them. It obliges controllers to put in place, both at the design *and* processing stage,[1173] technical and organisational measures 'that are designed to implement data protection principles.'[1174]

As pointed out in the elusiveness problem discussed in Section 4.3.2, little has been written what 'fair processing' really means[1175] and on the application of the fairness principle in practice.[1176] This renders the fairness principle incomputable. In addition, interdisciplinary research highlights that certain legally prohibited kinds of discrimination are too contextual, intuitive and open to judicial interpretation to be automated. Many of the available computational implementations of the fairness principle are thus not able to adequately reflect its legal requirements.[1177]

Uncertainties regarding the proper meaning of the accuracy principle render it incomputable, even when concepts of accuracy and information quality elaborated in the field of computer science are considered (see also Section 4.7.2). The incomputability of both the fairness and accuracy principles creates a Type 3 legal problem, both regarding the principles themselves as well as the concept of Data Protection by Design and Default ('DPbDD') according to Article 25 GDPR as introduced in Section 3.3.3.9. The computability of principles is an essential requirement to develop AI systems that implement data protection principles at both the design and processing stages. At first sight, the concept of DPbDD seems promising and relevant considering new technologies such as AI. However, this concept fails to deliver what it promises because it requires controllers to implement, by means of technical measures, data protection principles which are essentially *incomputable*. This is significant when considering that principles provide the basis for the protection of personal data in EU data protection law.[1178] Developers cannot implement these principles in the design phase and during the

[1173] Article 25, Recital 78 GDPR.
[1174] Article 25 GDPR.
[1175] Winston J Maxwell, 'Principle-based regulation of personal data: the case of fair processing' (2015) Vol 5 No 3 International Data Privacy Law 205.
[1176] Damian Clifford, Jef Ausloos 'Data Protection and the Role of Fairness' (2018) Vol 37 No 1 Yearbook of European Law 130, 184.
[1177] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 59 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1178] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.

actual use of AI systems. Thus, the DPbDD is not fit for purpose to protect the fundamental right to data protection. This constitutes a Type 3 legal problem. Incomputable principles are not fit for purpose to achieve the GDPR's aim to establish a strong and coherent data protection framework[1179] when considering that principles provide the basis for the protection of personal data[1180] in the GDPR. This Type 3 legal problem occurs regardless of which AI discipline the fairness and accuracy principles are applied to because the incomputability of these two principles causes the legal problem. Therefore, it is a general problem and relates to all AI disciplines as introduced in Chapter 2. To be clear, I do not suggest principle-based processing in AI systems is impossible as it cannot be computed abstractly. Instead, the incomputability is caused by the need for more guidance and more interdisciplinary research. Mathematical interpretations of principles are needed to render them computable.[1181]

The incomputability problem (Type 3)

The lack of guidance concerning the fairness and accuracy principle renders them incomputable. Developers cannot encode these principles in the design phase and during the actual use of AI systems as required by the concept of data protection by design and default which obliges controllers to implement the data protection principles by technical means. Incomputable principles are not fit for purpose to protect the fundamental right to data protection.

As outlined in Section 4.7.2, the accuracy principle is difficult to enforce in practice due to the absence of specific levels of accuracy that could be considered when assessing the accuracy of personal data in the context of AI. This is particularly problematic when considering that the accuracy principle is closely intertwined with the right to rectify personal data according to Article 16 GDPR.[1182] The AI disciplines ML and AC provide new means to generate inferences, predictions and other output. In Section 4.7.1 I have outlined that such outputs can be inaccurate. The lack of guidance regarding the accuracy principle makes it difficult for data subjects to enforce their right to rectification. I discuss this problem in Section 5.7.

4.8 Enhanced protection for 'special data'

The notion of special categories of personal data is broadly interpreted by the CJEU. It ruled that personal data *indirectly* revealing special categories of personal data defined in Article 9 (1) GDPR is also covered by the latter provision.[1183] In this ruling, the CJEU followed AG Pikamäe's opinion by stating that 'the verb "reveal" is consistent with taking into account processing of inherently

[1179] Recital 7 GDPR.
[1180] Cécile de Terwangne, Commentary of Article 5 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 313.
[1181] Asia Biega, Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44, 58 and 61 <https://arxiv.org/abs/2101.06203> accessed 8 February 2024.
[1182] See Recitals 6 and 10 GDPR, as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66.
[1183] Case C-184/20, *OT* [2022] ECR I-601, paras 117-128.

sensitive data, as well as data revealing information of that nature *indirectly*, following an intellectual operation involving deduction or cross-referencing'.[1184] In the context of AI, this ruling is quite important because ML might generate personal data that indirectly reveal special categories of personal data. ML models that apply dimensionality reduction (see Section 2.2.1.2) on easily accessible digital records of behaviour, for example, Facebook likes, may reveal and predict highly sensitive personal attributes such as sexual orientation, ethnicity, religious and political views and personality traits.[1185] It is now clear that the processing of such data falls under the scope of Article 9 GDPR. However, the broad interpretation of special categories of personal data does not solve all the legal problems that might arise due to AI. This is mainly due to the principle[1186] of enhancing protection for special data and the legislator's approach to enumerate such data exhaustively. In Section 4.8.3, I outline that this approach has significant consequences considering the technological developments facilitated by AI. Both GDPR and its predecessor use the term 'special categories' of personal data, but also refer to 'sensitive personal data' in the recitals. [1187] In order to avoid confusion, I will use the term 'special data' to refer to data that *are* in fact, protected under the GDPR and 'sensitive data' to refer to data that are, in fact, *not protected* under the GDPR (although they arguably should be).

As outlined in Section 3.3.1.2, the rationale for ensuring enhanced protection for special data stems from their particular sensitive nature (Recital 51 GDPR). Processing of special data can constitute a particularly serious interference with the fundamental rights to privacy and data protection.[1188] In view of the SAs, it is needed to specifically protect special data because misuse of such data may have more severe consequences for the data subjects than misuse of 'regular' personal data.[1189] This is underscored by Recital 51 GDPR, which states that 'processing [of sensitive personal data] could create significant risks to fundamental rights and freedoms'. Nevertheless, the principle[1190] of enhancing protection for special categories of personal data is not undisputed.[1191] This will be discussed in Section 6.3.

[1184] Case C-184/20, *OT* [2022] ECR I-601, paras 123, emphasis added; Case C-184/20, *OT* [2022] ECR I-601, Opinion of AG Pikamäe, para 85.
[1185] Michal Kosinski, David Stillwell, Thore Graepel, 'Private traits and attributes are predictable from digital records of human behaviour' (2013) Vol 110 No 15 PNAS, 5802.
[1186] For the purpose of this thesis, I regard this choice as a principle so that it neatly matches the approach taken, distinguishing between principles and rights.
[1187] See Recitals 10, 51 GDPR, Recitals 34 and 70 Data Protection Directive which refer to sensitive but not 'special' categories of personal data
[1188] Case C-184/20, *OT* [2022] ECR I-601, para 126; Case C-136/17, GC and Others [2019] ECR I-773 para 44; Recital 51 GDPR.
[1189] Art 29 Working Party, 'Advice paper on special categories of data ("sensitive data")' (20 April 2011) at 4.
[1190] Admittedly, this is not a traditional data protection principle. Nonetheless, it could be regarded as a principle in a broader sense, which then also aligns with the approach taken in this chapter.
[1191] Ludmila Georgieva, Christopher Kuner Commentary of Article 9 in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds), *The EU General Data Protection Regulation: A Commentary* (OUP 2020) 370; Lee A. Bygrave, *Data Privacy Law: An International Perspective* (OUP 2014) 165; Lokke Moerel, Corien Prins, 'Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things' (2016) p 11 and 56 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2784123> accessed 8 February 2024.

4.8.1 Legal problems: Type 1

As will be outlined in Section 4.8.3, the problem with respect to the approach to exhaustively enumerate special data arises because AI provides unprecedented means of generating and otherwise processing new types or categories of sensitive personal data. The exhaustive list of sensitive data contained in the GDPR does not keep up with technological developments facilitated by AI. This means that the strict rules concerning the processing of sensitive data do not apply to new types of sensitive personal data facilitated by AI. Nonapplicable or nonexisting provisions cannot be violated, and therefore no specific Type 1 legal problems arise.

4.8.2 Legal problems: Type 2

As outlined in Section 4.8.1, provisions that are not applicable or do not yet exist cannot be violated. Consequently, no specific Type 2 legal problems arise.

4.8.3 Legal problems: Type 3

AI provides an unprecedented means to generate and otherwise process arguably new types or categories of sensitive personal data. This causes legal problems regarding the principle of enhancing protection for special data and the legislator's approach to define special data exhaustively. Definitions contained in the current legal framework do not keep up with technological developments facilitated by AI. I demonstrate this issue by discussing emotion data, location data, neurodata and mental data, respectively.

Emotion data

By means of AC, machines may gain access to the emotional life of individuals, information that is highly personal, intimate and private.[1192] In fact, all emotions are by definition personal[1193] and revealing them makes an individual more vulnerable.[1194] A commonly agreed definition of emotion in any of the disciplines that study this phenomenon does not exist.[1195] For the purpose of this thesis, I define emotion data as information relating to emotions of an individual. To avoid lengthy discussions on what emotions are, I simply refer to the six most-used emotion categories[1196] in emotion research:

[1192] Rosalind W Picard, *Affective Computing* (MIT Press 1997) 118.
[1193] Not meaning personal in the sense of personal data but more to the common understanding of the notion.
[1194] Aaron Ben-Ze'Ev, *The Subtlety of Emotions* (MIT Press 2000) 183.
[1195] Kevin Mulligan, Klaus R. Scherer, 'Toward a Working Definition of Emotion' (2012) Vol. 4 No. 4 Emotion Review 345-537.
[1196] These six emotions refer to research conducted by psychologists in the early seventies that developed the methodology of 'basic emotions'; see Paul Ekman, Wallace v Friesen, 'Constants across cultures in the face and emotion' (1971) Vol 17 (2) Journal of Personality and Social Psychology 124.

anger, disgust, fear, happiness, sadness and surprise.[1197] These six 'basic emotions'[1198] are further described in Section 2.2.4.1. It should be noted that emotion data constitutes a subcategory of mental data (see Figure 2.1). Emotions are felt as personal because they relate to a person's values[1199] and express what a person cares about.[1200] Because there is an inherent relationship between emotions and personhood[1201] and privacy is considered fundamental to the maintenance of human dignity and the boundary to one's personhood,[1202] information regarding emotions is sensitive and intimate.[1203] When emotion data constitute personal data because the data subject is identified or identifiable, the question arises whether such data are specifically protected as 'special data'.

Considering the special categories of personal data defined in Article 9 (1) GDPR and its corresponding recitals,[1204] emotion data itself is never protected as a special category of personal data under the GDPR, despite its sensitive and intimate nature.[1205]

Ultimately, the approach taken in AC determines whether processing of *personal data used to detect or derive* emotion data falls under the scope of Article 9 GDPR. A distinction can be made between single-modal affect recognition and multimodal affect recognition approaches in AC. [1206] Single-modal approaches are divided into text sentiment analysis, audio emotion recognition, visual emotion recognition focussing on facial expression and body gestures and physiological-based emotion recognition systems. [1207] Physiologically-based emotion recognition systems include AC systems that detect emotional states from EEG and ECG. ECG-based emotion recognition systems record the physiological changes of the human heart in order to detect the corresponding waveform transformation, which provides information for emotion recognition.[1208] For example, ECG-based emotion recognition systems can be applied when listening to music.[1209] EEG is a non-invasive method consisting in detection

[1197] Lisa Feldman Barrett et al 'Emotional Expressions Reconsidered' (2019) Vol 20 (1) Psychological Science in the Public Interest 1, 52.
[1198] Eiman Kanjo et al, 'Emotions in context: examining pervasive affective sensing systems, applications, and analyses' (2015) Vol 19 Personal and Ubiquitous Computing 1197, 1204 <https://link.springer.com/content/pdf/10.1007/s00779-015-0842-3.pdf> accessed 8 February 2024.
[1199] Heather C Lench, Zakari Koebel Capenter, 'What Do Emotions Do for Us?' in Heather C Lench (ed) *The Function of Emotions* (Springer 2018) 1, 142.
[1200] Giovanni Stanghellini, René Rosfort, *Emotions and Personhood: Exploring Fragility – Making Sense of Vulnerability* (OUP 2013) 142.
[1201] Ibid 149.
[1202] William S Brown, 'Technology, Workplace Privacy and Personhood' (1996) Vol 15 Journal of Business Ethics 1237, 1243.
[1203] Andrew McStay, 'Emotion AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy' (2020) Vol 7 Iss 7 Big Data & Society 1, 4.
[1204] Recitals 51, 52, 53 GDPR.
[1205] Contrary to Clifford's view that argues this 'will clearly result in the processing of sensitive personal data'; see Damian Clifford, 'Citizen-consumers in a personalised Galaxy: Emotion influenced decision-making, a true path to the dark side?' (2017) CiTiP Working Paper 31/2017, 21 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037425> accessed 8 February 2024.
[1206] Yan Wang et al, 'A systematic review on affective computing: emotion models, databases, and recent advances' (2022) Volumes 83-84 Information Fusion 19-52.
[1207] Ibid 19, 21.
[1208] Ibid 19, 35-36.
[1209] Yu-Liang Hsu et al, 'Automatic ECG-Based Emotion Recognition in Music Listening' (2020) Vol 11 No 1 IEEE Transactions on Affective Computing 85-99.

and registration of electrical activity occurring in the brain.[1210] EEG-based emotion recognition systems directly measure changes in brain activities, which provides internal features of emotional states.[1211]

Only physiologically-based emotion recognition systems in AC involve the processing of special data as defined in the GDPR. Information processed by these systems falls under the definition of health data, which covers not only physical or mental health, but also 'any information (…) on the *physiological* or biomedical state of the data subject independent of its source'.[1212] Consider, for example, AC applications that derive emotion data from physiological data such as heart rate, blood pressure and skin conductance. Research has shown that heart rate variability provides a novel marker to recognise emotions in humans.[1213] Information relating to heart rate, blood pressure and skin conductance falls under the definition of health data and is protected as a special category of personal data according to the GDPR.[1214] Automated face analysis systems (AFA) that try to detect depression from analysing an individual's facial expressions in videos arguably process (mental) health data, even if the data subject concerned is completely healthy.[1215]

Most of the single-modal affect recognition systems pursued in AC do not amount to the processing of special data. AC systems deploying approaches such as text sentiment analysis, audio emotion recognition and visual recognition of emotion focussing on facial expressions and body gestures do *not* involve the processing of special categories of personal data.[1216] Information processed within these approaches and derived emotion data are thus not protected as special personal data under the GDPR, despite their sensitive and intimate nature.[1217] This also holds true when biometric data are used for AC to detect the emotional state of the individual concerned, for example in the context of AFA systems and emotion detection based on an individual's voice and speech.[1218] Biometric data according to Article 9 (1) GDPR is only protected as special personal data if it is used for the *purpose*

[1210] Szczepan Paszkiel, *Analysis and Classification of EEG Signals for Brain–Computer Interfaces* (Springer Nature 2020) 3.
[1211] Yan Wang et al, 'A systematic review on affective computing: emotion models, databases, and recent advances' (2022) Volumes 83-84 Information Fusion 19, 35; Jianhua Zhang et al, 'Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review' (2020) Vol 59 Information Fusion 103-126.
[1212] Recital 35 GDPR (emphasis added).
[1213] Quintana Daniel et al. 'Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the automatic nervous system and social cognition' (2012) Vol 86 No 2 International Journal of Psychophysiology 168.
[1214] Article 3 (15) and 9 (1) GDPR; Recital 15 GDPR.
[1215] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 9.
[1216] Recitals 51, 52, 53 GDPR.
[1217] Contrary to Clifford's view that argues this 'will clearly result in the processing of sensitive personal data'; see Damian Clifford, 'Citizen-consumers in a personalised Galaxy: Emotion influenced decision-making, a true path to the dark side?' (2017) CiTiP Working Paper 31/2017, 21 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037425> accessed 8 February 2024.
[1218] Note that Article 29 WP considered voice as biometric data, Art 29 Working Party, 'Opinion 4/2007 on the concept of personal data' (WP 136, 20 June 2007) at 8.

of uniquely *identifying* an individual. This means 'processed through a specific technical means allowing the unique identification or authentication of a natural person'.[1219]

According to regulatory guidance adopted by EU supervisory authorities, biometric identification typically involves 'the process of comparing biometric data of an individual (acquired at the time of the identification) to a number of other biometric templates stored in a data database (i.e. a one-to-many matching process)'.[1220] For example, HumeAI[1221] provides AC-powered tools helping recruiters to assess personality traits and detect emotional states of job candidates disclosed during automated video assessments based on facial expressions. This system does not process biometric data in the form of facial expressions to uniquely identify the job candidate, as required by Article 9 (1) GDPR. Rather, it detects the emotional states the candidate portrays during the automated video assessment. Identification is achieved through other means beforehand: when the candidate reveals its name, the other identifiable information. The same applies to any other AC system aiming to detect emotional states from facial expressions,[1222] for instance those offered by the companies Realeyes[1223] or Tawny.[1224]

This also holds true when AC systems use biometric data in the form of speech, as discussed in Section 2.2.4.2 to detect the emotions of the individual concerned. Consider an AC system that advises a call centre agent to speak with more empathy because the customer seems to be angry according to the automated speech and voice analysis. Such a system does not process biometric data for identification purposes. Regulatory guidance generally considers voice to be biometric data[1225] as defined in Article 4 (14) GDPR, i.e. personal data 'resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'. However, according to Article 9 (1) GDPR, biometric data *only* qualifies as a special category of personal data if it is *used* for the *purpose* of uniquely *identifying* an individual.[1226] AG Pikamäe has termed this the 'purposive approach'.[1227] This purposive approach causes the inapplicability of Article 9 GDPR when biometric data are processed for purposes other than uniquely identifying an individual. The GDPR thus links the use of biometric data *exclusively* to the purpose of identification and therefore excludes

[1219] Recital 51 GDPR, the same recital states that processing of photographs should not systematically be considered to be processing of special categories of personal data.
[1220] Art 29 Working Party, 'Opinion 3/2012 on developments in biometric technologies' (WP 193, 27 April 2012) at 5.
[1221] See < https://hume.ai/products/facial-expression-model/ > and < https://gethume.com/blog5/artificial-intelligence-for-recruiting > accessed 26 March 2023. > accessed 8 February 2024.
[1222] Provided that identification is not based on biometric data.
[1223] See <https://www.realeyesit.com/> accessed 8 February 2024.
[1224] See <https://www.tawny.ai/product> accessed 8 February 2024.
[1225] Art 29 Working Party, 'Opinion 4/2007 on the concept of personal data' (WP 136, 20 June 2007) at 8; European Data Protection Board, 'Guidelines 02/2021 on Virtual Voice Assistants' (16 May 2011) at 31.
[1226] Article 9 (1) GDPR.
[1227] Case C-184/20 [2021] *OT* ECR I-991 Opinion of AG Pikamäe para 86.

all biometric data processed for *other purposes*,[1228] such as emotion recognition purposes. Hence, emotion data are not protected as special data under the GDPR, nor as biometric data defined in Article 4 (14) GDPR. This interpretation is also in line with the regulatory enforcement pursued by the Hungarian SA. In this case, a Hungarian bank used an AI system with the aim to detect and measure emotions of customers that called the bank's customer service.[1229] In its decision, the Hungarian SA reached the conclusion that emotion data did not constitute special data according to Article 9 (1) GDPR. Voice recordings (biometric data) were not used to identify the data subject, nor did the inferences drawn by the AI system reveal data with respect to physical or mental health.[1230]

In some cases, AC systems process special personal data to *derive or detect* emotion data. This applies to physiological-based emotion recognition systems that process information like heart rate, blood pressure and skin conductance. Such information constitutes health data, which is a special category of personal data in the GDPR. Nevertheless, the highly sensitive detected emotion data itself *never* constitutes special data under the GDPR, irrespective of which affect recognition (single-modal or multimodal) approach in AC is deployed. Thus, inherently sensitive personal data are not specifically protected in EU data protection law. This leads to a significant gap in legal protection.

The EU Commission's proposed ePrivacy Regulation[1231] as well as the compromise text[1232] used for the EU's trilogue procedure label information relating to emotions as highly sensitive. This implies that emotion data might be subject to different levels of protection depending on the applicable laws. In case both the GDPR[1233] and the future ePrivacy Regulation are triggered, emotion data will be protected as sensitive data according to the ePrivacy Regulation, but not according to the GDPR.[1234] Such a situation might be confusing and disadvantageous for data subjects, but also for companies that need to comply with the GDPR and the ePrivacy Regulation. In addition, regulating emotion data by means of different levels of protection does not seem to contribute to legal certainty.

[1228] Gloria González Fuster, Michalina Nadolna Peeters, 'Person identification, human rights and ethical principles. Rethinking biometrics in the era of artificial intelligence' (2021) 2, 20, 25 <https://www.europarl.europa.eu/RegData/etudes/STUD/2021/697191/EPRS_STU(2021)697191_EN.pdf> accessed 8 February 2024.
[1229] Sebastião Barros Vale, Gabriela Zanfir-Fortuna, 'Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities' (2022) 48 <https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf> accessed 8 February 2024.
[1230] Cesar Manso-Sayao, Summary of Hungarian SA Decision NAIH-85-3/2022 < https://gdprhub.eu/NAIH_(Hungary)_-_NAIH-85-3/2022 > accessed 8 February 2024.
[1231] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) Recital 2 <https://data.consilium.europa.eu/doc/document/ST-12633-2019-INIT/en/pdf > accessed 8 February 2024.
[1232] Council of the EU 6087/21 recital 2 <https://data.consilium.europa.eu/doc/document/ST-6087-2021-INIT/en/pdf > accessed 8 February 2024.
[1233] Namely, where emotion data must be considered personal data because the data subject is identified or identifiable.
[1234] Provided that the proposed ePrivacy Regulation will not be amended with regard to the sensitivity of emotion data.

Emotion data are inherently sensitive due to the intrinsic relationship between emotions and personhood[1235] and therefore merit specific protection as 'special data' according to the GDPR. Furthermore, the processing of emotion data could create significant risks to fundamental rights and freedoms in the sense of Recital 51 GDPR, the personal autonomy of the data subject, in particular. As outlined in Section 4.3.3, information concerning the emotional state of an individual might be particularly helpful to manipulate this individual because emotions play an important role in the elicitation of autonomous motivated behaviour[1236] and reasoning.[1237] AC provides access to emotion data of individuals and may affect people's decisions and lives in unprecedented ways. This holds particularly true regarding manipulation that operates by relying on facts about the subject's psychology such as knowledge about its emotions and desires.[1238] Emotions can have significant effects on economic transactions and play a powerful role in everyday economic choices.[1239] This affects personal autonomy, i.e. the idea 'that people should make their own lives'[1240] when facing freely both existential and every day's choices.[1241]

The fact that emotion data do not receive specific protection under the GDPR despite its highly sensitive nature and the risks relating to the data subject's personal autonomy leads to a Type 3 legal problem. The approach to exhaustively enumerate special categories of personal data creates a protection gap with regard to the processing of new kinds of sensitive personal data facilitated by AI. Therefore, this approach is not fit for purpose to effectively[1242] protect the fundamental right to data protection as it fails to specifically protect inherently sensitive data. In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect[1243] the data subject's

[1235] Giovanni Stanghellini, René Rosfort, *Emotions and Personhood: Exploring Fragility – Making Sense of Vulnerability* (OUP 2013) 149, Andrew McStay, 'Emotion AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy' (2020) Vol 7 Iss 7 Big Data & Society 1, 4; William S Brown, 'Technology, Workplace Privacy and Personhood' (1996) Vol 15 Journal of Business Ethics 1237, 1243.
[1236] Leen Vandercammen et al, 'On the Role of Specific Emotions in Autonomous and Controlled Behaviour' (2014) Vol 28 Iss 5 European Journal of Personality 413, 445.
[1237] Steffen Steinert, Orsolya Friedrich, 'Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces' (2020) Vol 26 Science and Engineering Ethics 351, 352.
[1238] J S Blumenthal-Barby, 'A Framework for Assessing the Moral Status of Manipulation' in Christian Coons, Michael Weber (eds) *Manipulation* (Oxford University Press 2014) 123, 127.
[1239] Jennifer S Lerner, Deborah A Small, George Loewenstein, 'Heart Strings and Purse Strings' (2004) Vol 15 No 5 American Psychology Society 337-340.
[1240] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1986) 369.
[1241] Daniel Susser, Beate Roessler, Helen Nissenbaum 'Technology, autonomy, and manipulation' (2019) Vol 8 Iss 2 Internet Policy Review 1, 8.
[1242] Recital 11.
[1243] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.

personal data against risk of misuse.[1244] It can neither ensure a high level of protection[1245] nor a strong and coherent data protection framework[1246] when considering the gap of protection it creates.

The emotion data problem (Type 3)

The AI discipline AC facilitates the processing of emotion data, information that is highly sensitive and intimate. Despite the sensitive nature, it is not protected as special data under the GDPR because the approach to enumerate special categories of personal data exhaustively cannot keep up with developments in AI. Consequently, this principle creates a significant gap of protection and is therefore not fit for purpose to protect the fundamental right to data protection.

Location data

Location data reveals where individuals live, work and shop, which bars and restaurants they visit, which political events they attend and which medical services they need,[1247] providing a very intimate insight into the private life of individuals.[1248] Therefore, location data are of sensitive nature.[1249] It is considered to be a valuable asset with a variety of commercial and public uses.[1250] As opposed to emotion data, mental data and neurodata, location data are not a 'new' type of personal data. Rather, when processed by means of AI, location data become personal data of a sensitive nature. Based on historical patterns, modelling applications that analyse user location data can predict where a user will be located at a particular time of the day. The prediction of a user's location is often based on ML,[1251] using techniques such as regression, clustering and ANNs as described in Section 2.2.1. Research has shown that the current location of a smartphone user can be predicted with an average of 90% accuracy by exploiting ML techniques to develop a hybrid AI system for location prediction with smartphone logs.[1252] ML and probabilistic reasoning techniques can infer daily activities of an individual from location data.[1253] Collecting, storing and analysing location data can have significant privacy implications and enables to infer a detailed picture of a person's routine, lifestyle and social

[1244] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1245] See Recitals 6 and 10 GDPR, as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44.
[1246] Recital 7 GDPR.
[1247] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1248] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 18.
[1249] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 13; Article 29 Working Party, 'Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation 2002/58/EC' (WP 247, 4 April 2017) at 30.
[1250] Andrej Savin, EU Telecommunications Law (Elgar 2018) 296.
[1251] Eran Toch et al, 'Analyzing large-scale human mobility data: a survey of machine learning methods and applications (2019) Vol 58 Knowledge and Information Systems 501, 512, 513.
[1252] Sung-Bae Cho, 'Exploiting machine learning techniques for location recognition and prediction with smartphone logs' (2016) Vol 176 Neurocomputing 98-106.
[1253] Lin Liao, 'Location-Based Activity Recognition' Dissertation University of Washington 2006.

network.[1254] Location-related information such as everyday habits, daily movements, and activities can help to establish a profile of the individuals concerned. From a privacy perspective, such profiles are no *less sensitive* than the actual content of electronic communications, according to the CJEU.[1255] Key locations such as the home or workplace of a mobile user can be inferred even from pseudonymous location data.[1256] By analysing widely available location metadata in public data streams like Twitter, such key locations can be pinpointed with a high level of accuracy, making it a trivial task to identify the individual concerned.[1257]

Despite its sensitive nature, location data are not listed in the definition of special data according to Article 9 (1) GDPR. Furthermore, the ePD does not provide protection against processing sensitive location data performed by information society providers. As outlined in Section 3.4.3.3, the processing of location data is specifically regulated by Article 9 (1) ePD and requires the consent of the user or subscriber or is allowed where location data are made anonymous when processed by electronic communications services (ECS). The latter covers access services, interpersonal communications services and services consisting wholly or mainly of the conveyance of signals[1258] and over-the-top (OTT) services such as VoIP[1259] solutions, messaging services and web-based email services which are functionally equivalent to traditional voice telephony and text message services.[1260] The strict regulation of Article 9 (1) ePD however does *not* apply where location data are processed by providers of information society services, even when such processing is performed via public electronic communication networks.[1261] Information society services are defined broadly and include any 'service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services'. [1262] Whereas the installation of an app on a mobile device itself requires consent according to Article 5 (3) ePD,[1263] the processing of location data itself is not regulated by the ePD in case of information society services.

[1254] Eran Toch et al, 'Analyzing large-scale human mobility data: a survey of machine learning methods and applications (2019) Vol 58 Knowledge and Information Systems 501, 517.
[1255] Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 89, 99.
[1256] Julien Freudiger, Reya Shokri, Jean-Pierre Hubaux, 'Evaluating the Privacy Risk of Location-Based Services' in Danezis Georg (ed) *Financial Cryptography and Data Security* (Springer 2012) 36.
[1257] Drakonakis Kostas et al, 'Please Forget Where I Was Last Summer: The Privacy Risks of Public Location (Meta) Data' (2019) 2 <https://arxiv.org/pdf/1901.00897.pdf> accessed 8 February 2024.
[1258] Article 2 (4) EECC.
[1259] VoIP solutions, for example, enable individuals to call via computer without the call being routed on to a number in the regular telephony numbering plan
[1260] Recital 15 EECC.
[1261] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 9.
[1262] Defined as 'any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services'. See Directive (EU) 2015/1535 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services ('Information Society Services Directive').
[1263] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 14.

Consequently, information society service providers must comply with the GDPR for the further processing of sensitive location data gained by means of mobile devices and can legitimise such processing by a variety of lawful grounds according to Article 6 GDPR.[1264] Given that consent is one of the main legislative tools for giving individuals control over the processing of personal data[1265] – if not the 'ultimate expression of control'[1266] – data subjects seem to have few ways to exercise control over the processing of their location data (apart from exercising their rights). Controllers and particularly information society service providers may rely on a variety of legal bases other than consent. They can legally argue that there is no need to ask permission from individuals to process their location data,[1267] information that is of sensitive nature.[1268] Given that location data are not considered special data under the GDPR, controllers may deploy ML approaches to infer daily activities, behavioural patterns and predict the location of individuals in a particular time period without the need to obtain consent from the individuals concerned. Notably, also the CJEU acknowledges the sensitive nature of profiles that may derived from location-related information.[1269]

The current legal framework does not effectively[1270] protect the fundamental rights to privacy and data protection, because sensitive location data are only regulated strictly under the ePD when it is processed by ECSs,[1271] excluding a broad range of information society services. This fails to achieve the ePD's goal of protecting users from risks regarding their personal data and privacy.[1272] It also fails to fulfil the GDPR's aim to respect the fundamental right to privacy[1273] considering that location data provide a very intimate insight into the private life of individuals[1274] as it reveals where they live, work and shop, which bars and restaurants they visit, which political events they attend and which medical services they need.[1275] Thus, the approach to exhaustively enumerate special categories of personal data creates a gap of protection with regard to the processing of sensitive location data facilitated by

[1264] Note however that regulatory guidance sees informed consent as the main applicable legal ground for the processing of location data Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 13.
[1265] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1266] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona para 73.
[1267] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1268] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 13, 18; Article 29 Working Party, 'Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation 2002/58/EC' (WP 247, 4 April 2017) at 30.
[1269] Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 89, 99.
[1270] Recital 11 GDPR; Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Svergie* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1271] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 7.
[1272] Recital 6 ePD.
[1273] Recital 4 GDPR.
[1274] Article 29 Working Party, 'Opinion 13/2011 on Geolocation services on smart mobile devices' (WP 185, 16 May 2011) at 18.
[1275] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024, see also Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 89, 99.

AI. It also fails to empower data subjects for exercising control regarding the processing of their personal data.[1276] Consent is considered to be one of the main legislative tools for giving individuals control over the processing of their personal data,[1277] if not the 'ultimate expression of control'.[1278] Because information society services do not fall under the scope of the ePD, controllers may rely on a variety of legal bases other than consent for processing location data. They can argue that there is no need to ask permission from individuals to process sensitive location data.[1279] Therefore, the approach to exhaustively enumerate special data and the restricted scope of the ePD are not fit for purpose to protect the fundamental rights to data protection and privacy when considering the gap of protection they create.

Note that locational privacy, i.e. the privacy of information about someone's physical (geographic) location[1280] is protected as such under the fundamental right to privacy. The processing of location data can be regarded as an interference with an individual's fundamental right to privacy.[1281]

The location data problem (Type 3)

ML can infer daily activities of an individual from location data and the processing of such data may have significant privacy implications, allowing to draw a detailed picture about a person's routine, lifestyle and social network. Information society service providers are not obliged to obtain consent for the processing of location data according to Article 9 (1) ePD. Likewise, sensitive location data is not protected as such according to Article 9 (1) GDPR. Consequently, these provisions create significant gaps of protection and are therefore not fit for purpose to protect the fundamental rights to privacy and data protection.

Neurodata

AI is a powerful driver for neurotechnologies which interface with the brain and that sense information about or produced by the brain function and/or offer input or 'write' information into the brain to modulate function.[1282] Advancements in human neuroscience and neurotechnology facilitate unprecedented means for accessing, collecting, sharing and otherwise processing neurodata. Neurodata is any information with respect to brain functions, neural activity, brain signals and any other information relating to the human brain ('neurodata').[1283] This broad definition includes brain signals

[1276] Recital 7 GDPR.
[1277] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1278] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona para 73.
[1279] Article 29 Working Party, 'Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation 2002/58/EC' (WP 247, 4 April 2017) at 30.
[1280] Bert-Jaap Koops et al 'A Typology of Privacy' (2017) Vol. 38 Iss. 2 University of Pennsylvania Journal of International Law 438, 500.
[1281] *Uzun v Germany United* App no 35623/05 (ECtHR 2 December 2010) paras 51-52
[1282] Karen S Rommelfanger et al, 'Mind the Gap: Lessons Learned from Neurorights' AAAS Center for Science Diplomacy (2022) < https://www.sciencediplomacy.org/article/2022/mind-gap-lessons-learned-neurorights > accessed 8 February 2024.
[1283] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1.

measured by means of electroencephalography (EEG) and neuroimaging. The latter refers to the various techniques used to create images of the structures and/or functioning of the nervous system.[1284] EEG is a non-invasive method consisting of detection and registration of electrical activity occurring in the brain. It relies on electrodes attached to the scalp that register changes of electric potential on the skin surface caused by the activity of cerebral neurons. After their amplification, they form a record, namely, an encephalogram.[1285]

Brain-computer interfaces (BCIs), also known as mind-machine interfaces, are designed to translate brain signals into computer commands. They facilitate communication between the human brain and devices.[1286] BCIs enable their users to send commands to computers by means of brain signals alone which are usually measured by means of electroencephalography (EEG).[1287] In the beginning, BCIs have largely focussed on medical assistive applications to improve the quality of life for patients, for example on applications that enable advanced communications with paralysed patients.[1288] Recently, BCIs have been developed for non-clinical applications, such as for the purpose of entertainment, mental state monitoring, virtual reality and in Internet of Things (IoT) services,[1289] device control or real-time neuromonitoring, neurosensory-based vehicle operator systems, wearables for mental well-being and virtual reality systems.[1290] Kernel intends to 'hack the human brain'[1291] and Facebook wants to develop means of controlling devices directly with neurodata.[1292]

All these BCI applications process neurodata. Data acquisition methods facilitating the collection of neurodata used for BCI applications vary and include EEG, magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI).[1293] Non-invasive BCIs, which currently are most widely used in BCI research, place sensors on the scalp to acquire EEG signals.[1294] The development

[1284] Damian Eke et al, 'Pseudonymisation of neuroimages and data protection: Increasing access to data while retaining scientific utility' (2021) Vol 1 Iss 4 Neuroimage 1-12.
[1285] Szczepan Paszkiel, *Analysis and Classification of EEG Signals for Brain–Computer Interfaces* (Springer Nature 2020) 3.
[1286] Hongchang Shan, 'Towards High Performance and Efficient Brain Computer Interface Character Speller: Convolutional Neural Network based Methods' Dissertation Universiteit Leiden 2020, 1.
[1287] Camille Jeunet, Bernard N'Kaoua, Fabien Lotte, 'Chapter 1 - Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates' in Damien Coyle (ed) *Progress in Brain-Computer Interfaces: Lab Experiments to Real-World Applications* (Elsevier 2016) 4.
[1288] Brent J. Lance et al, 'Brain-Computer Interface Technologies in the Coming Decades' (2012) Vol 100 Proceedings of the IEE 1585.
[1289] Hongchang Shan, 'Towards High Performance and Efficient Brain Computer Interface Character Speller: Convolutional Neural Network based Methods' Dissertation Universiteit Leiden 2020, 1.
[1290] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 4.
[1291] Nick Statt, 'Kernel is Trying to Hack the Human Brain—But Neuroscience has a Long Way to Go' *The Verge* (New York 22 February 2017) <https://www.theverge.com/2017/2/22/14631122/kernel-neuroscience-bryanjohnson-human-intelligence-ai-startup> accessed 8 February 2024.
[1292] John Constine, 'Facebook is building brain-computer interfaces for typing and skin-hearing' TechCrunch (San Francisco 19 April 2017) < https://techcrunch.com/2017/04/19/facebook-brain-interface/ > accessed 8 February 2024.
[1293] Szczepan Paszkiel, *Analysis and Classification of EEG Signals for Brain–Computer Interfaces* (Springer Nature 2020) 1.
[1294] Hongchang Shan, 'Towards High Performance and Efficient Brain Computer Interface Character Speller: Convolutional Neural Network based Methods' Dissertation Universiteit Leiden 2020, 2.

of consumer-directed wearable devices to record brain activity based on EEGs will likely lead to the analysis of neurodata at a large scale.[1295] Before neurodata can be useful for specific purposes, it must be 'de-coded' meaning that features must be extracted and classified according to known particularities of a specific brain activity.[1296] AI proves to be very helpful for such de-coding.[1297] BCI applications use different ML techniques for the classification of EEG signals.[1298] For example, researchers have used a convolutional neural network[1299] (CNN) to decode movement-related information from EEG data.[1300] ML and particularly DL approaches modelled on ANNs will be useful for this and allow fine-grained decoding of neurodata.[1301] Classification techniques[1302] used for supervised ML[1303] introduced in Section 2.2.1.1 as well as feature extraction techniques from the AI discipline CV[1304] can adaptively decode neurodata.[1305] Because most existing EEG decoding methods separate feature extraction from classification, it has been suggested to develop deep convolutional networks from DL to decode neurodata[1306] which combine feature extraction and classification. In addition, neurodata may be used for the purpose of artificially generating speech by means of NLP. Because neurodata associated with speech can be recorded from specific articulatory motor areas in the brain, unvoiced speech can be reconstructed and realised synthetically via a speaker.[1307]

Developments of ML, CV, NLP and DL applied to BCI open the possibility to analyse neurodata. It is very likely that processing of neurodata constitutes processing of personal data, in particular due to

[1295] Philipp Kellermayr, 'Big Neurodata: On the Responsible Use of Neurodata from Clinical and Consumer-Directed Neurotechnological Devices' (2018) Vol 14 Neuroethics 83, 84 <https://link.springer.com/article/10.1007/s12152-018-9371-x> accessed 8 February 2024.
[1296] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 2 < https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.
[1297] Karen S Rommelfanger et al, 'Mind the Gap: Lessons Learned from Neurorights' AAAS Center for Sience Diplomacy (2022) < https://www.sciencediplomacy.org/article/2022/mind-gap-lessons-learned-neurorights > accessed 8 February 2024.
[1298] Mamunir Rashid et al, 'The classification of EEG Signal Using Different Machine Learning Techniques for BCI Application' in J.-H. Kim et al (Eds) *Robot Intelligence Technology and Applications* (Springer 2018) 207-221.
[1299] Type of network architecture in DL.
[1300] Philipp Kellermayr, 'Big Neurodata: On the Responsible Use of Neurodata from Clinical and Consumer-Directed Neurotechnological Devices' (2018) Vol 14 Neuroethics 83, 86 <https://link.springer.com/article/10.1007/s12152-018-9371-x> accessed 8 February 2024.
[1301] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 2, 3 < https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.
[1302] Camille Jeunet, Bernard N'Kaoua, Fabien Lotte, 'Chapter 1 - Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates' in Damien Coyle (ed) *Progress in Brain-Computer Interfaces: Lab Experiments to Real-World Applications* (Elsevier 2016) 4, 5.
[1303] Szczepan Paszkiel, *Analysis and Classification of EEG Signals for Brain–Computer Interfaces* (Springer Nature 2020) 42.
[1304] Mark Nixon, Alberto Aguado, *Feature Extraction & Image Processing for Computer Vision* (3rd edn Elsevier 2012).
[1305] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 3 < https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.
[1306] Implementing a joint space–time–frequency feature extraction scheme for EEG decoding see Dongye Zhao et al, 'Learning joint space–time–frequency features for EEG decoding on small labeled data' (2019) Vol 114 Neural Networks 67.
[1307] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 11 < https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.

the personal nature of the brain itself: brain characteristics are largely determined by genetic factors that are often unique to individuals.[1308] Additionally, certain forms of neurodata remain unique to one specific individual regardless of attempts to segregate the link between neurodata and this specific individual.[1309] Neurodata is said to provide unique insights into people[1310] and their behaviour.[1311] Neurodata are a particularly sensitive class of data due to their direct link with mental processes[1312] and the strong link to the individual's personhood.[1313] Despite this, it is clear that neurodata as such is not considered a special category of personal data according to the GDPR.[1314] However, in some cases and depending on the context, the processing of neurodata could reveal data that is protected as a special category such as genetic data,[1315] racial and ethnic origin,[1316] health data[1317] or biometric data.[1318] Apart from these very specific cases, highly sensitive neurodata do not fall under the definition of special categories of personal data. Neurodata relates to processes of the human mind, which represents a uniquely sensitive and intimate space in the individual's private sphere. Neurodata is not only sensitive because of what can be concluded from it in terms of mental states, but also in view of inferred data, such as insights into a data subject's personality, cognitive capacity and future behaviour.[1319] It may also reveal sensitive neuronal states that are associated with below average functioning something that is not health data as such. When revealed, such data may result in discrimination. For example, someone may be labelled or classified as 'stupid' simply due to the detection of uncommon neuronal states.[1320] Because of its sensitive nature[1321] and the sensitive information that can be inferred

[1308] Therefore, neurodata could be used for so called 'brain-fingerprinting'. See Kuldeep Kumar et al, 'Multi-modal brain fingerprinting: A manifold approximation based framework' (2018) Vol 183 Neuro-Image 212-226.
[1309] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 65 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024; See also Damian Eke et al, 'Pseudonymisation of neuroimages and data protection: Increasing access to data while retaining scientific utility' (2021) Vol 1 Iss 4 Neuroimage 1-12.
[1310] Neurodata are of highly personalised nature and allows for identification ('brain fingerprinting').
[1311] Brent J. Lance et al, 'Brain-Computer Interface Technologies in the Coming Decades' (2012) Vol 100 Proceedings of the IEE 1587.
[1312] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 14; Marcello Ienca, Karolina Ignatiadis, 'Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges' (2020) Vol 11 Iss 2 AJOB Neuroscience 77-87; Rafael Yuste et al, 'Four ethical priorities for neurotechnolgies and AI' (2017) Vol 551 Nature 159-163.
[1313] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 14.
[1314] Because the lawmaker arguably did not anticipate the use of this novel type of data, since it is not mentioned in Article 9 (1) GDPR or in corresponding recitals.
[1315] When revealing genetic features such as biomarkers.
[1316] Morphological differences between various sections of the brain in different individuals allows the identification of different ethnical groups; see Wei Liang Chee et al, 'Brain Structure in Young and Old East Asians and Westerners: Comparison of Structural Volume and Cortical Thickness' (2011) Vol 23 Iss 5 Journal of Cognitive Neuroscience <www.ncbi.nlm.nih.gov/pmc/articles/PMC3361742/> accessed 8 February 2024.
[1317] When neurological problems or brain diseases are detected.
[1318] When neurodata are used to identify an individual.
[1319] Karen S Rommelfanger et al, 'Mind the Gap: Lessons Learned from Neurorights' AAAS Center for Sience Diplomacy (2022) < https://www.sciencediplomacy.org/article/2022/mind-gap-lessons-learned-neurorights > accessed 8 February 2024; Ryan H Purcell, Karen S Rommelfanger, 'Internet-Based Brain Training Games, Citizen Scientists, and Big Data: Ethical Issues in Unprecedented Virtual Territories' (2015) Vol 86 Iss 2 Neuron 356, 357.
[1320] Jan-Hendrik Heinrichs, 'The Sensitivity of Neuroimaging Data' (2012) Vol 5 Iss 2 Neuroethics 185, 193.
[1321] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 14; Marcello Ienca, Karolina Ignatiadis, 'Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges' (2020) Vol 11 Iss 2 AJOB Neuroscience 77-87; Rafael Yuste et al, 'Four ethical priorities for neurotechnolgies and AI' (2017) Vol 551 Nature 159-163.

from it, neurodata should receive specific protection under the GDPR. The high level of protection of neurodata (as special data) should include neural activity occurring in the human brain which generates the neurodata.[1322] This means that neurodata would already be protected before it is 'de-coded', revealing for instance mental data (see mental data problem later in this section). This is needed to protect sensitive information that might be inferred from it. Inferences derived from neurodata can be used to influence an individual's commercial, social and political behaviour. For example, information derived from neurodata may be used to tailor content or experiences in a way that is more addictive for individuals concerned based on psychology.[1323]

Article 9 (1) GDPR does not list neurodata. Because neurodata itself does not receive specific protection under the GDPR, the approach to exhaustively enumerate special data is not fit for purpose to effectively[1324] protect the fundamental right to data protection. In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect[1325] the data subject's personal data against risk of misuse.[1326] Such risk of misuse is high when considering that inferences drawn from neurodata may be used to influence an individual's commercial, social and political behaviour. Due to its direct link with mental processes[1327] and an individual's personhood,[1328] neurodata is highly sensitive and provides unique insights into an individual's behaviour. [1329] Therefore, the processing of neurodata can pose significant risks to the fundamental rights and freedoms of individuals. By virtue of its content, neurodata carries the risk of infringing the individual's fundamental right to privacy (see also the mental data problem discussed later in this section and Section 5.4) that the GDPR envisages to protect.[1330] Article 9 (1) GDPR can neither ensure a high level of protection[1331] nor a strong and coherent data protection framework[1332] when considering the gap of protection it creates.

[1322] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 14.
[1323] Karen S Rommelfanger et al, 'Mind the Gap: Lessons Learned from Neurorights' AAAS Center for Sience Diplomacy (2022) < https://www.sciencediplomacy.org/article/2022/mind-gap-lessons-learned-neurorights > accessed 8 February 2024.
[1324] Recital 11 GDPR.
[1325] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[1326] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1327] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 14; Marcello Ienca, Karolina Ignatiadis, 'Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges' (2020) Vol 11 Iss 2 AJOB Neuroscience 77-87; Rafael Yuste et al, 'Four ethical priorities for neurotechnolgies and AI' (2017) Vol 551 Nature 159-163.
[1328] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 14.
[1329] Brent J. Lance et al, 'Brain-Computer Interface Technologies in the Coming Decades' (2012) Vol 100 Proceedings of the IEE 1587.
[1330] Recital 4 GDPR.
[1331] See Recitals 6 and 10 GDPR, as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66; Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[1332] Recital 7 GDPR.

The neurodata problem (Type 3)

ML, CV, NLP and DL facilitate the processing of neurodata. Neurodata provide unique insights into people and are particularly sensitive due to their direct link with mental processes and an individual's personhood. Despite this, neurodata is not protected as special data under the GDPR because the approach to enumerate special categories of personal data exhaustively cannot keep up with the developments in AI. Consequently, this approach creates a significant gap of protection and is therefore not fit for purpose to protect the fundamental right to data protection.

Mental data

Neurotechnologies powered by AI have an unprecedented ability to decode information about mental states or processes by analysing data concerning neural activity patterns and 'transcribe' mental states by modulating neural computation.[1333] When processed by AI systems, neurodata as described earlier in this section may reveal mental data, which is any information about mental states and processes of individuals ('mental data').[1334] Mental states and processes include information related to all conscious and non-conscious mental representations, events, propositional attitudes, including thoughts, beliefs, emotions, moods and underlying psychological mechanisms.[1335] Mental data constitutes information relating to the core of an individual's private sphere,[1336] including information such as thoughts, memories and intentions. The processing of neurodata by AI systems, in particular ML and DL, allows one to derive insights in an individual's mental domain[1337] and particularly insights in 'real-time' mental processes.[1338] ML and DL approaches offer powerful capabilities (e.g. to detect patterns and make predictions) to infer a variety of highly sensitive information[1339] from neurodata, including dimensions of an individual's thoughts, intentions and sometimes even information that is not known to an individual herself or beyond her control.[1340] Through the processing of neurodata by means of AI, mental data becomes accessible. This indicates a partial overlap between the two categories of

[1333] Abel Wajnerman Paz, 'Is Your Neural Data Part of Your Mind? Exploring the Conceptual Basis of Mental Privacy' (2022) Vol 32 Minds and Machines 395, 396.
[1334] Ibid; see a similar definition by Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 4.
[1335] Jan-Christoph Bublitz, 'The Nascent Right to Psychological Integrity and Mental Self-Determination' in Andreas von Arnauld, Kerstin von der Decken, Mart Susi (eds) *The Cambridge Handbook of New Human Rights* (Cambridge University Press 2020) 30; Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 4.
[1336] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 65 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.
[1337] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 3.
[1338] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 65 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.
[1339] Marcello Ienca, Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotech- nology' (2017) Vol 13 Iss 1 Life Science, Society and Policy 1, 24.
[1340] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 65 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.

data. However, not all neurodata constitute mental data and vice versa. In addition, emotion data as discussed in the first part of this section can be seen as a subcategory of mental data. The relationship between neurodata, mental data and emotion data is illustrated in Figure 2.1.

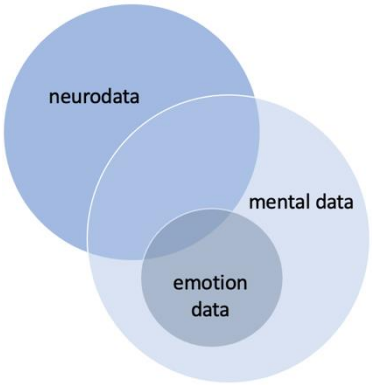

Figure 2.1
Overlaps between neurodata, mental data and emotion data.

Neurodata may be used to predict future behaviour, brain states and other aspects of an individual.[1341] When processed by AI systems, neurodata facilitates the inference of mental states of individuals. It should be noted that mental data may be generated from both neurodata and other data.[1342] Therefore, mental data and neurodata only partially overlap,[1343] as shown in Figure 2.1. For example, information regarding the emotional states of individuals might be inferred by approaches developed within the AI discipline AC as introduced in Sections 2.2.4.1 and 2.2.4.2, which do not comprise the processing of neurodata.[1344] This is illustrated in Figure 2.1 which shows that emotion data, seen as a subcategory of mental data, partially overlaps with neurodata. In addition, mental data may be inferred from digital footprints such as Facebook likes, tweets or credit card records when analysed by AI (for example, ML).[1345] Mental data form the core of an individual's private sphere[1346] and are therefore of a particularly sensitive nature. Risks associated with the processing of mental data are considerable because mental representations are the closest psychological substrate of fundamental ethical-legal notions[1347] such as personal autonomy. By using insights gained from the processing of mental data, BCI systems may influence the development of an individual's reasons by altering options to act independently, which has a negative impact to the self-determination of the individual concerned.[1348] Affective BCIs

[1341] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 3, 12, 14 < https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.
[1342] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 4.
[1343] Andrea Lavazza, 'Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis' (2018) Vol 12 Frontiers in Neuroscience 1-10.
[1344] The notions 'emotion data' and mental data partly overlap as the latter also covers emotion data. However, this section focusses on thoughts and other mental states.
[1345] Sandra C Matz et al, 'Privacy in the age of psychological targeting' (2020) Vol 31 Current Opinion in Psychology 116-221.
[1346] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 68 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.
[1347] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 5.
[1348] Orsolya Friedrich et al, 'An Analysis of the Impact of Brain-Computer Interfaces on Autonomy' (2021) Vol 14 Iss 1 Neuroethics 17, 27.

use neurodata to extract features related to affective states such as emotions and may even stimulate and influence emotions.[1349] For example, an affective BCI system was developed to detect the current emotional state with the aim to modulate it accordingly by moving individuals from one emotional state to another.[1350] Such affective BCIs are problematic in terms of personal autonomy because they monitor, influence and directly stimulate emotional states of individuals.[1351] Influencing individuals may include manipulative forms of nudging. Nudges are 'interventions that steer people in particular directions but that also allows them to go their own way'.[1352] Nudging may be manipulative, for instance, if it is used to subvert an individual's decision-making powers.[1353]

Mental data may contain very sensitive information with respect to unexecuted behaviour such as unuttered thoughts and intended actions,[1354] information that previously was inaccessible to others. The developments in neurotechnology powered by AI can bypass the cognitive process of filtering and selectively sharing information that people typically perform to control the flow of information about them. Thus, information a person decided not to share may become available to others anyway.[1355] For example, thoughts and intentions can be disclosed by interpreting neurodata and decode it by ML and DL approaches. Researchers have achieved translating brain activity into text by means of ML and ANN approaches.[1356] Developments in neurotechnology, powered by ML and DL approaches, have unlocked the human brain to some extent.[1357] Neurodata in the form of connection patterns and activation of nerve cells are believed to constitute partial correlates of mental states an individual has at any given time.[1358] AI proves helpful to de-code such neurodata. A study has achieved to decode what the brain is neurally representing by means of CNN.[1359] However, current applications can often only decode a rather limited set of predetermined mental states from available neurodata.[1360] They are not yet able to decode mental information per se, but are sophisticated enough to establish statistically significant relations between certain patterns of neurodata and other data on the one hand,

[1349] Steffen Steinert, Orsolya Friedrich, 'Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces' (2020) Vol 26 Science and Engineering Ethics 351, 353.
[1350] Ian Daly et al, 'Affective brain–computer music interfacing' (2016) Vol 13 No 4 Journal of Neural Engineering
[1351] Steffen Steinert, Orsolya Friedrich, 'Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces' (2020) Vol 26 Science and Engineering Ethics 351, 355.
[1352] Cass R Sunstein, 'The Ethics of Nudging' (2015) Vol 32 Yale Journal of Regulation 413, 417.
[1353] Ibid, 446.
[1354] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 6.
[1355] Abel Wajnerman Paz, 'Is Mental Privacy a Component of Personal Identity?' (2021) Vol 15 Frontiers in Human Neuroscience 2.
[1356] Joseph G Makin, David A Moses, Edward F Chang, 'Machine translation of cordial activity to text with an encoder-decoder framework' (2020) Vol 23 Nature Neuroscience 575.
[1357] Marcello Ienca, Roberto Andorno, 'Towards new human rights in the age of neuroscience and neurotechnology. Life Sciences, Society and Policy' (2017) Vol 13 Life Sciences, Society and Policy 5 <https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-017-0050-1> accessed 8 February 2024.
[1358] Andrea Lavazza, 'Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis' (2018) Vol 12 Frontiers in Neuroscience 1, 3.
[1359] Haiguang Wen et al, 'Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision' (2018) Vol 28 Iss 12 Cerebral Cortex 4136-4160.
[1360] Abel Wajnerman Paz, 'Is Your Neural Data Part of Your Mind? Exploring the Conceptual Basis of Mental Privacy' (2022) Vol 32 Minds and Machines 395, 397.

and the actual occurrence of mental states on the other hand. Information inferred from mental data and neurodata[1361] may have considerable (mental) privacy implications (see Section 5.4).

Mental data falls, as such, not under the definition of special data in the GDPR[1362] despite its highly intimate and sensitive nature. Because mental data does not receive specific protection under the GDPR, the approach to enumerate special data exhaustively is not fit for purpose to effectively[1363] protect the fundamental right to data protection. In its case law, the CJEU has repeatedly stressed that EU data protection law aims to effectively protect[1364] the data subject's personal data against risk of misuse.[1365] Such risk of misuse seems relatively high when considering that AI, sooner or later, be able to decode neurodata in a way that discloses an individual's mental states, their thoughts in particular. Thus, there is a clear conceptual and normative gap regarding the protection of mental data. It is difficult to assert that thoughts, and mental data more generally, are less sensitive than the special categories of personal data[1366] listed in the GDPR. Processing inherently sensitive mental data is prone to create significant risks to the fundamental rights and freedoms of individuals. Mental data carries the risk of infringing an individual's fundamental right to privacy (see also Section 5.4) that the GDPR also envisages to protect.[1367] Additionally, Article 9 (1) GDPR can neither ensure a high level of protection[1368] nor a strong and coherent data protection framework[1369] when considering the gap of protection it creates.

The mental data problem (Type 3)

ML and AC facilitate the processing of mental data, i.e. any data used to infer mental states of individuals including thoughts, beliefs and underlying mechanisms and processes. Mental data are inherently sensitive and form the core of an individual's private sphere. Despite this, mental data are not specifically protected under the GDPR because the approach to enumerate special categories of personal data exhaustively cannot keep up with the developments in AI. This principle creates a significant gap of protection and is not fit for purpose to protect the fundamental right to data protection.

[1361] Marcello Ienca, Gianclaudio Malgieri, 'Mental data protection and the GDPR' (2022) Vol 9 Iss 1 Journal of Law and the Biosciences 1, 6.
[1362] Stephen Rainey et al, 'Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?' (2020) Journal of Law and the Biosciences 16 <https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsaa051/5864051?searchresult=1> accessed 8 February 2024.
[1363] Recital 11 GDPR.
[1364] Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-319/20, *Meta Platforms Ireland Limited* [2022] ECR I-322 para 73.
[1365] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014] ECR-I 238 paras 54 and 66; Joined Cases C-203/15 and C-698/15 [2016] *Tele 2 Sverige* ECR-1 970 paras 109 and 122; Case C-362/14 *Schrems v Data Protection Commissioner* [2015] ECR I-650 para 91.
[1366] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 67 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.
[1367] Recital 4 GDPR.
[1368] See Recitals 6 and 10 GDPR, as well as CJEU case law, such as Case C-534/20, *Leistritz AG* [2022] ECR I-594 para 26; Case C-645/19 [2021] *Facebook Ireland* ECR I-483 para 45; Case C-511/18, *La Quadrature du Net* [2020] ECR I-791 para 106; Case C-131/12, *Google Spain* [2014] ECR I-317 para 66. Case C-534/20, Case C-154/21 *Österreichische Post AG* [2023] ECR I-3 para 44; Case C-132/21 *Nemzeti* [2023] ECR I-2 para 42.
[1369] Recital 7 GDPR.

A Type 3 legal problem may also be identified with regard to the protection of human minds per se, namely, the forum internum, which is denoted as a layer of the private sphere that describes the mental world of an individual.[1370] Whereas in human rights law the forum internum theoretically enjoys absolute and unconditional protection,[1371] it is doubtful whether this in fact applies in practice because the absolute, unimpugnable and fundamental nature of the forum internum seems to be undermined since individuals are not able to enforce their rights with regard to the forum internum.[1372] This will be discussed in the context of mental privacy (Section 5.4).

4.9 Confidentiality of communication

AI and people's interactions with it do not fit neatly into paradigms of communication theory that have long focussed on human–human communication.[1373] As I outline in this section, the same can be said about the legal protection concerning the confidentiality of human-machine communication. The GDPR regulates the processing of personal data, but not specifically the confidentiality of communication. This is regulated by the ePrivacy Directive ('ePD') as introduced in Section 3.4 and potentially the future ePrivacy Regulation.[1374] However, the obligation to ensure the confidentiality of communications and the general prohibition of listening, tapping, storing or other kinds of surveillance of communications and traffic data according to Article 5 (1) ePD *solely* applies to providers of publicly available electronic communication services (ECS) and providers of public electronic communication networks[1375] in the EU. Companies that provide virtual assistant services are not subject to Article 5 (1) ePD because they do not qualify as an ECS. As outlined in Section 3.4.1, an ECS covers Internet access services, interpersonal communications services and services consisting wholly or mainly in the conveyance of signals.[1376]

Clearly, virtual assistant services do not constitute Internet access services. In addition, they are not interpersonal communication services,[1377] because these services do not relate to communication

[1370] Dara Hallinan et al, 'Neurodata and Neuroprivacy: Data Protection Outdated?' (2014) Vol 12 Iss 1 Surveillance and Society 68 <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/neurodata/neurodata4> accessed 8 February 2024.
[1371] Article 9 ECHR.
[1372] Paul M Taylor, *Freedom of Religion UN and European Human Rights Law and Practice* (2005 Cambridge University Press) 202.
[1373] Andrea L Guzman, Seth C Lewis, 'Artificial intelligence and communication: A Human-Machine Communication agenda' (2020) Vol 22 Iss 1 New Media & Society 70-86.
[1374] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) Recital 2 <https://data.consilium.europa.eu/doc/document/ST-12633-2019-INIT/en/pdf > accessed 8 February 2024.
[1375] Defined as 'electronic communications network used wholly or mainly for the provision of publicly available electronic communications services which support the transfer of information between network termination points'; see Article 2 (8) EECC.
[1376] Article 2 (4) EECC.
[1377] As defined in Article 2 (5) EECC: 'service normally provided for remuneration that enables *direct interpersonal and interactive exchange of information* via *electronic communications networks* between a *finite number of persons*' emphasis added.

between natural persons.[1378] Rather, they relate to communications between natural persons and a *machine*. Recital 17 EECC clarifies what interpersonal communication means: communication between *natural persons*. Communications involving legal persons fall within the definition only to a limited extent, for instance, if natural persons act on behalf of those legal persons.[1379] Therefore, human-machine communications fall outside the scope of interpersonal communication services defined in Article 2 (5) EECC. In addition, virtual assistant services do also not qualify as machine-to-machine services under the EECC. Recital 249 EECC says such services involve 'an automated transfer of data and information between devices or software-based applications with limited or no human interaction.' Virtual assistant services involve more than only limited human interaction.

A service provider is *responsible* vis-à-vis the end-users for *transmission* of the *signal* which ensures that users are supplied with the service to which they have subscribed.[1380] Clearly, providers of virtual assistant services are not responsible for the transmission of the signal. Rather, the Internet Access Providers (IAPs) and the *operators* of the *various networks* of which the open web is constituted are responsible for this.[1381]

Services facilitating human-machine communications do not qualify as an ECS which is problematic with regard to confidentiality. As will be described in Section 4.9.3, this applies particularly to the confidentiality of human-machine communications enabled by the AI disciplines NLP and AC when embedded in virtual assistants and smart devices connected to the Internet of Things ('IoT'). The IoT is the cyber-physical ecosystem of interconnected physical and potentially virtual sensors and actuators.[1382] It consists of devices such as smartphones, wearables and even toothbrushes which are connected together.[1383] The growing use of virtual assistants and smart home devices causes serious concerns about the confidentiality of communication and how related data are processed and controlled.[1384] For example, Amazon's virtual assistant Alexa is bound to be embedded in toilets, e-bikes, beds, cars and other everyday objects.[1385]

To be clear, providers of human-machine communication services need to adhere to the GDPR when processing personal data. Whereas both the GDPR and the ePD aim to protect fundamental rights and

[1378] Article 2 (5) EECC and Recital 17.
[1379] It seems unclear what the phrase 'or are at least involved on one side of the communication' contained in Recital 15 EECC precisely means.
[1380] Case C-475/12, *UPC* [2014] ECR I-285 para 43.
[1381] Case C-193/18, *Google LLC* [2019] ECR I-498 para 36.
[1382] European Union Agency for Network and Information Security, 'Good Practices for Security of Internet of Things' (2018) 45 <https://www.enisa.europa.eu/publications/good-practices-for-security-of-iot/@@download/fullReport> accessed 8 February 2024.
[1383] Matt Burgess, 'What is the Internet of Things? WIRED explains' *Wired* (New York, 16 February 2018) <https://www.wired.co.uk/article/internet-of-things-what-is-explained-iot> accessed 8 February 2024.
[1384] Tine Munk, 'Does Online Privacy Exist in the GDPR Era? The Google Voice Assistant Case' in Tatiana-Eleni Synodiou et al (eds) *EU Internet Law in the Digital Single Market* (Springer Nature 2021) 489.
[1385] Amrita Khalid, 'Alexa was everywhere at CES' *Quartz* (New York, 10 January 2020) <https://qz.com/1783414/amazons-alexa-was-everywhere-at-ces-2020/> accessed 8 February 2024.

freedoms,[1386] the GDPR sets general rules for processing personal data, and the ePD regulates the fundamental right to privacy *and* data protection in the electronic communications sector.[1387] Thus, merely because providers of human-machine communication services fall outside the scope of the ePD does not lead to a complete lacuna in legal protection. However, the provisions of the GDPR are less strict than Article 5 (1) ePD, which requires consent for the surveillance of interpersonal communications. Arguably, and as outlined in Section 4.9.3, human-machine communications deserve the same level of confidentiality as interpersonal communications.

4.9.1 Legal problems: Type 1

As will be outlined in Section 4.9.3, problems arise with respect to the confidentiality of human-machine and interpersonal communication in human-machine communication services, such as virtual assistant services, which do not qualify as ECS. Such services are therefore excluded from the scope of Article 5 (1) ePD, which prohibits surveillance of communications and related traffic data without consent of the user. Provisions that are not applicable to the company processing data cannot be violated. Therefore, no specific Type 1 legal problems arise.

4.9.2 Legal problems: Type 2

As outlined in Section 4.9.1, provisions enshrined in the current legal framework that are not applicable to providers of human-machine communication services, cannot be violated. Consequently, no specific Type 2 legal problems arise because provisions that are not applicable to a certain processing cannot be violated, and thus they also do not need to be enforced.

4.9.3 Legal problems: Type 3

Because of the restricted material scope of the ePD, the prohibition of listening, tapping, storage or other kinds of interception or surveillance of communications without consent of the individual concerned does not apply to human-machine and interpersonal communications occurring in the context of virtual assistants and smart home technologies powered by NLP and ML. Omission to subject such services to the material scope of the ePD creates a loophole for the providers of the services. A loophole exists where a failure to include something in the law allows someone to do something generally considered illegal.[1388] This occurs here due to the omission of not including virtual assistant and smart home services in the scope of the ePD. Due to this omission, providers of such services are not subject

[1386] In the case of the GDPR, the fundamental right to the protection of personal data, and in the case of the ePrivacy Directive, both the fundamental right to privacy (Recital 12) and data protection (Recital 2). Note that the Directive which amended the ePrivacy Directive also refers to the fundamental right to privacy and confidentiality (Recital 51) and the fundamental right to the protection of personal data (Recital 56).
[1387] Christina Etteldorf, 'EDPB on the Interplay between the ePrivacy Directive and the GDPR' (2019) Iss 5 No 2 European Data Protection Law Review 224, 226.
[1388] See <https://dictionary.cambridge.org/dictionary/english/loophole> accessed 8 February 2024.

to the confidentiality obligation enshrined in the ePD. They may intercept human-machine and interpersonal communications without needing to seek consent for intercepting such communications.[1389] This is particularly problematic when considering the extensive use of virtual assistants, smart home applications and similar services. Today, people routinely communicate with virtual assistants such as Amazon Alexa, Siri or Google Assistant.

In essence, virtual assistants are *software applications* equipped with the capabilities to interpret human speech as a question or instruction, perform tasks and respond using synthesised voices.[1390] Virtual assistants are made of several components designed to resolve specific challenges, for example, understanding and producing speech.[1391] They employ sophisticated NLP capabilities enabling users to interact with them conversationally. Put simply, virtual assistants work as follows. The virtual assistant permanently analyses every sound in its environment to recognise its 'wake word', which activates the recording of the user's request.[1392] A request is sent to the virtual assistant's service platform (thus *not* kept on the device) where speech is converted into text by means of speech recognition powered by NLP which translates the text into machine-readable instructions.[1393] Because virtual assistants permanently listen to detect the wake word which activates recording, virtual assistants are referred to as 'always-on' microphone-enabled devices.[1394] Accidental recordings are common in virtual assistant services and occur where virtual assistants activate, transmit and/or record audio from their environment when the wake word is *not* spoken.[1395] Such recordings are caused by accidental triggers, namely, sounds that wrongfully trigger virtual assistants, and occur within the whole range of virtual assistants available on the market, including Amazon Alexa, Google Assistant and Siri.[1396] Activating the wake word by accidental triggers is problematic because it leads to the recording (and upload to the cloud) of potentially sensitive audio data.[1397]

NLP provides powerful means to analyse voice and speech data obtained by means of virtual assistants (VA), in particular when combined with classification techniques adopted in the AI discipline

[1389] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1390] See, for an overview Roberto Pieraccini, *AI Assistants* (MIT Press 2021).
[1391] Roberto Pieraccini, *AI Assistants* (MIT Press 2021) 7.
[1392] Lea Schönherr et al, 'Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers' (2020) at 1 < https://arxiv.org/pdf/2008.00508.pdf > accessed 8 February 2024.
[1393] Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3.
[1394] Yousra Javed, Shashank Sethi, Akshay Jadoun, 'Alexa's Voice Recording Behavior: A Survey of User Understanding and Awareness' (ARES '19, Canterbury 26-29 August 2019) 3 <https://dl.acm.org/doi/10.1145/3339252.3340330> accessed 8 February 2024.
[1395] Daniel J Dubois et al, 'When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers' (2020) Iss 4 Proceedings on Privacy Enhancing Technologies 255-276.
[1396] Daniel J Dubois et al, 'When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers' (2020) Iss 4 Proceedings on Privacy Enhancing Technologies 255-276; Lea Schönherr et al, 'Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers' (2020) < https://arxiv.org/pdf/2008.00508.pdf > accessed 8 February 2024.
[1397] Lea Schönherr et al, 'Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers' (2020) at 2 < https://arxiv.org/pdf/2008.00508.pdf > accessed 8 February 2024.

ML (see Section 2.2.1.1). With NLP and ML, rather sensitive information can be derived from human speech and other acoustic elements in recorded audio. In addition to the linguistic content of speech, a speaker's voice characteristics and manner of expression may contain a rich array of personal information, including clues about the speaker's biometric identity, personality, physical traits, geographical origin, level of intoxication/sleepiness, age, gender, health condition and even an individual's socioeconomic status.[1398]

As outlined in Section 2.2.4.2 regarding the AI discipline AC, speech-based emotion recognition systems measure and quantify emotions of a person by observing speech signals.[1399] Research has demonstrated specific associations between emotions such as fear, anger, sadness, joy and features of speech such as pitch, voice level and speech rate.[1400] Human-machine communication intercepted by means of virtual assistants can therefore also be used to detect the emotional state of the user. Amazon's patented technology enabling Alexa to detect the user's emotional state derived from the user's voice underscores this claim.[1401] Another real-world application is Amazon's wearable 'Halo', which analyses voice tones to detect user emotions.[1402] Information concerning the emotional state of an individual might be particularly helpful to manipulate this individual because emotions play an important role in the elicitation of autonomous motivated behaviour.[1403] According to research in behavioural sciences, especially psychology, emotions constitute powerful, pervasive and predictable drivers of decision-making.[1404] Emotions can therefore have significant effects on economic transactions and play a powerful role in everyday economic choices.[1405]

Companies such as Apple, Amazon, Google and the like offer virtual assistants and intercept, analyse and otherwise process human-machine communication for a plethora of purposes and infer sensitive information by means of ML, NLP and AC, without falling under the scope of the ePD. It should be noted that this lacuna in the current legal framework does not solely apply to human-machine

[1398] Jacob Leon Kröger, Otto Hans-Martin Lutz, Philip Raschke, 'Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference' in Michael Friedewald et al (eds) *Privacy and Identity management. Data for Better Living: AI and Privacy* (Springer 2020) 242.
[1399] Chi-Chun Lee et al, 'Speech in Affective Computing' in Rafael Calvo et al (eds), *The Oxford Handbook of Affective Computing* (OUP 2015) 171.
[1400] Christina Sobn and Murray Alpert, 'Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sandess, and Joy' (1999) Vol 28 No 4 Journal of Psycholinguistic Research, 347.
[1401] Huafeng Jin, Shuo Wang 'Voice-Based Determination of Physical and Emotional Characteristics of Users' US Patent Number US 10096319 B1 (Assignee: Amazon Technologies, Inc.) October 2018 <https://patentimages.storage.googleapis.com/f6/a2/36/d99e36720ad953/US10096319.pdf> accessed 8 February 2024.
[1402] Alex Hern, 'Amazon's Halo wristband: the fitness tracker that listens to your mood' *The Guardian* (London, 28 August 2020) < https://www.theguardian.com/technology/2020/aug/28/amazons-halo-wristband-the-fitness-tracker-that-listens-to-your-mood > accessed 8 February 2024; Austin Carr, 'Amazon's New Wearable Will Know If I'm Angry. Is That Weird?' *Bloomberg* (New York, 31 August 2020) < https://www.bloomberg.com/news/newsletters/2020-08-31/amazon-s-halo-wearable-can-read-emotions-is-that-too-weird > accessed 8 February 2024.
[1403] Leen Vandercammen et al, 'On the Role of Specific Emotions in Autonomous and Controlled Behaviour' (2014) Vol 28 Iss 5 European Journal of Personality 413, 445.
[1404] Jennifer S Lerner et al, 'Emotion and Decision Making' (2015) Vol 66 Annual Review of Psychology 799, 802.
[1405] Jennifer S Lerner, Deborah A Small, George Loewenstein, 'Heart Strings and Purse Strings' (2004) Vol 15 No 5 American Psychology Society 337-340.

communication but also to *interpersonal communications.* All major players in the virtual assistant market (Amazon, Google, Microsoft and Apple) revealed that audio recordings made by their virtual assistants were listened to by either employees or subcontractors to categorise utterances, improve the quality of wake word detection and the performance of speech transaction.[1406] For example, in 2019 Google Assistant recordings were leaked to the Belgian news site VRT NWS. The corresponding report published by the news site outlined that Google employees systematically listened to audio files recorded by Google Home smart speakers and the Google Assistant smartphone app.[1407] Unavoidably, these audio recordings also include interpersonal communications. In the case of Google, the audio snippets contained a wide range of highly sensitive recordings, including private conversations about health status, domestic violence, sexual relationships and drug deals.[1408] In addition, a former Apple employee revealed that he had listened to hundreds of Siri recordings every day, including unintended recordings, for the purpose of quality control.[1409] These recordings concerned sensitive interpersonal communications such as discussions between doctors and patients, business deals, seemingly criminal acts and sexual encounters.[1410] Press coverage points to similar practices at Amazon.[1411]

Of course, providers of human-machine communication services must comply with the provisions enshrined in the GDPR when processing personal data in this context. However, the provisions of the GDPR are less strict than Article 5 (1) ePD, which requires consent for the surveillance of interpersonal communications. According to the GDPR, consent is only one of six legal bases. As outlined in Section 4.4.2, consent is one of the main legislative tools for giving individuals control over the processing of their personal data,[1412] if not the 'ultimate expression of control'.[1413] By excluding human-machine communication services from its scope, the ePD fails to meet its legislative aims to guarantee the confidentiality of communications,[1414] to protect natural persons with respect to the automated storage and processing of data[1415] and ultimately to protect personal data and the privacy of

[1406] CNIL, 'Exploring the ethical, technical and legal issues of voice assistants' (2020) 40 <https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf> accessed 8 February 2024.
[1407] Tim Verheyden et al, 'Hey Google, are you listening?' *VRTB* (Brussels 10 July 2019) <https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms/ > accessed 8 February 2024.
[1408] Tine Munk, 'Does Online Privacy Exist in the GDPR Era? The Google Voice Assistant Case' in Tatiana-Eleni Synodiou et al (eds) *EU Internet Law in the Digital Single Market* (Springer Nature 2021) 497.
[1409] Alex Hern, 'Apple contractors regularly hear confidential details on Siri recordings' *The Guardian* (London, 26 July 2019) < https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings > accessed 8 February 2024.
[1410] Ibid.
[1411] Alex Hern, 'Amazon staff listen to customers' Alexa recordings, report says' *The Guardian* (London, 11 April 2019) <https://www.theguardian.com/technology/2019/apr/11/amazon-staff-listen-to-customers-alexa-recordings-report-says> accessed 8 February 2024.
[1412] Giovanni Butarelli, 'The urgent case for a new ePrivacy law' (2018) <https://edps.europa.eu/press-publications/press-news/blog/urgent-case-new-eprivacy-law_fr> accessed 8 February 2024.
[1413] Case C-300/21 *UI* [2022] ECR I-756 Opinion AG Manuel Campos Sánchez-Bordona para 73.
[1414] Recital 3 ePD.
[1415] Recital 7 ePD.

users.[1416] Human-machine communications deserve the same level of confidentiality as applicable to interpersonal communications under of Article 5 (1) ePD when considering the sensitivity of information captured by human-machine communications and the sensitive information that can be derived from it. Due to this gap in legal protection, Article 5 (1) ePD is not fit for purpose to ensure the confidentiality of human-machine communication and interpersonal communication facilitated by current human-machine communication services and similar future services. This creates a Type 3 legal problem regarding the fundamental rights to privacy and data protection.

The communication surveillance problem (Type 3)

ML, NLP and AC facilitate the surveillance of both human-machine and interpersonal communication. Major tech companies that offer human-machine communication services, such as virtual assistants, may easily intercept and otherwise process such communication. Providers of these services do not fall under the strict regime of Article 5 (1) ePD, which regulates the confidentiality of communications. This creates a significant gap in legal protection and outlines that the ePD is not fit for purpose to ensure the confidentiality of both interpersonal and human-machine communication.

Likewise, the requirement to obtain consent for the storage of information or gaining access to information already stored in the terminal equipment of a subscriber or user according to Article 5 (3) ePD as introduced in Section 3.4.3.2 is likely not applicable to virtual assistant services. With virtual assistants, the information (e.g., voice recordings) is *not* stored on the terminal equipment, nor does the service gain access to information stored on the terminal equipment. This holds true regardless of whether the virtual assistant service is embedded in a smartphone or in a smart home device such as 'Amazon Echo'. Rather, information is stored and otherwise processed within the service platform of the provider, namely, in the cloud.[1417] Regulatory guidance neglects the technical functioning of virtual assistant services such as Amazon Alexa when stating that 'consent as required by Article 5 (3) ePD would be necessary for the storing or gaining of access to information for any purpose other than executing a user request (e.g., user profiling)'.[1418] Leading virtual assistants do not store the voice recording of their users on the terminal equipment, but rather on the service platform of the provider, mostly in the cloud. [1419] Major providers of virtual assistants (e.g. Amazon, Apple, Google, Cortana) rely on cloud environments to store data processed in the context of virtual assistants.[1420] This is

[1416] Recitals 2, 5 ePD.
[1417] Daniel J Dubois et al, 'When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers' (2020) Iss 4 Proceedings on Privacy Enhancing Technologies 255; Lea Schönherr et al, 'Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers' (2020) at 1 < https://arxiv.org/pdf/2008.00508.pdf > accessed 8 February 2024.
[1418] European Data Protection Board, 'Guidelines 02/2021 on Virtual Voice Assistants' (16 May 2011) at 29.
[1419] Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3; Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8.
[1420] Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8.

different when compared with another 'always-on' service, namely, Amazon's wearable 'Halo', which analyses voice tones to detect user emotions.[1421] According to Amazon, the recordings are never uploaded to the cloud but instead analysed on the user's device and then deleted.[1422]

For most virtual assistants, the computing performed locally on the device focusses on listening for a wake word[1423] and sampling subsequent audio information for transportation to the cloud.[1424] Computing necessary for automatic speech recognition, natural language understanding, natural language generation and ultimately speech generation[1425] are thus *not* performed or stored locally on the device used by the virtual assistant service. Most virtual assistants do not require storing information or accessing information on the user's device. Rather, by uttering the voice command, the user initiates the streaming of the voice recordings to the servers of the provider *via* the device. This does not mean that the provider retrieves the voice recording *from* the device or gains access to voice recordings *stored on* the device of the user.[1426] Moreover, virtual assistants are software applications[1427] consisting of several components[1428] and layers. They are, as such, *not* terminal equipment as referred to in Article 5 (3) ePD. Like any other software, virtual assistants rely on hardware in order to function, for example, devices like computers, smartphones, tablets or on purpose-built speaker devices.[1429] When activated by the voice command of the user, the device usually sends the speech recording directly to the service platform of the provider where it is subsequently stored.[1430] Hence, the device *solely* opens a stream to the cloud[1431] but does not store the voice recording (e.g., voice command).

[1421] Austin Carr, 'Amazon's New Wearable Will Know If I'm Angry. Is That Weird?' *Bloomberg* (New York, 31 August 2020) < https://www.bloomberg.com/news/newsletters/2020-08-31/amazon-s-halo-wearable-can-read-emotions-is-that-too-weird > accessed 8 February 2024.
[1422] Alex Hern, 'Amazon's Halo wristband: the fitness tracker that listens to your mood' *The Guardian* (London, 28 August 2020) < https://www.theguardian.com/technology/2020/aug/28/amazons-halo-wristband-the-fitness-tracker-that-listens-to-your-mood > accessed 8 February 2024.
[1423] Lea Schönherr et al, 'Unacceptable, where is my privacy? Exploring Accidental Triggers of Smart Speakers' (2020) at 2 < https://arxiv.org/pdf/2008.00508.pdf > accessed 8 February 2024.
[1424] Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3; Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 11.
[1425] Roberto Pieraccini, *AI Assistants* (MIT Press 2021) 8.
[1426] Centre for Information Policy Leadership, 'Comments by the Centre for Information Policy Leadership on the EDPB Guidelines 02/2021 on Virtual Voice Assistants' (2021) 5 < https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_comments_on_the_edpb_guidelines_on_virtual_voice_assistants__23_april_2021_.pdf > accessed 8 February 2024.
[1427] Matthew B Hoy, 'Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants' (2018) Vol 37 No 1 Medical Reference Services Quarterly 81, 82; Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1.
[1428] Roberto Pieraccini, *AI Assistants* (MIT Press 2021) 7.
[1429] Matthew B Hoy, 'Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants' (2018) Vol 37 No 1 Medical Reference Services Quarterly 81, 82; Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1.
[1430] Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3; Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8.
[1431] Centre for Information Policy Leadership, 'Comments by the Centre for Information Policy Leadership on the EDPB Guidelines 02/2021 on Virtual Voice Assistants' (2021) 5 < https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_comments_on_the_edpb_guidelines_on_virtual_voice_assistants__23_april_2021_.pdf > accessed 8 February 2024.

In addition, many virtual assistants are designed as distributed web services with application services provided by different companies, organisations and developers,[1432] which indicates significant sharing of data. Furthermore, it is unclear whether the user in fact can be aware of what actually is being stored[1433] and there seems to be no accurate mechanism for the user to exercise control regarding the sharing of such stored data.[1434] Providers of virtual assistant services, along with the different actors involved in providing the service, can further process the recorded speech of their users and other data to infer a rich array of personal information without the need to obtain consent from the users. Such information includes clues about the user's biometric identity, personality, physical traits, geographical origin, level of intoxication and sleepiness, age, gender, health condition and even an individual's socioeconomic status.[1435]

Regulatory guidance implies that processing in the context of virtual assistants occurs locally on the device[1436] and that providers of virtual assistant services gain access to information stored on the user's device. However, this is not correct. The actual speech recordings, namely, the command given to the virtual assistant, is directly transmitted to the platform of the provider. Further processing, as well as storage, occurs there.[1437] For this reason, it cannot be concluded that Article 5 (3) ePD is applicable. Virtual assistant services do not store information, nor gain access to information already *stored*, *in the terminal equipment*, as is the case with cookies. Major providers of these services, such as Amazon, Apple and Google store data processed in the context of virtual assistants in the cloud, not on the device.[1438] Moreover, providers can also link speech data with other datasets (e.g. social media meta data, browsing behaviour, purchase histories) in order to draw further sensitive inferences.[1439] As explained in the communication surveillance problem, there is a loophole in the current legal framework that specifically ensures the confidentiality of human-machine communication[1440] and

[1432] Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8; Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3.
[1433] Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1, 16.
[1434] Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8; Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1, 8.
[1435] Jacob Leon Kröger, Otto Hans-Martin Lutz, Philip Raschke, 'Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference' in Michael Friedewald et al (eds) *Privacy and Identity management. Data for Better Living: AI and Privacy* (Springer 2020) 242.
[1436] European Data Protection Board, 'Guidelines 02/2021 on Virtual Voice Assistants' (16 May 2011) at 16, 29.
[1437] This is acknowledged by regulatory guidance, which also remarks that Article 5 (3) ePrivacy Directive might need to be amended in the future. See European Data Protection Board, 'Guidelines 02/2021 on Virtual Voice Assistants' (16 May 2011) at 16 and Footnote 12 on page 12; see also Tom Bolton et al, 'On the Security and Privacy Challenges of Virtual Assistants' (2021) Vol 21 Iss 7 Sensors 1-3; Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8.
[1438] Allan de Barcelos Silva et al, 'Intelligent personal assistants: A systematic literature review' (2020) Vol 147 Expert Systems With Applications 1, 8.
[1439] Jacob Leon Kröger, Otto Hans-Martin Lutz, Philip Raschke, 'Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference' in Michael Friedewald et al (eds) *Privacy and Identity management. Data for Better Living: AI and Privacy* (Springer 2020) 2523.
[1440] However, the GDPR regulates human-machine communications provided this relates to the processing of personal data.

prevents surveillance as well as further processing thereof without consent.[1441] Therefore, Article 5 (3) ePD is not fit for purpose to protect the fundamental rights to privacy and data protection, in particular the privacy of communications, when considering the gap of protection it creates with regard to human-machine communications. It also fails to meet the ePD's legislative aims to guarantee the confidentiality of communications,[1442] to protect natural persons concerning the automated storage and processing of data[1443] and ultimately to protect personal data and the privacy of users.[1444] This constitutes a Type 3 legal problem regarding the fundamental rights to privacy and data protection. As indicated in the communication surveillance problem, this lacuna in the current legal framework does not solely apply to human-machine communication, but also to sensitive *interpersonal communications*, including conversations about health status, domestic violence, sexual relationships, drug deals,[1445] discussions between doctors and patients and business deals.[1446]

The storage problem (Type 3)

ML, NLP and AC facilitate the provision of virtual assistant services. Providers may analyse and otherwise process human-machine and interpersonal communication without needing to obtain consent from the user. Article 5 (3) ePD does not apply to virtual assistant services as they do not store information, or gain access to information already stored, in the device of the user. This provision is not fit for purpose to protect the fundamental rights to privacy and data protection because it creates a significant gap of protection as processing of both human-machine and interpersonal communication may reveal sensitive information and that likely is to be shared with various actors.

4.10 Conclusions

Chapter 4 aimed to answer Subquestion 3, namely, what legal problems arise or may arise when the principles enshrined in the current legal framework are applied to AI. In this chapter, I have outlined that all AI disciplines as described in Section 2.2 raise or may raise legal problems when they are applied to the principles enshrined in the current legal framework as introduced in Chapter 3. Three types of legal problems were identified: (1) legal provisions that are violated, (2) legal provisions that cannot be enforced and (3) legal provisions that are not fit for purpose to protect the fundamental right at stake. These legal problems may be caused by AI disciplines *or* by the principles themselves when they are applied in the context of AI. Table 4.3 provides an overview of the legal problems identified in this chapter.

[1441] For instance, as it is the case with consent for cookies as required by Article 5 (3) ePrivacy Directive.
[1442] Recital 3 ePD.
[1443] Recital 7 ePD.
[1444] Recitals 2, 5 ePD.
[1445] Tine Munk, 'Does Online Privacy Exist in the GDPR Era? The Google Voice Assistant Case' in Tatiana-Eleni Synodiou et al (eds) *EU Internet Law in the Digital Single Market* (Springer Nature 2021) 497.
[1446] Alex Hern, 'Apple contractors regularly hear confidential details on Siri recordings' *The Guardian* (London, 26 July 2019) < https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings > accessed 8 February 2024.

Problem	Principles	Type	AI Disciplines
Balancing	Lawfulness, Proportionality, Accountability	1	AR
Probability	Fairness, Accountability	1	ML, AR
Facial recognition	Fairness, Accountability	1	CV
Inaccuracy	Fairness, Accuracy, Accountability	1	ML, AC
Sensitivity	Fairness, Accountability	1	AC
Elusiveness	Fairness	2, 3	ML, NLP, CV, AC, AR
Manipulation	Fairness	3	ML, AC
Sabotage	Fairness	3	ML, NLP, CV, AC, AR
Opacity	Transparency, Accountability	1	ML, NLP, CV, AC, AR
Interpretability	Transparency	1, 2	ML
Inference	Transparency	3	ML, AC
Profiling	Transparency	3	ML, AC
Inexplicitness	Purpose limitation, Accountability	1	ML, NLP, CV, AC, AR
Function creep	Purpose limitation, Accountability	1	ML, NLP, CV, AC, AR
Restriction	Purpose limitation	3	ML, NLP, CV, AC, AR
Compatible use	Purpose limitation	3	ML
Data appetite	Data minimisation, Accountability	1	ML, NLP, CV, AC, AR
Necessity	Data minimisation, Accountability	1	ML
Verification	Data minimisation	2	ML, NLP, CV
Trade-off	Data minimisation, Accuracy, Fairness	3	ML, NLP, CV, AC, AR
Rebuttal	Accuracy, Accountability	1	AC
Common sense	Accuracy, Accountability	1	AR
Guidance	Accuracy	2, 3	ML, NLP, CV, AC, AR
Incomputability	Accuracy, Fairness	3	ML, NLP, CV, AC, AR
Emotion data	Enhanced protection for special data	3	AC
Location data	Enhanced protection for special data	3	ML
Neurodata	Enhanced protection for special data	3	ML (DL), CV, NLP
Mental data	Enhanced protection for special data	3	ML, AC
Communication surveillance	Confidentiality	3	ML, NLP, AC
Storage	Confidentiality	3	ML, NLP, AC

Table 4.3 Overview of legal problems related to the principles contained in the legal framework. The brackets surrounding DL indicate that this *specific kind* of ML causes the legal problem in question.

Table 4.3 illustrates the broad range of legal problems that arise or may arise in the context of AI. In total, 30 problems are identified.

The *lawfulness principle* does not appear to be particularly problematic when applied to AI; after all, only one legal problem relates to this principle. The reason for this is that the meaning of the lawfulness principle is substantively clear, as is further substantiated in Article 6 GDPR, which enumerates six lawful bases that can be relied upon for the processing of personal data. According to Table 4.3, only the AI discipline *AR* causes a Type 1 legal problem when applied to the lawfulness principle.

The *fairness principle* causes the most legal problems when applied to AI: 10 out of 30 problems relate to the fairness principle. In addition, it causes all three types of legal problems. When interpreted as substantive fairness to prevent adverse effects on data subjects, the principle of fairness can be violated by processing facilitated by AI. This is underscored by the fact that four legal problems regarding the fairness principle relate to Type 1 problems. The main issue with the fairness principle lies in its elusive meaning. The substantively unclear meaning of the fairness principle reduces legal certainty and makes it less likely that it will be enforced by individuals or regulators (Type 2 problem). Ultimately, this also leads to Type 3 legal problems because a substantively unclear principle is not fit for purpose to protect the fundamental right to data protection. *ML* and *AC* seem to be the most problematic AI disciplines when applied to the fairness principle.

Regarding the *transparency* principle, I have identified four legal problems of either Type 1, 2 or 3 when applied to AI. Because AI systems may be rather ubiquitous, all AI disciplines potentially clash with the transparency principle. Nevertheless, *ML* is the main driver for legal problems relating to the transparency principle: all the four legal problems relate to this AI discipline. This is mainly caused by the fact that ML is widely used to infer and derive data from existing data and because AI systems deploying DL and ANN approaches are likely to produce noninterpretable outputs.

Regarding the *purpose limitation* principle, I have identified four legal problems of either Type 1 or Type 3 when applied to AI. Generally, all AI disciplines process personal data from various sources for a plethora of purposes and are therefore in conflict with the purpose limitation principle. ML serves as a typical example: unsupervised ML processes personal data for unspecific and inexplicit purposes. Thus, the processing itself determines the purpose and future use of personal data, which causes Type 1 legal problems. The purpose limitation principle also causes Type 3 legal problems when applied to AI because it does not restrict the processing of personal data and allows further processing for compatible purposes.

Regarding the *data minimisation* principle, I have identified four legal problems of Type 1, 2 or 3 when applied to AI. Type 1 problems are mainly caused by the data appetite of AI – regardless of which discipline of AI is used to process personal data. When consequently applied in the context of AI, the data minimisation principle may create trade-offs regarding the accuracy and fairness principles, which leads to Type 3 legal problems. Here as well, *ML* is the main driver for the legal problems

with respect to the data minimisation principle: all four legal problems are caused by this single AI discipline.

Regarding the *accuracy* principle, I have identified six legal problems of Type 1, 2 or 3 when applied to AI. The main issue with the accuracy principle is caused by the fact that the required level of accuracy depends on the purpose of the processing, as suggested by relevant case law ('relative accuracy'). Such relative accuracy does not outline specific levels of accuracy that personal data processed in the context of AI must reach: there is no one-size-fits-all approach. Thus, the precise substantive requirements of the accuracy principle remain an underexplored topic in academia and case law, which is highly problematic when considering the developments in AI, causing Type 2 and 3 legal problems. *ML* and *AC* are the most problematic AI disciplines because they are likely to generate inaccurate personal data.

Regarding the principle of *enhancing protection for special categories of personal data*, I have identified four Type 3 legal problems when applied to AI. The main issue of this principle is caused by the legislators' approach to exhaustively enumerate special data in Article 9 GDPR. This exhaustive list of special personal data contained in the GDPR does not keep up with the technological developments facilitated by AI. The stringent rules concerning the processing of sensitive data do not apply to the processing of new types of sensitive personal data facilitated by AI, such as emotion data, neurodata and mental data. As apparent from Table 4.3, *ML, NLP* and *AC* are the most problematic AI disciplines in the context of this principle.

Regarding the *confidentiality of communications* principle, I have identified two Type 3 legal problems when applied to AI. Due to the restricted material scope of the ePD, the prohibition of listening, tapping, storage or other kinds of interception or surveillance does not apply in the context of virtual assistants and smart home technologies which are powered by the AI disciplines ML, NLP and AC. Likewise, the ePD does not require providers of virtual assistant services to obtain consent from their users in order to analyse and otherwise process human-machine communication because virtual assistant services typically do not store information, or gain access to information already stored, in the device of the user as required by the ePD. As apparent from Table 4.3, *ML, NLP* and *AC* are the most problematic AI disciplines in the context of this principle.

In terms of the *types of legal problems* caused by AI, Table 4.3 shows that 14 out of 30 legal problems identified within this chapter relate to *Type 3* legal problems. Thus, there is a clear mismatch between the principles enshrined in the current legal framework and the AI disciplines introduced in Chapter

2.[1447] This means that legislative measures may be needed to address said mismatch. Furthermore, almost half of the problems relate to *Type 1* legal problems. Thus, AI is likely to violate the principles enshrined in the current legal framework. *Type 2* legal problems seem to be rare: only four legal problems identified within this chapter relate to the enforcement of the provisions enshrined in the current legal framework. Therefore, more enforcement seems to be needed, both with respect to private enforcement initiated by data subjects or representative bodies and with respect to regulatory enforcement pursued by SAs.

In terms of which AI disciplines cause *how many legal problems* when applied to the principles enshrined in the current legal framework, Table 4.3 shows that ML leads to twenty-four, NLP fourteen, CV thirteen, AC nineteen and AR thirteen legal problems, respectively. The prominent role of ML is not surprising, as this AI discipline is the most widely used and often combined with other AI disciplines. In addition, AC seems to be the main driver of legal problems which only causes slightly less legal problems when compared to ML. The amounts of legal problems associated to the AI disciplines NLP, CV and AR are distributed almost equally.

[1447] This is in line with other research, e.g. Tal Z Zarsky 'Incompatible: The GDPR in the Age of Big Data' (2017) Vol 47 Iss 4 Seton Hall Law Review 995-1020.