# 7

# AI Meets the GDPR

## *Navigating the Impact of Data Protection on AI Systems*

*Pierre Dewitte*

### 7.1 INTRODUCTION

To state that artificial intelligence ("AI") has seen drastic improvements since the age of expert systems is rather euphemistic at a time when language models have become so powerful they could have authored this piece – hint, they didn't. If, conceptually speaking, AI systems refer to the ability of a software to mimic the features of human-like reasoning, most are used to draw predictions from data through the use of a trained model, that is, an algorithm able to detect patterns in data it has never encountered before. When such models are used to derive information relating to individuals, personal data are likely involved somewhere in the process, whether at the training or deployment stage. This can certainly result in many benefits for those individuals. However, as abundantly illustrated throughout this book, the link between personal information and natural persons also exposes them to real-life adverse consequences such as social exclusion, discrimination, identity theft or reputational damage, all the while directly contributing to the opacification of the decision-making processes that impact their daily lives. For all these reasons, specific legal guarantees have been adopted at various levels to minimize these risks by regulating the processing of personal data and equipping individuals with the appropriate tools to understand and challenge the output of AI systems.

In Europe, the General Data Protection Regulation ("GDPR")[1] is the flagship piece of legislation in that regard, designed to ensure both the protection of natural persons and the free movement of personal data. Reconciling the intrinsic characteristics of AI systems with the principles and rules contained therein is a delicate exercise, though. For two reasons. First, the GDPR has been conceived as a technology-neutral instrument comprised of voluntarily open-ended provisions meant to carry their normative values regardless of the technological environment

[1] Regulation 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 ELI: http://data.europa.eu/eli/reg/2016/679/oj.



https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

they are applied in.[2] Such is the tradeoff necessary to ensure resilience and future-proofness when technological progresses have largely outpaced the capacity of regulators to keep up with the unbridled rhythm of innovation.[3] In turn, navigating that ecosystem comprised of multiple layers of regulation designed to reconcile flexibility and legal certainty can prove particularly daunting. Second, AI systems have grown more and more complex, to the point where the opacity of their reasoning process has become a common ground for concern.[4] This reinforces the need for interdisciplinary collaboration, as the proper understanding of their functioning is essential for the correct application of the law. In short, regulating the processing of personal data in AI systems requires to interpret and apply a malleable regulatory framework to increasingly complex technological constructs. This, in itself, is a balancing act between protecting individuals' fundamental rights and guaranteeing a healthy environment for innovation to thrive.

The purpose of this chapter is not to provide a comprehensive overview of the implications of the GDPR for AI systems. Nor is it to propose concrete solutions to specific problems arising in that context.[5] Rather, it aims to walk the reader though the core concepts of EU data protection law, and highlight the main tensions between its principles and the functioning of AI systems. With that goal in mind, Section 7.2 first sketches the broader picture of the European privacy and data protection regulatory framework, and clarifies the focus for the remainder of this chapter. Section 7.3 then proceeds to delineate the scope of application of the GDPR and its relevance for AI systems. Finally, Section 7.4 breaks down the main friction

[2] This is recalled in Recital 15 GDPR: "In order to prevent creating a serious risk of circumvention, the protection of natural persons should be technologically neutral and should not depend on the techniques used."

[3] This is most commonly referred to as the "pacing problem" of the law. See Roger Brownsword, *Rights, Regulation, and the Technological Revolution* (Oxford University Press, 2008); Larry Downes, *The Laws of Disruption: Harnessing the New Forces That Govern Life and Business in the Digital Age* (Basic Books, 2009); Gary E Marchant, "The growing gap between emerging technologies and the law" in Gary E Marchant, Braden R Allenby, and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight*, vol. 7 (Springer Netherlands, 2011) 20–22, http://link.springer.com/10.1007/978-94-007-1356-7_2, accessed December 4, 2019.

[4] For instance, in the context of predictive policing, where algorithms are used to assess the likelihood of defendants becoming recidivists. See ProPublica's analysis of the COMPAS algorithm used by US courts: Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias – There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks" *ProPublica* (May 23, 2016), www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed January 14, 2023. Their calculation is also available on GitHub at the following address: https://github.com/propublica/compas-analysis.

[5] For that, I redirect the reader to dedicated reference manuscripts and studies such as, among many others: Dara Hallinan, Ronald Leenes, and Paul De Hert (eds), *Data Protection and Privacy: Data Protection and Artificial Intelligence* (Hart Publishing, 2021); Giovanni Sartor and Francesca Lagioia, "The impact of the general data protection regulation (GDPR) on artificial intelligence" (European Parliamentary Research Service, 2020) Think Tank: European Parliament, Study, www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530, accessed January 11, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

points between the former and the latter and illustrates each of these with examples of concrete data protection challenges raised by AI systems in practice.

## 7.2 SETTING THE SCENE – THE SOURCES OF PRIVACY AND DATA PROTECTION LAW IN EUROPE

While the GDPR is the usual suspect when discussing European data protection law, it is but one piece of a broader regulatory puzzle. Before delving into its content, it is therefore crucial to understand its position and role within that larger ecosystem. Not only will this help clarify the different sources of privacy and data protection law, but it will also equip the reader with keys to understand the interaction between these texts. The goal of this section is hence to contextualize the GDPR in order to highlight its position within the hierarchy of legal norms.

In Europe, two coexisting legal systems regulate the processing of personal data.[6] First, that of the Council of Europe ("CoE") through Article 8 of the European Convention on Human Rights ("ECHR")[7] as interpreted by the European Court of Human Rights ("ECtHR").[8] Second, that of the European Union ("EU") through Articles 7 and 8 of the Charter of Fundamental Rights of the European Union ("CFREU")[9] as interpreted by the Court of Justice of the European Union ("CJEU").[10] While these systems differ in scope and functioning, the protection afforded to personal data is largely aligned as the case law from both Courts influence each other.[11] National legislation constitutes an extra layer of privacy and data protection law, bringing the amount of regulatory silos up to three (see Figure 7.1).

[6] Juliane Kokott and Christoph Sobotta, "The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR" (2013) *International Data Privacy Law*, 3: 222, 222–223. See, for further information on these two systems: European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Data Protection Law – 2018 Edition* (2018), https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law-2018-edition, accessed January 16, 2023.

[7] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended by Protocols n° 11, 14, and 15 and supplemented by Protocols n° 6, 7, 12, 13, and 16).

[8] An overview of the jurisprudence of the ECtHR on Article 8 is available here: Registry of the European Court of Human Rights, "Guide on Article 8 of the European Convention on Human Rights. Right to respect for private and family life, home and correspondence" (April 9, 2024), https://ks.echr.coe.int/documents/d/echr-ks/guide_art_8_eng, accessed July 30, 2024.

[9] Charter of Fundamental Rights of the European Union, O.J.E.U., December 18, 2000, C 364/01.

[10] See, for an overview of the main relevant cases: Research and Documentation Directorate, "Fact Sheet: Protection of Personal Data" (Court of Justice of the European Union, 2021), https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-10/fiche_thematique_-_donnees_personnelles_-_en.pdf, accessed January 16, 2023.

[11] More specifically, Article 52(3) CFREU states that "*in so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention*."

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

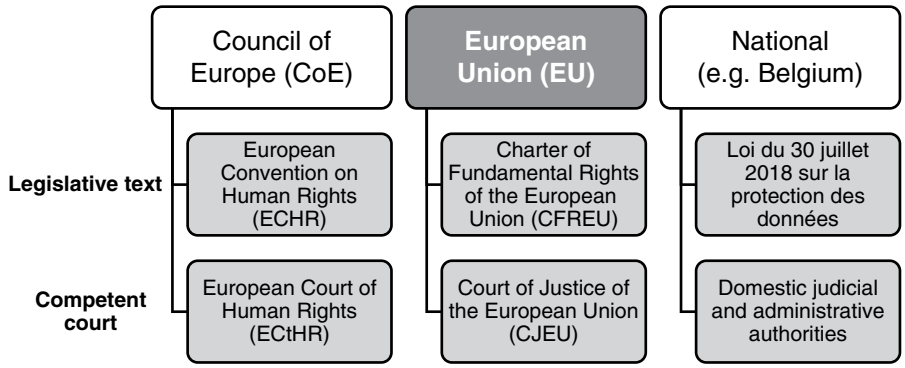



FIGURE 7.1 A fundamental rights perspective on the sources of privacy and data protection law

For the purpose of this chapter, let's zoom in on the EU legal order, comprised of primary and secondary legislation. While the former sets the foundational principles and objectives of the EU, the latter breaks them down into actionable rules that can then be directly applied or transposed by Member States into national law. This is further supplemented by "soft law" instruments issued by a wide variety of bodies to help interpret the provisions of EU. While these are not strictly binding, they often have quasi-legislative authority.[12] As illustrated in Figure 7.2, the GDPR is only *a* piece of secondary EU law meant to protect all data subjects' fundamental rights – including but not limited to privacy and data protection – when it comes to the processing of their personal data. As illustrated in the following sections, the Guidelines issued by the Article 29 Working Party ("WP29") and its successor the European Data Protection Board ("EDPB") are particularly helpful when fleshing out the scope and substance of the rules contained in the GDPR.[13] While all three of the silos detailed above impact – to a certain extent – the processing of personal data by AI systems, the remainder of this chapter focuses exclusively on the EU legal order, more specifically on the GDPR and its accompanying soft law instruments.

[12] See, for an overview of the GDPR soft law ecosystem and its limitations: Athena Christofi, Pierre Dewitte, and Charlotte Ducuing, "Erosion by standardisation: Is ISO/IEC 29134:2017 on privacy impact assessment up to (GDPR) standard?" in Maria Tzanou (ed), *Personal Data Protection and Legal Developments in the European Union* (IGI Global, 2020) 145–148, http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-9489-5, accessed January 16, 2023.

[13] The Article 29 Working Party (WP29) and its successor the European Data Protection Board (EDPB) are independent EU bodies composed of representative from national supervisory authorities tasked with ensuring the consistent interpretation of the GDPR throughout the Union. More specifically, the Board now plays a central role in the cooperation and consistency mechanism outlined in Chapter VII GDPR by issuing the so-called "binding decisions" in cases where national supervisory authorities disagree on substance of a draft decision (Article 65(1)a GDPR). The duties of the Board are detailed in Article 70 GDPR.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

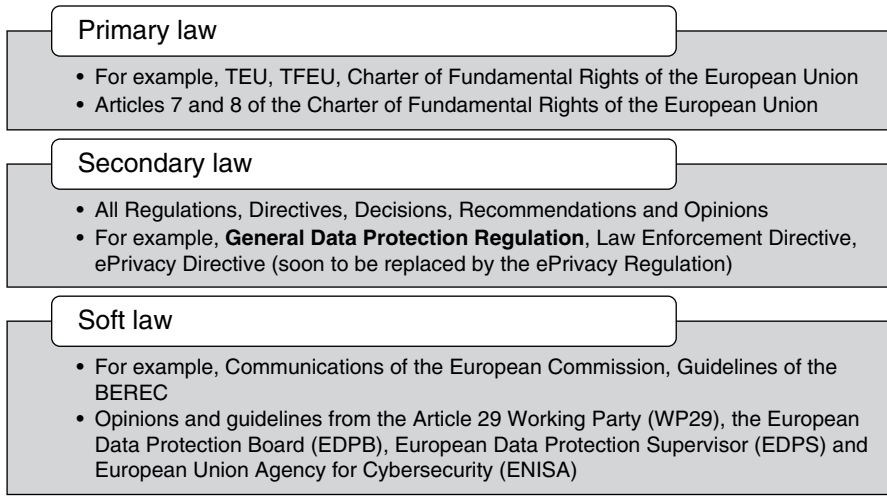



FIGURE 7.2 The EU legal order – general and data protection specific

## 7.3 OF PERSONAL DATA, CONTROLLERS AND PROCESSORS – THE APPLICABILITY OF THE GDPR TO AI SYSTEMS

As hinted at earlier, the GDPR is likely to come into play when AI systems are trained and used to make predictions about natural persons. Turning that intuition into a certainty nonetheless requires a careful analysis of its precise scope of application. In fact, this is the very first reflex anyone should adopt when confronted to *any* piece of legislation, as it typically only regulates certain types of activities (i.e., its "material scope") by imposing rules on certain categories of actors (i.e., its "personal scope"). Should the situation at hand fall outside the remit of the law, there is simply no need to delve into its content. Before discussing the concrete impact of the GDPR on AI systems in Section 7.4, it is therefore crucial to clarify whether (Section 7.3.1) and to whom it applies (Section 7.3.2).

### 7.3.1 *Material Scope of Application – The Processing of Personal Data*

#### 7.3.1.1 The Notion of Personal Data and the Legal Test of Identifiability

Article 2(1) GDPR limits the applicability of the Regulation "to the processing of personal data wholly or partly by automated means." Equally important, Article 4(1) defines the concept of personal data as "any information relating to an identified or identifiable natural person." The reference to "any information" implies that the qualification as personal data is nature-, content-, and format-agnostic,[14] while

[14] See the examples in: Lee A Bygrave and Luca Tosoni, "Article 4(1). Personal data" in Christopher Kuner et al. (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (Oxford University Press, 2020) 109–110, https://doi.org/10.1093/oso/9780198826491.003.0007, accessed January 17, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

"relating to" must be read as "linked to a particular person."[15] As such, the notion of personal data is not restricted to "information that is sensitive or private, but encompasses all kinds of information, not only objective but also subjective, in the form of opinions or assessments."[16] The term "natural persons," then, refers to human beings, thereby excluding information relating to legal entities, deceased persons, and unborn children from the scope of protection of the Regulation.[17]

The pivotal – and most controversial – element of that definition is the notion of "identified or identifiable." According to the WP29's Opinion 4/2007, a person is "identified" when "within a group of persons, he or she is 'distinguished' from all other members of the group." This can be the case when that piece of information is associated with a name, but any other indirect identifier or combination thereof, such as a telephone number or a social security number, might also lead to the identification of that individual. A person is "identifiable" when, "although he or she has not been identified yet, it is possible to do so."[18] "To determine whether a natural person is identifiable," states Recital 26 GDPR, "account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly." In turn, "to ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments." This makes the qualification of "personal data" a dynamic, context-sensitive assessment that calls for a case-by-case analysis of the reidentification potential.

Such an assessment was conducted by the CJEU in the *Breyer* case,[19] in which it held that a dynamic IP address collected by a content provider was to be considered as a piece of personal data, even though that provider was not able, by itself, to link the IP address back to a particular individual. German law indeed allowed content providers, in the context of criminal proceedings following cyberattacks for instance, to obtain from the internet service provider the information

[15] C-434/16 *Nowak* [2017] ECLI:EU:C:2017:994, para 35.

[16] Ibid., para 34. In that case, the CJEU held that the written answers submitted by a candidate at a professional examination as well as any comments made by an examiner with respect to those answers constitute personal data, within the meaning of Article 4(1) GDPR.

[17] On post-mortem privacy, see: Edina Harbinja, "Post-mortem privacy 2.0: Theory, law, and technology" (2017) *International Review of Law, Computers & Technology*, 31: 26. The author offers a deeper analysis of these issues in her doctoral thesis: Edina Harbinja, "Legal Aspects of Transmission of Digital Assets on Death" (University of Strathclyde, Law School, 2017), https://scholar.archive.org/work/owjux2fhlbbjnkiar2tfiowkki/access/wayback/https://stax.strath.ac.uk/downloads/pz50gw38v, accessed May 16, 2023.

[18] Article 29 Working Party, "Opinion 4/2007 on the concept of personal data" 12, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf, accessed January 16, 2023.

[19] C-582/14, *Patrick Breyer* v *Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779, para 49.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

necessary to turn that dynamic IP address back to its static form, and therefore link it to an individual user. That means of reidentification was considered "reasonably likely" to be used, thereby falling under the scope of Article 4(1) read in combination with Recital 26 GDPR. On the contrary, that likelihood test would not have been met if such reidentification was "prohibited by law or practically impossible on account of the fact that it requires disproportionate efforts in terms of time, cost, and workforce, so that the risk of identification appears in reality to be insignificant."[20] By investigating the actual means of reidentification at the disposal of the content provider to reidentify the data subject to whom the dynamic IP address belonged, the Court embraced a "risk-based" approach to the notion of personal data, as widely supported in legal literature and discussed in Section 7.4.3.[21]

Data for which the likelihood of reidentification falls below that "reasonable" threshold are considered "anonymous" and are not subject to the GDPR. Lowering the risk of reidentification to meet the GDPR standard of anonymity is no small feat, however, and depends on multiple factors such as the size and diversity of the dataset, the categories of information it contains, and the effectiveness of the techniques applied to reduce the chances of reidentification.[22] For instance, swapping names for randomly generated number-based identifiers might not be sufficient to reasonably exclude the risk of reidentification if the dataset at stake is limited to the employees of a company paired with specific categories of data such as hobbies, gender, or device fingerprints. In that case, singling someone out, linking two records, or deducing the value of an attribute based on other attributes – in this example, the name of a person based on a unique combination of the gender and hobbies – remains possible. For the same reason, hashing the license plate of a car entering a parking before storing it into the payment system, even when the hash function used is strictly nonreversible, might not reasonably shield the driver from reidentification if the hash value is stored alongside other information such as the time of arrival or departure, which might later be combined with unblurred CCTV

[20] Ibid., para 46.

[21] Michèle Finck and Frank Pallas, "They who must not be identified – distinguishing personal from nonpersonal data under the GDPR" (2020) *International Data Privacy Law*, 10(11): 34–36; Daniel Groos and Evert-Ben van Veen, "Anonymised data and the rule of law" (2020) *European Data Protection Law Review*, 6(498): 5; Sophie Stalla-Bourdillon, "Anonymising personal data: Where do we stand now?" (2019) *Privacy & Data Protection*, 19(3): 3–5.

[22] For examples of anonymization techniques and their robustness, see Article 29 Working Party, "Opinion 05/2014 on Anonymisation Techniques," 11–19, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, accessed January 16, 2023. It is worth noting that these guidelines, which have been abundantly criticized in legal literature for their extremely strict understanding of anonymization, are being revised as the time of writing. See Finck and Pallas (n 21) 15; Sophie Stalla-Bourdillon, "Anonymous data v. personal data – false debate: An EU perspective on anonymization, pseudonymization and personal data" (2016) *Wisconsin International Law Journal*, 34(384): 306–320.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

footages to retrieve the actual plate number.[23] These techniques are therefore considered as "pseudonymization" rather than "anonymization,"[24] with the resulting "pseudonymized data" falling under the scope of the GDPR in the same way as regular personal data. As detailed in Section 7.4.3, pseudonymization techniques nonetheless play a critical role as mitigation strategies in the risk-based ecosystem of the Regulation.[25]

#### 7.3.1.2 The Processing of Personal Data in AI Systems

AI systems, and more specifically machine learning algorithms, process data at different stages, each of which is likely to involve information that qualifies as personal data. The first of these is the training stage, if the target and predictor variables are sufficiently granular to allow a third party to reidentify the individuals included in the training dataset.[26] This could be the case, for instance, when training a model to detect tax fraud based on taxpayers' basic demographic data, current occupation, life history, income, or previous tax returns, the intimate nature of which increases the risk of reidentification. Anonymization – or pseudonymization, depending on the residual risk – techniques can be used to randomize variables by adding noise (e.g., replacing the exact income of each taxpayer by a different yet comparable amount) or permutating some of them (e.g., randomly swapping the occupation of two taxpayers).[27] Generalization techniques such as *k*-anonymity (i.e., ensuring that the dataset contains at least *k*-records of taxpayers with identical predictors by decreasing their granularity, such as replacing the exact age with a range) or *l*-diversity

[23] Agencia Española de Protección de Datos and European Data Protection Supervisor, "Introduction to the hash function as a personal data pseudonymisation technique" (October 2019), https://edps.europa.eu/sites/default/files/publication/19-10-30_aepd-edps_paper_hash_final_en.pdf, accessed January 16, 2023.

[24] Defined in Article 4(5) GDPR as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

[25] For an overview of the state of the art on pseudonymization, see European Union Agency for Cybersecurity, "Data pseudonymisation: Advanced techniques and use cases," www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases, accessed January 16, 2023.

[26] The target variable being the variable that the model, once trained, will be able to predict, and the predictor variables being the information on the basis of which the model will ground its prediction. For a simplified overview of the functioning of supervised and unsupervised machine learning, see Datatilsynet, "Artificial intelligence and privacy," 7–14, www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf, accessed January 11, 2023.

[27] The Information Commissioner's Office, UK's supervisory authority, provides a solid introduction to anonymization techniques in: Information Commissioner's Office, "Anonymisation: Managing data protection risk code of practice." See also: Information Commissioner's Office, "Big data, artificial intelligence, machine learning and data protection," paras 130–138, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf, accessed January 18, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

(i.e., extending *k*-anonymity to make sure that the variables in each set of *k*-records have at least *l*-different values) are also widely used in practice. Synthetic data, namely artificial data that do not relate to real individuals but are produced using generative modeling, can serve as an alternative to actual, real-life personal data to train machine learning models.[28] Yet, doing so is only a workaround, as the underlying generative model also needs to be trained on personal data. Plus, the generated data might reveal information about the natural persons who were included in the training dataset in cases where one or more specific variable stand out.

Second, a trained machine learning model might leak some of the personal data included in the training dataset. Some models might be susceptible to model inversion or membership inference attacks, which respectively allow an entity that already knows some of the characteristics of the individuals who were part of the training dataset to infer the value of other variables simply by observing the functioning of the said model, or to deduce whether a specific individual was part of that training dataset.[29] Other models might leak by design.[30] The qualification of trained models as personal – even if pseudonymized – data means that the GDPR will regulate their use, as the mere sharing of these models with third parties, for instance, will be considered as a "processing" of personal data within the meaning of Article 4(2) GDPR.

As detailed in Section 7.3.1.1, the criteria used for the identifiability test of Article 4(1) lead to a broad understanding of the notion of personal data; so much so that the GDPR has been coined as the "law of everything."[31] This is especially true when it comes to the role of "the available technology" in assessing the risk of reidentification, the progress of which increases the possibility that a technique considered as proper anonymization at time $t$ is reverted and downgraded to a mere pseudonymizations method at time $t + 1$.[32] Many allegedly anonymous datasets have already been reidentified using data that were not available at the time of their

[28] For an overview of generative (adversarial) modeling, see Fida K Dankar and Mahmoud Ibrahim, "Fake it till you make it: Guidelines for effective synthetic data generation" (2021) *Applied Sciences*, 11(2158): 3–5. For a real-life example of a generative adversarial network, check the website, https://thispersondoesnotexist.com/.

[29] Michael Veale, Reuben Binns, and Lilian Edwards, "Algorithms that remember: Model inversion attacks and data protection law" (2018) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376: 20180083.

[30] Such as support vector machines and *k*-nearest neighbors algorithms, as mentioned and explained in: Information Commissioner's Office, "Guidance on AI and Data Protection," 58, https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-artificial-intelligence-and-data-protection/, accessed January 11, 2023.

[31] Nadezhda Purtova, "The law of everything. Broad concept of personal data and future of EU data protection law" (2018) *Law, Innovation and Technology*, 10: 40.

[32] Authors have even suggested that the current technological progress implies that 99.98% of Americans would be correctly reidentified in any dataset using 15 demographic attributes. See: Luc Rocher, Julien M Hendrickx, and Yves-Alexandre de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models" (2019) *Nature Communications*, 10: 1.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

release, or by more powerful computational means.[33] This mostly happens through linkage attacks, which consist in linking an anonymous dataset with auxiliary information readily available from other sources, and looking for matches between the variables contained in both datasets. AI makes these types of attacks much easier to perform, and paves the way for even more efficient reidentification techniques.[34]

### 7.3.2 *Personal Scope of Application – Controllers and Processors*

#### 7.3.2.1 The Controller–Processor Dichotomy and the Notion of Joint Control

Now that Section 7.3.1 has clarified *what* the GDPR applies to, it is crucial to determine *who* bears the burden of compliance.[35] "Controllers" are the primary addressees of the Regulation, and are responsible to comply with virtually all the principles and rules it contains. Article 4(7) defines the controller as "the natural or legal person that, alone or jointly with others, determines the purposes and means of the processing of personal data." The EDPB provides much needed clarifications on how to interpret these notions.[36] First, the reference to "natural or legal person" – in contrast with a mere reference to the former in Article 4(1) GDPR – implies that both individuals and legal entities can qualify as controllers. The capacity to "determine" then refers to "the controller's influence over the processing, by virtue of an exercise of decision making power." That influence can either stem from a legal designation, such as when national law specifically appoints a tax authority as the controller for the processing of the personal data necessary to calculate citizens' tax returns, or follow from a factual analysis. In the latter case, the EPBD emphasizes that the notion of controller is a "functional concept" meant to "allocate responsibilities according to the actual roles of the parties." It is therefore necessary to look past any existing

[33] Two examples are worth a mention. First, the linkage attack performed on mobility data that suggests that four spatiotemporal points are enough to uniquely identify 95% of individuals. See: Yves-Alexandre de Montjoye et al., "Unique in the crowd: The privacy bounds of human mobility" (2013) *Scientific Reports*, 3(1): 2. Second, the reidentification attack performed on Netflix's user ratings dataset that uncovered that six ratings are sufficient to reidentify 84% of individuals. See: Arvind Narayanan and Vitaly Shmatikov, "How to break anonymity of the Netflix Prize dataset" (arXiv, November 22, 2007) 12, http://arxiv.org/abs/cs/0610105, accessed January 18, 2023.

[34] See, for instance: Stefan Vamosi, Thomas Reutterer, and Michael Platzer, "A deep recurrent neural network approach to learn sequence similarities for user-identification" (2022) *Decision Support Systems*, 155: 113718.

[35] See, for more a more detailed overview of the allocation of responsibilities under the GDPR, the seminal work of Brendan Van Alsenoy, *Data Protection Law in the EU: Roles, Responsibilities and Liability*, vol 6 (Intersentia, 2019), www.larcier-intersentia.com/en/data-protection-law-the-eu-roles-responsibilities-liability-9781780688282.html, accessed January 16, 2023.

[36] European Data Protection Board, "Guidelines 07/2020 on the concepts of controller and processor in the GDPR" (July 2021), https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-072020-concepts-controller-and-processor-gdpr_en, accessed January 17, 2023. For the remainder of Section 3.2.1, reference is made to these guidelines. The notion of controller is covered in paras 15–45, that of joint control in paras 46–72 and that of processor in paras 73–84.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

formal designation – in a contract, for instance – and to analyze the factual elements or circumstances indicating a decisive influence over the processing.

Next, the "purposes" and "means" relate, respectively, to the "why's" and "how's" of the processing. An entity must exert influence over both those elements to qualify as a controller, although there is a margin of maneuver to delegate certain "non-essential means" without shifting the burden of control. This would be the case, for instance, for the "practical aspects of implementation." For example, a company that decides to store a backup copy of its customers' data on a cloud platform remains the controller for that processing even though it does not determine the type of hardware used for the storage, nor the transfer protocol, the security measures or the redundancy settings. On the contrary, decisions pertaining to the type of personal data processed, their retention period, the potential recipients to whom they will be disclosed, and the categories of data subjects they concern typically fall within the exclusive remit of the controller; any delegation of these aspects to another actor would turn that entity into a (joint) controller in its own right.

Finally, the wording "alone or jointly with others" hints at the possibility for two or more entities to be considered as joint controllers. According to the EDPB, the overarching criterion for joint controllership to exist is "the joint participation of two or more entities in the determination of the purposes and means of a processing operation." This is the case when the entities at stake adopt "common" or "converging" decisions. Common decisions, on the one hand, involve "a common intention." Converging decisions, on the other, "complement each other and are necessary for the processing to take place in such a manner that they have a tangible impact on the determination of the purposes and the means of the processing." Another indication is "whether the processing would not be possible without both parties' participation in the sense that the processing by each party is inseparable, i.e. inextricably linked." The CJEU has, for instance, recognized a situation of joint controllership between a religious community and its members for the processing of the personal data collected in the course of door-to-door preaching, as the former "organized, coordinated and encouraged" the said activities despite the latter being actually in charge of the processing.[37] The Court held a similar reasoning with regard to Facebook and the administrator of a fan page, as creating such a page "gives Facebook the opportunity" to place cookies on visitors' computer that can be used to both "improve its system of advertising" and to "enable the fan page administrator to obtain statistics from the visit of the page."[38] Lastly, the Court also considered Facebook and Fashion ID, an online clothing retailer that had embedded Facebook's "Like" plugin on its page, as joint controllers for the collection and transmission of the visitors' IP address and unique browser string, since both entities

[37] Case C-25/17 *Tietosuojavaltuutettu* [2018] ECLI:EU:C:2018:551, paras 70–75.
[38] Case C-210/16 *Wirtschaftsakademie Schleswig-Holstein GmbH* [2018] ECLI:EU:C:2018:388, paras 25–44.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

benefitted from that processing. Facebook, because it could use the collected data for its own commercial purpose. And Fashion ID, because the presence of a "Like" button would contribute to increasing the publicity of its goods.[39]

Next to "controllers," "processors" also fall within the scope of the GDPR. These are entities distinct from the controller that process personal data on its behalf (Article 4(8) GDPR). This is typically the case for, say, a call center that processes prospects' phone numbers in the context of a telemarketing campaign organized by another company. The requirement to be a separate entity implies that internal departments, or employees acting under the direct authority of their employer, will – at least in the vast majority of cases – not qualify as processors. Besides, processors can only process personal data upon the documented instructions and for the benefit of the controller. Should a processor go beyond the boundaries set by the controller and process personal data for its own benefit, it will be considered as a separate controller for the portion of the processing that oversteps the original controller's instructions. If the said call center decides, for instance, to reuse the phone numbers it has obtained from the controller to conduct its own marketing campaign or to sell it to third parties, it will be considered as a controller for those activities. Compared to controllers, processors must only comply with a subset of the rules listed in the GDPR, such as the obligation keep a record of processing activities (Article 30(2)), to cooperate with national supervisory authorities (Article 31), to ensure adequate security (Article 32), to notify data breaches to controllers (Article 33(2)), and to appoint a Data Protection Officer (DPO) when certain conditions are met (Article 44).

#### 7.3.2.2 The Allocation of Responsibilities in AI Systems

The CJEU has repeatedly emphasized the importance to ensure, through a broad definition of the concept of controller, the "effective and complete protection of data subjects."[40] The same goes for the notion of joint control, which the Court now seems to have extended to any actor that has made the processing possible by contributing to it.[41] In the context of complex processing operations involving multiple actors intervening at different stages of the processing chain, such as the ones at stake in AI systems, an overly broad interpretation of the notion of joint control might lead to situations where everyone is considered as a joint controller.[42] Properly allocating responsibilities is therefore essential, as the qualification of each

[39] Case C-40-17 *Fashion ID* [2019] ECLI:EU:C:2019:629, paras 64–85.

[40] Case C-131/12 *Google Spain* [2014] ECLI:EU:C:2014:317, para 34; Case C-210-16 (n 38), para 28; Case C-25/17 (n 37), para 21; ibid., para 66.

[41] See, on that note, the remark of Advocate General Bobek in his Opinion on the Fashion ID case. Case C-40/17 (n 39), Opinion of Advocate General Bobek, ECLI:EU:C:2018:1039, para 74.

[42] Concerns have been voiced by, for instance: Jiahong Chen et al., "Who is responsible for data processing in smart homes? Reconsidering joint controllership and the household exemption" (2020)

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

party will drastically impact the scope of their compliance duties. Doing so requires the adoption of a "phase-oriented" approach, by slicing complex sets of processing operations into smaller bundles that pursue an identical overarching purpose before proceeding with the qualification of the actors involved.[43] Machine learning models, for instance, are the products of different activities ranging from the gathering and cleaning of training datasets, to the actual training of the model and its later use to make inferences in concrete scenarios. The actors involved do not necessarily exert the same degree of influence over all these aspects. As a result, their qualification might differ depending on the processing operation at stake. This makes it particularly important to circumscribe the relevant processing activities before applying the criteria detailed in Section 7.3.2.1.[44]

Let's illustrate the above by breaking down the processing operations typically involved in machine learning, starting with the collection and further use of the training datasets. Company X might specialize in the in-house development and commercialization of trained machine learning models. When doing so, it determines why the training datasets are processed (i.e., to train their model with the view of monetizing it) as well as the essential and nonessential means of the processing (e.g., which personal data are included in the training dataset and the technical implementation of the training process). It will therefore be considered as the sole controller for the processing of the training datasets. Company X might also decide to collaborate with Company Y, the latter providing the training dataset in exchange for the right to use the model once trained. This could be considered as converging decisions leading to a situation of joint controllership between Companies X and Y. Looking at the inference stage, then, Company X might decide to offer its trained model to Company Z, a bank, that will use it to predict the risk of default before granting loans. By doing so, Company Z determines the purposes for which it processes its clients' personal data (i.e., calculating the risk of default), as well as the essential means of the processing (e.g., the granularity of the data fed to the model). As a result, Company Z will be considered as the sole controller for the processing of its customers' data, regardless of whether Company X retains a degree of influence over how the algorithm works under the hood. Company X could also be considered as a processor in case it computes the risk score on behalf of Company Z using its own hardware and software infrastructure. This is a common scenario in the context of software- or platform-as-a-service cloud-based solutions.

*International Data Privacy Law*, 10: 279; Christopher Millard, "At this rate, everyone will be a [joint] controller of personal data!" (2019) *International Data Privacy Law*, 9: 217.

[43] René Mahieu and Joris van Hoboken, "Fashion-ID: Introducing a phase-oriented approach to data protection?" (*European Law Blog*, September 30, 2019), https://europeanlawblog.eu/2019/09/30/fashion-id-introducing-a-phase-oriented-approach-to-data-protection/, accessed January 19, 2023.

[44] See, for more examples, the ICO Guidance on AI and data protection, more specifically under the section "How should we understand controller/processor relationships in AI?" Information Commissioner's Office, "Guidance on AI and Data Protection" (n 30) 23–27.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

## 7.4 AI SYSTEMS MEET THE GDPR – OVERVIEW AND FRICTION POINTS

Controllers – and, to a certain extent, processors – that process personal in the context of the development and/or use of AI systems must comply with the foundational principles detailed in Article 5 GDPR, namely lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability. These are the pillars around which the rest of the Regulation is articulated. While AI systems are not, *per se*, incompatible with the GDPR, reconciling their functioning with the rules of the Regulation is somewhat of a balancing act. The following sections aim at flagging the most pressing tensions by contrasting some of the characteristics of AI systems against the guarantees laid down in Article 5 GDPR.

### 7.4.1 *The Versatility of AI Systems v. the Necessity and Compatibility Tests*

#### 7.4.1.1 Lawfulness and Purpose Limitation at the Heart of the GDPR

In order to prevent function creep, Article 5(1)a introduces the principle of "lawfulness," which requires controllers to justify their processing operations using one of the six lawful grounds listed in Article 6. These include not only the consent of the data subject – often erroneously perceived as the only option – but also the alternatives such as the "performance of a contract" or the "legitimate interests of the controller." Relying on any of these lawful grounds (except for consent) requires the controller to assess and demonstrate that the processing at stake is "objectively necessary" to achieve the substance of that lawful ground. In other words, there is no other, less-intrusive way to meet that objective. As recently illustrated by the Irish regulator's decision in the Meta Ireland case,[45] the processing of Facebook and Instagram users' personal data for the purpose of delivering targeted advertising is not, for instance, objectively necessary to fulfil the essence of the contractual relationship between these platforms and their users.[46] As a result, the processing cannot be based on Article 6(1)b, and it has to rely on another lawful ground. Consent, on the other hand, must be "freely given, specific, informed and unambiguous," thereby undermining its validity when obtained in a scenario that involves

[45] Full decision still to be published; see: www.dataprotection.ie/en/news-media/data-protection-commission-announces-conclusion-two-inquiries-meta-ireland, accessed January 23, 2023.

[46] See, for other examples: European Data Protection Board, "Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects," paras 23–29, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf, accessed January 17, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

unbalanced power or information asymmetries, such as when given by an employee to its employer.[47]

With that same objective in mind, Article 5(1)b lays down the principle of "purpose limitation," according to which personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes."[48] In practice, this requires controllers to, first, determine the exact reasons why personal data are collected and, then, assess the compatibility of every subsequent processing activity in light of the purposes that were specified at the collection stage. Doing so requires to take into account various criteria such as, for instance, the context in which the personal data have been collected and the reasonable expectations of the data subjects.[49] While compatible further processing can rely on the same lawful ground used to justify the collection, incompatible processing must specify a new legal basis. Reusing a postal address originally collected to deliver goods purchased online for marketing purposes is a straightforward example of an incompatible further processing. The purposes specified during the collection also serve as the basis to assess the amount of personal data collected (i.e., "data minimization"), the steps that must be taken to ensure their correctness (i.e., "accuracy") and their retention period (i.e., "storage limitation").

Lawfulness and purpose limitation are strongly interconnected, as the purposes specified for the collection will influence the outcome of both the necessity test required when selecting the appropriate lawful ground – with the exception of consent, for which the purposes delimit what can and cannot be done with the data – and the compatibility assessment that must be conducted prior to each further processing. Ensuring compliance with these principles therefore calls for a separate analysis of each "personal data – purpose(s) – lawful ground" triad, acting as a single, indissociable whole (see Figure 7.3).

Severing the link between these three elements would empty Articles 5(1)a and 5(1)b from their substance and render any necessity or compatibility assessment meaningless. Whether a webshop can rely on its legitimate interests (Article 6(1)f) to profile its users and offers targeted recommendations, for instance, heavily depends on the actual personal data used to tailor their experience, and therefore the intrusiveness of the processing.[50]

[47] European Data Protection Board, "Guidelines 05/2020 on Consent under Regulation 2016/679," paras 13–54, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf, accessed January 15, 2023.

[48] For a thorough overview of that principle, see: Article 29 Working Party, "Opinion 03/2013 on purpose limitation," https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf, accessed January 16, 2023.

[49] Recital 50 GDPR also highlights the relevance of other criteria such as "the nature of the personal data, the consequences of the intended further processing for data subjects, and the existence of appropriate safeguards in both the original and intended further processing operations."

[50] More examples can be found in Annex 2 of: Article 29 Working Party, "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC," www.dataprotection.ro/servlet/ViewDocument?id=1086, accessed January 14, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

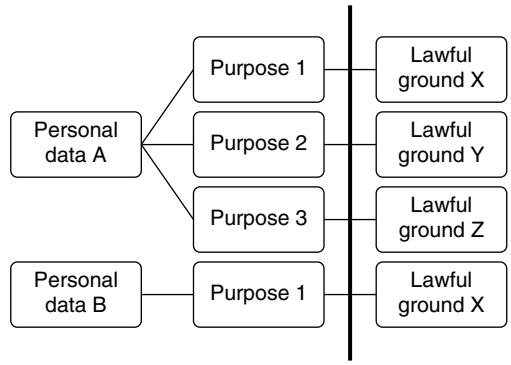



FIGURE 7.3 Lawfulness and purpose limitation, combined

#### 7.4.1.2 Necessity and Compatibility in AI Systems

While complying with the principles of lawfulness and purpose limitation is already a challenge in itself, the very nature of AI systems spices it up even more. The training of machine learning models, for example, often involves the reuse, as training datasets, of personal data originally collected for completely unrelated purposes. While it is still unclear whether scraping publicly accessible personal data should be regarded as *a further processing* activity subject to the compatibility assessment pursuant to Articles 6(1)b and 6(4) GDPR, or as a *new collection* for which the said entity would automatically need to rely on a *different* lawful ground than the one used to legitimize the original collection, this raises the issue of function creep and loss over one's personal data. The case of Clearview AI is a particularly telling example. Back in 2020, the company started to scrape the internet, including social media platforms, to gather images and videos to train its facial recognition software and offer its clients – among which law enforcement authorities – a search engine designed to look up individuals on the basis of another picture. After multiple complaints and a surge in media attention, Clearview was fined by the Italian,[51] Greek,[52] French,[53]

[51] Garante per la protezione dei dati personali, Ordinanza ingiunzione nei confronti di Clearview AI [2022], www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9751362, accessed January 24, 2023.

[52] Αρχή προστασίας δεδομένων προσωπικού χαρακτήρα, Επιβολή προστίμου στην εταιρεία Clearview AI, Inc [2022], www.dpa.gr/el/enimerwtiko/prakseisArxis/epiboli-prostimoy-stin-etaireia-clearview-ai-inc, accessed January 24, 2023.

[53] Commission nationale de l'informatique et des libertés, Délibération de la formation restreinte n° SAN-2022-019 du octobre 17, 2022 concernant la société Clearview AI [2022], www.legifrance.gouv.fr/cnil/id/CNILTEXT000046444859, accessed January 24, 2023. See also, more recently, the 5.2 million penalty payment issued by the CNIL against Clearview AI for non-compliance with the above-mentioned injunction: Commission nationale de l'informatique et des libertés, Délibération de la formation restreinte n° SAN-2023-005 du 17 avril 2023 concernant la société Clearview AI [2023], www.legifrance.gouv.fr/cnil/id/CNILTEXT000047527412, accessed June 15, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

and UK[54] regulators for having processed these images without a valid lawful ground. The Austrian regulator issued a similar decision, if not paired with a fine.[55] As detailed in Section 7.4.1.1, the fact that these images are *publicly accessible* does not, indeed, mean that they are *freely reusable* for any purpose. All five authorities noted the particularly intrusive nature of the processing at stake, the amount of individuals included in the database, and the absence of any relationship between Clearview AI and the data subjects who could therefore not reasonably expect their biometric data to be repurposed for the training of a facial recognition algorithm.

The training of Large Language Models ("LLMs") such as OpenAI's GPT-4 or EleutherAI's GPT-J raises similar concerns, which the Garante recently flagged in its decision to temporarily ban[56] – then conditionally reauthorize –[57] ChatGPT on the Italian territory.[58] This even prompted the EDPB to set up a dedicated task force to "foster cooperation and to exchange information on possible enforcement actions conducted by data protection authorities."[59] Along the same lines, but looking at the

[54] Information Commissioner's Office, Monetary Penalty Notice to Clearview AI Inc of May 26, 2022 [2022], https://ico.org.uk/media/action-weve-taken/mpns/4020436/clearview-ai-inc-mpn-20220518.pdf, accessed June 15, 2023; see also, for the order to stop obtaining and using the personal data of UK residents that is publicly available on the internet, and to delete the data of UK residents from its systems: Information Commissioner's Office, Enforcement Notice to Clearview AI Inc. of May 26, 2022 [2022], https://ico.org.uk/media/action-weve-taken/enforcement-notices/4020437/clearview-ai-inc-en-20220518.pdf, accessed June 15, 2023.

[55] Datenschutzbehörde, Decision of May 9, 2023 against Clearview AI [2023], https://noyb.eu/sites/default/files/2023-05/Clearview%20Decision%20Redacted.pdf.

[56] Garante per la protezione dei dati personali, Provvedimento del 30 marzo 2023 [9870832] [2023], www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832, accessed June 15, 2023. An earlier decision issued against Luka Inc., the company behind Replika, also questioned the lawful ground applicable in the context of companion chatbots. See: Garante per la protezione dei dati personali, Provvedimento del 2 febbraio 2023 [9852214] [2023], www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9852214, accessed June 15, 2023.

[57] Garante per la protezione dei dati personali, ChatGPT: Garante privacy, limitazione provvisoria sospesa se OpenAI adotterà le misure richieste. L'Autorità ha dato tempo allá società fino al 30 aprile per mettersi in regola [2023], www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9874751, accessed June 15, 2023; ChatGPT: OpenAI riapre la piattaforma in italia garantendo più trasparenza e più diritti a utenti e non utenti europei, www.gpdp.it/home/docweb/-/docweb-display/docweb/9881490. For an overview of the new controls added by ChatGPT following the Garante's ban, see the dedicated Help Centre Article on OpenAI's website: https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed, accessed June 15, 2023. Yet, OpenAI did not offer any solution to remedy the unlawfulness of the processing of the personal data contained in the dataset used to train ChatGPT.

[58] It is also worth noting that OpenAI now faces a class action in California for a breach of both data protection and copyright law. See: Gerrit De Vynck, "ChatGPT maker OpenAI faces a lawsuit over how it used people's data" (2023) *Washington Post* (June 28), www.washingtonpost.com/technology/2023/06/28/openai-chatgpt-lawsuit-class-action/, accessed July 4, 2023.

[59] The EDPB announced the creation of the task force back in April 2023. See: www.edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt. In May 2024, it published a meager interim report documenting the results of the said taskforce that "reflect[s] the common denominator agreed by the Supervisory Authorities in their interpretation of the applicable provisions of the GDPR in relation to the matters that are within the scope of their investigation."

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

inference rather than the training phase, relying on algorithmic systems to draw predictions might not always be proportional – or even necessary – to achieve a certain objective. Think about an obligation to wear a smart watch to dynamically adjust a health insurance premium, for instance.

As hinted at earlier, the principle of "data minimization" requires to limit the amount of personal data processed to what is objectively necessary to achieve the purposes that have been specified at the collection stage (Article 5(1)c GDPR). At first glance, this seems to clash with the vast amount of data often used to train and tap into the potential of AI systems. It is therefore essential to reverse the "collect first, think after" mindset by laying down the objectives that the AI system is supposed to achieve *before* harvesting the data used to train or fuel its predictive capabilities. Doing so, however, is not always realistic when such systems are designed outside any concrete application area and are meant to evolve over time. Certain techniques can nonetheless help reduce their impact on individuals' privacy. At the training stage, pseudonymization methods such as generalization and randomization – both discussed in Section 7.3.1.2 – remain pertinent. Standard feature selection methods can also assist controllers in pruning their training datasets from variables that are of little added-value in the development of their model.[60] In addition, federated machine learning, which relies on the training, sharing and aggregation of "local" models, is a viable alternative to the centralization of training datasets in the hands of a single entity, and reduces the risks associated with their duplication.[61] At the inference stage, running the machine learning model on the device itself rather than hosting it on the cloud is also an option to cut on the need to share personal data with a central entity.[62]

### 7.4.2 *The Complexity of AI Systems v. Transparency and Explainability*

#### 7.4.2.1 Ex-ante and Ex-post Transparency Mechanisms

As a general principle, transparency percolates through the entire Regulation and plays a critical role in an increasingly datified society. As noted in Recital 39 GDPR,

See: European Data Protection Board, "Report of the Work Undertaken by the ChatGPT Taskforce," www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf. Looking beyond the EU, ChatGPT is also on the radar of the Office of the Privacy Commissioner of Canada. See: Office of the Privacy Commissioner of Canada, Announcement of April 4, 2023, www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230404/, accessed June 15, 2023.

[60] For an overview of these methods: Jason Brownlee, "How to choose a feature selection method for machine learning" (MachineLearningMastery.com, November 26, 2019), https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/, accessed January 25, 2023.

[61] Stephanie Rossello, Luis Muñoz-González, and Roberto Díaz Morales, "Data protection by design in AI? The case of federated learning" (2021) *Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht*, 3: 273.

[62] For other relevant examples of minimization techniques that can be deployed at the inference stage, see: Information Commissioner's Office, "Guidance on AI and Data Protection" (n 30) 66–68.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

"it should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed." To meet that objective, Articles 13 and 14 detail the full list of information controllers must provide to data subjects. It includes, among others, the contact details of the controller and its representative, the purposes and legal basis of the processing, the categories of personal data concerned, any recipient, and information on how to exercise their rights.[63] Article 12 then obliges controllers to communicate that information in a "concise, transparent, intelligible and easily accessible way, using clear and plain language," in particular for information addressed to children. This requires them to tailor the way they substantiate transparency to their audience by adapting the tone and language to the targeted group. Beyond making complex environments observable, this form of *ex-ante* transparency also pursues an instrumental goal by enabling other prerogatives.[64] As pointed out in literature, "neither rectification or erasure […] nor blocking or objecting to the processing of personal data seems easy or even possible unless the data subject knows exactly what data [are being processed] and how."[65] Articles 13 and 14 therefore ensure that data subjects are equipped with the necessary information to later exercise their rights.

In this regard, Articles 15 to 22 complement Articles 13 and 14 by granting data subjects an arsenal of prerogatives they can use to regain control or balance information asymmetries. These include the right to access, to rectify, to erase, restrict, and move one's data, as well as the right to challenge and to object to certain types of automated decision-making processes. More specifically, Article 15 grants data subjects the right to request a confirmation that personal data concerning them are being processed, more information on the relevant processing operations and a copy of the personal data involved. As a form of *ex-post* transparency mechanism, it allows data subjects to look beyond what is provided in a typical privacy policy and obtain an additional, individualized layer of transparency. Compared to the information provided in the context of Articles 13 and 14, controllers should, when answering an access request, tailor the information provided to the data subject's specific situation. This would involve sharing the recipients to whom their personal data have *actually* been disclosed, or the sources from which these have *actually* been obtained – a point of information that might not always be clear at the time

[63] For a detailed overview of Articles 12, 13, and 14 GDPR, see: Article 29 Working Party, "Guidelines on Transparency under Regulation 2016/679," https://ec.europa.eu/newsroom/article29/redirection/document/51025, accessed January 16, 2023.

[64] Laurens Naudts, Pierre Dewitte, and Jef Ausloos, "Meaningful transparency through data rights: A multidimensional analysis" (2022) *Research Handbook on EU Data Protection Law* 530, 540.

[65] Jef Ausloos and Pierre Dewitte, "Shattering one-way mirrors – data subject access rights in practice" (2018) *International Data Privacy Law*, 8: 7, https://academic.oup.com/idpl/advance-article/doi/10.1093/idpl/ipy001/4922871, accessed May 16, 2023. See also the many references therein.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

the privacy policy is drafted.[66] By allowing data subjects to verify controllers' practices, Article 15 paves the way for further remedial actions, should it be necessary. It is therefore regarded as one of the cornerstones of data protection law, and is one of the few guarantees explicitly acknowledged in Article 8 CFREU.

#### 7.4.2.2 Algorithmic Transparency – And Explainability?

AI systems are increasingly used to make or support decisions concerning individuals based on their personal data. Fields of applications range from predictive policing to hiring strategies and healthcare, but all share a certain degree of opacity as well as the potential to adversely affect the data subjects concerned. The GDPR seeks to address these risks through a patchwork of provisions regulating what Article 22(1) defines as "decisions based solely on automated processing, including profiling, which produce legal effects concerning [the data subject] or similarly significantly affect him or her." This would typically include, according to Recital 71, the "automatic refusal of an online credit applications" or "e-recruiting practices without any form of human intervention." Based *solely*, in this case, suggests that the decision must not necessarily be *taken* by an automated system for it to fall within the scope of Article 22(1). The routine usage of a predictive system by a person who is not in a position to exercise any influence or meaningful oversight over its outcome would, for instance, also fall under Article 22(1).[67] While fabricating human involvement is certainly not a viable way out, national data protection authorities are still refining the precise contours of that notion.[68]

Controllers that rely on such automated decision-making must inform data subjects about their existence, and provide them with "meaningful information about the logic involved," as well as their "significance and the envisaged consequences." This results from the combined reading of Articles 13(2)f, 14(2)g, and 15(1)h. Additionally, Article 22(3) and Recital 71 grant data subjects the right to obtain human intervention, express their point of view, contest the decision and – allegedly – obtain an explanation of the decision reached. Over the last few years, these provisions have fueled a lively debate as to the existence of a so-called "right to

[66] The fact that the elements listed in Article 15 partially overlap with the ones listed in Articles 13 and 14 does not mean that the controller can always answer an access request by recycling elements from its privacy policy or record of processing. See: European Data Protection Board, "Guidelines 01/2022 on data subject rights – right of access," para 111, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012022-data-subject-rights-right-access_en, accessed January 16, 2023.

[67] Article 29 Working Party, "Guidelines on data protection impact assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of regulation 2016/679" 21, https://ec.europa.eu/newsroom/document.cfm?doc_id=47711, accessed January 25, 2022.

[68] See, for the interpretation proposed by national supervisory authorities across Europe: Sebastião Barros Vale and Gabriela Zanfir-Fortuna, "Automated decision-making under the GDPR: Practical cases from courts and data protection authorities" (Future of Privacy Forum, 2022), https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf, accessed January 11, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

explanation" that would allow data subjects to enquire about how a *specific* decision was reached rather than only about the overall *functioning* of the underlying system.[69] Regardless of these controversies, it is commonly agreed that controllers should avoid "complex mathematical explanations" and rather focus on concrete elements such as "the categories of data that have been or will be used in the profiling or decision-making process; why these categories are considered pertinent; how the profile is built, including any statistics used in the analysis; why this profile is relevant and how it is used for a decision concerning the data subject."[70] The "right" explanation will therefore strongly depend on the sector and audience at stake.[71] A media outlet that decides to offer users a personalized news feed might, for instance, need to explain the actual characteristics taken into account by its recommender system, as well as their weight in the decision-making process and how past behavior has led the system to take a specific editorial decision.[72]

### 7.4.3 *The Dynamicity of AI v. the Risk-Based Approach*

#### 7.4.3.1 Accountability, Responsibility, Data Protection by Design and DPIAs

Compared to its predecessor,[73] one of the main objectives of the GDPR was to move away from compliance as a mere ticking-the-box exercise – or window dressing[74] – by incentivizing controllers to take up a more proactive role in the

[69] See, among others: Bryce Goodman and Seth Flaxman, "European Union Regulations on algorithmic decision-making and a 'right to explanation'" (2017) *AI Magazine*, 38, http://arxiv.org/abs/1606.08813; Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Why a right to explanation of automated decision-making does not exist in the general data protection regulation" (2017) *International Data Privacy Law*, 7: 76; Gianclaudio Malgieri and Giovanni Comandé, "Why a right to legibility of automated decision-making exists in the general data protection regulation" (2017) *International Data Privacy Law*, 7: 243.

[70] See Annex 1 of Article 29 Working Party, "WP29, guidelines on DPIA" (n 67) 31.

[71] The British regulator has provided a solid overview of the different types of explanations controllers could provide. See, more specifically, the Section "What goes into an explanation" from the Information Commissioner's Office and Alan Turing Institute, "Explaining decisions made with AI," https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf, accessed January 25, 2023.

[72] Max van Drunen, Natali Helberger, and Mariella Bastian, "Know your algorithm: What media organizations need to explain to their users about news personalization" (2019) *International Data Privacy Law*, 9: 220.

[73] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31 ELI: http://data.europa.eu/eli/dir/1995/46/oj.

[74] The EDPS indeed noted that "in the past, privacy and data protection have been perceived by many organisations as an issue mainly related to legal compliance, often confined to the mere formal process of issuing long privacy policies covering any potential eventuality and reacting to incidents in order to minimise the damage to their own interests." See: European Data Protection Supervisor, "Opinion 5/2018 – Preliminary Opinion on Privacy by Design," para 13, https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf, accessed January 15, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

implementation of appropriate measures to protect individuals' rights and freedoms. This led to the abolition of the antique, paternalistic obligation for controllers to notify their processing operations to national regulators in favor of a more flexible approach articulated around the obligation to maintain a record of processing activities (Article 30), to notify data breaches to competent authorities and the affected data subjects (Articles 33 and 34) and to consult the former in cases where a data protection impact assessment ("DPIA") indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk (Article 36). The underlying idea was to responsibilize controllers by shifting the burden of analyzing and mitigating the risks to data subject's rights and freedoms onto them. Known as the "risk-based approach," it ensures both the flexibility and scalability needed for the underlying rules to remain pertinent in a wide variety of scenarios. As noted in legal literature, the risk-based approach "provides a way to carry out the shift to accountability that underlies much of the data protection reform, using the notion of risk as a reference point in light of which we can assess whether the organisational and technical measures taken by the controller offer a sufficient level of protection."[75]

The combined reading of Articles 5(2) ("accountability"), 24(1) ("responsibility"), and 25(1) ("data protection by design") now requires controllers to take into account the state of the art, the cost of implementation, and the nature, scope, context, and purposes as well as the risks posed by the processing. They should implement, both at the time of determination of the means for processing and at the time of the processing itself, appropriate technical and organizational measures to ensure and demonstrate compliance with the Regulation. In other words, they must act responsibly as of the design stage, and throughout the entire data processing lifecycle. Data protection-specific risks are usually addressed in a DPIA, which should at least provide a detailed description of the relevant processing activities, an assessment of their necessity and proportionality, as well as an inventory of the risks and corresponding mitigation strategies (see Figure 7.4).[76] While Article 35(1) obliges controllers to conduct a DPIA for processing activities that are "likely to result in a high risk for rights and freedoms of natural persons," such an exercise, even if succinct, is also considered as best practice for all controllers regardless of the level of risk.[77]

[75] Claudia Quelle, 'Enhancing Compliance under the General Data Protection Regulation: The Risky Upshot of the Accountability- and Risk-Based Approach' (2018) 9 *European Journal of Risk Regulation* 502, 505.

[76] See, for a detailed overview of the steps involved in a DPIA: Article 35(7) GDPR and Annex 2 of the Article 29 Working Party, "WP29, Guidelines on DPIA" (n 67).

[77] European Data Protection Board, "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default," para 32, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf, accessed May 3, 2022.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

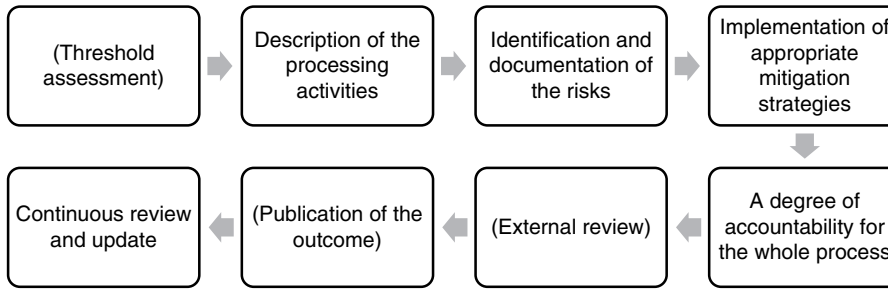



FIGURE 7.4 Overview of the main steps of a Data Protection Impact Assessment

#### 7.4.3.2 From DPIAs to AIAs, and the Rise of Algorithmic Governance

The development and use of AI systems are often considered as processing likely to result in a "high risk," for which a DPIA is therefore mandatory. In fact, Article 35(3) GDPR, read in combination with the Guidelines from the WP29 on the matter,[78] extends that obligation to any processing that involves, among others, the evaluation, scoring or systematic monitoring of individuals, the processing of data on a large scale, the matching or combining of datasets or the innovative use or application of new technological or organizational solutions. All these attributes are, in most cases, inherent to AI systems and therefore exacerbate the risks for individuals' fundamental rights and freedoms. Among these is, for instance, the right not to be discriminated. This is best illustrated by the Dutch "Toeslagenaffaire," following which the national regulator fined the Tax Administration for having unlawfully created erroneous risk profiles using a machine learning algorithm in an attempt to detect and prevent child care benefits fraud, which led to the exclusion of thousands of alleged fraudsters from social protection.[79] Recent research has also uncovered the risk of bias in predictive policing and offensive speech detection systems, both vulnerable to imbalanced training datasets, and susceptible to reflect past discrimination.[80]

Addressing these risks requires more than just complying with the principles of lawfulness, purpose limitation, and data minimization. It also goes beyond the provision of explanations, however accessible and accurate these may be. In fact, that issue largely exceeds the boundaries of the GDPR itself which, as hinted in Section 7.3, is but one regulatory angle among many others. The AI Act is, for

[78] Article 29 Working Party, "WP29, Guidelines on DPIA" (n 67) 9–12.

[79] Autoriteit Persoonsgegevens, "Boete Belastingdienst voor zwarte lijst FSV" April 12, 2022, https://autoriteitpersoonsgegevens.nl/nl/nieuws/boete-belastingdienst-voor-zwarte-lijst-fsv, accessed January 25, 2023.

[80] Competition and Market Authority and others, "Auditing algorithms: The existing landscape, role of regulators and future outlook" (Digital Regulation Cooperation Forum) Findings from the DRCF Algorithmic Processing workstream – Spring 2022, www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook, accessed January 26, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

instance, a case in point.[81] More generally, this book is a testimony to the diversity of the regulatory frameworks applicable to AI systems. This calls for a drastic rethinking of how AI systems are designed and deployed to mitigate their adversarial impact on society. This has led to the development of *Algorithmic* – rather than *Data Protection* – Impact Assessments ("AIAs"), conceived as broader risk management approaches that integrate but are not limited to data protection concerns.[82] While these assessments can assist controllers in developing their own technology, they are also relevant for controllers relying on off-the-shelf AI solutions offered by third parties, who are increasingly resorting to auditing and regular testing to ensure that these products comply with all applicable legislation. All in all, the recent surge in awareness of AI's risks has laid the groundwork for the rise of a form of algorithmic accountability.[83] Far from an isolated legal exercise, however, identifying and mitigating the risks associated with the use of AI systems is, by nature, an interdisciplinary exercise. Likewise, proper solutions will mostly follow from the research conducted in fora that bridge the gap between these different domains, such as the explainable AI ("XAI") and human–computer interaction ("HCI") communities.

## 7.5 CONCLUSION

As pointed out from the get go, this chapter serves as an entry point into the intersection of AI and data protection law, and strives to orient the reader toward the most authoritative sources on each of the subjects it touches upon. It is hence but a curated selection of the most relevant data protection principles and rules articulated around the most salient characteristics of AI systems. Certain important issues therefore had to be left out, among which the obligation to ensure a level of security appropriate to the risks at stake, the rules applicable to special categories of personal data, the exercise of data subjects rights, the role of certification mechanisms and codes of conduct, or the safeguards surrounding the transfers of personal data to third countries. Specific sources on these issues are, however, plentiful.

There is no doubt that AI systems, and the large-scale processing of personal data that is often associated with their development and use, has put a strain on individuals' fundamental rights and freedoms. The goal of this chapter was to highlight the

[81] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) [2024] OJ L144/1 ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

[82] See, for a use case in the healthcare sector: Lara Groves, "Algorithmic impact assessment: A case study in healthcare" (Ada Lovelace Institute, 2022), www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/, accessed January 26, 2023.

[83] Christian Katzenbach and Lena Ulbricht, "Algorithmic governance" (2019) *Internet Policy Review*, 8(4), https://policyreview.info/concepts/algorithmic-governance, accessed January 26, 2023.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press

role of the GDPR in mitigating these risks by clarifying its position and function within the broader EU regulatory ecosystem. It also aimed to equip the reader with the main concepts necessary to decipher the complexity of its material and personal scope of application. More importantly, it ambitioned to debunk the myth according to which the applicability of the GDPR to AI systems would inevitably curtail their deployment, or curb innovation altogether. As illustrated throughout this contribution, tensions do exist. But the open-ended nature of Article 5, paired with the interpretation power granted to European and national supervisory authorities, provide the flexibility needed to adapt the GDPR to a wide range of scenarios. As with all legislation that aims to balance competing interests, the key mostly – if not entirely – lies in ensuring the necessity and proportionality of the interferences of the rights at stake. For that to happen, it is crucial that all stakeholders are aware of both the risks raised by AI systems for the fundamental rights to privacy and data protection, and of the solutions that can be deployed to mitigate these concerns and hence guarantee an appropriate level of protection for all the individuals involved.

https://doi.org/10.1017/9781009367783.010 Published online by Cambridge University Press