Finale Doshi-Velez and Been Kim, 'Towards a rigorous science of interpretable machine learning'. arXiv preprint arXiv:1702.08608 (2017).

Alex A. Freitas, 'Comprehensible classification models: a position paper' (2014) 15(1) ACM SIGKDD explorations newsletter, 1–10.

Bryce Goodman and Seth Flaxman, 'EU regulations on algorithmic decision-making and a "right to explanation"'. ICML workshop on human interpretability in machine learning (2016).

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Dino Pedreschi and Fosca Giannotti, 'A survey of methods for explaining black box models' (2018) 51(5) ACM computing surveys 1–42.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Dino Pedreschi and Fosca Giannotti, 'Factual and counterfactual explanations for black box decision making'. IEEE Intelligent Systems (2019).

Stefan Haufe et al., 'On the interpretation of weight vectors of linear models in multivariate neuroimaging' (2014) 87 Neuroimage 96.

Johan Huysmans et al., 'An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models' (2011) 51(1) Decision Support Systems 141.

Been Kim, Rajiv Khanna and Oluwasanmi O. Koyejo, 'Examples are not enough, learn to criticize! Criticism for interpretability'. Advances in neural information processing systems (2016).

John K. C. Kingston, 'Artificial intelligence and legal liability'. International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cham, 2016.

Joshua A. Kroll et al., 'Accountable algorithms' (2016) 165 U. Pa. L. Rev. 633.

Ana Lucic, Hinda Haned and Maarten de Rijke, 'Why does my model fail? Contrastive local explanations for retail forecasting'. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (2020).

Gianclaudio Malgieri and Giovanni Comandé, 'Why a right to legibility of automated decision-making exists in the general data protection regulation' (2017) 7(4) International Data Privacy Law 243.

Tim Miller, 'Explanation in artificial intelligence: Insights from the social sciences' (2019) 267 Artificial Intelligence 1–38.

Grégoire Montavon et al., 'Explaining nonlinear classification decisions with deep Taylor decomposition' (2017) 65 Pattern Recognition 211.

Anh Nguyen, Jason Yosinski and Jeff Clune, 'Deep neural networks are easily fooled: High confidence predictions for unrecognizable images'. Proceedings of the IEEE conference on computer vision and pattern recognition (2015).

Frank Pasquale, *The Black Box Society* (Harvard University Press 2015).

Dino Pedreschi et al., 'Meaningful explanations of black box AI decision systems' (2019)

33 Proceedings of the AAAI Conference on Artificial Intelligence, 9780–4.

J. Ross Quinlan, *C4. 5: Programs for Machine Learning* (Elsevier 2014).

Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin, '"Why should I trust you?" Explaining the predictions of any classifier'. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge discovery and data mining (2016).

Salvatore Ruggieri, Dino Pedreschi and Franco Turini, 'Data mining for discrimination discovery' (2010) 4(2) ACM Transactions on Knowledge Discovery from Data 1–40.

Karen Simonyan, Andrea Vedaldi and Andrew Zisserman, 'Deep inside convolutional networks: Visualising image classification models and saliency maps'. arXiv preprint arXiv:1312.6034 (2013).

Christian Szegedy et al., 'Intriguing properties of neural networks'. arXiv preprint arXiv:1312.6199 (2013).

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) 7(2) International Data Privacy Law, 76–99.

Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual explanations without opening the black box: Automated decisions and the GDPR' (2017) 31 Harv. JL & Tech. 84.

31 Fairness

31.1 Algorithm and 'unfair' biases

Machine learning techniques as drivers of businesses and public administrations' activities provide a data-driven basis upon which data subjects and clusters of data subjects are categorised. These categorisations occur through statistical differentiation procedures regarding the data sets employed, which arrange the inputs around 'differentiating' patterns and correlations. Upon these same patterns and correlations, decisions regarding considered individuals and groups are ultimately taken.

The differentiation process, first of all, regards the data sets employed as input, in machine learning the features that are relevant for the purposes of the decision-making, on the basis of the given data set. The enactment of the decision-making process is the second stage of machine learning–driven differentiation, naturally implying the making of choices among considered individuals and groups.

With the spread of these forms of automated decision-making processes, the risk of unfair treatment of data subjects involved has also increased. Unfair treatment directly results from the presence of biases in the artificial intelligence (AI) systems employed.

The algorithms used indeed may entail historical biases, failures of incompleteness and be based on bad governance models. In these regards, the literature[805] distinguishes two major algorithmic biases: biased training data and proxy discrimination.

With reference to biases affecting training data, it needs to be recalled that training data is the informational basis upon which machine learning algorithms optimise a statistical model that links input to output data.

Training data is thus the benchmark that assures that the predictions and correlations made on the basis of a certain input are right. Errors in collected data may make patterns harder to identify or lead to false pattern recognition.[806]

One common source of biased training data is given by sampling bias. This bias occurs when some strands of the population are misrepresented, because there is not a sufficient representation of the features of these strands of the population in the data sets used. There may also be cases where data sets have data referring to these groups, but these data are less valid and prone to error.

Biases in training data are also likely to generate biases in the subsequent analytical processing of collected data. Algorithms build a score from the given inputs (target variable, such as a risk score). However, this output, and thus the scores of the target variable, may excessively correlate with membership in a protected group.[807]

A different kind of bias is related to the 'unequal ground truth' stemming from the untruthful distribution of capacities or risks among protected groups. Such bias causes a 'proxy discrimination', that is a statistical discrimination, given by 'untrue' statistical associations and subsequent scientific inferences.[808]

Through these biases, algorithms reproduce and reinforce unfair inequalities regarding 'protected' classes in society. Indeed, the presence of these biases may generate prejudices to certain categories of people and groups, directly impacting on both 'individual' and 'group fairness'.

'Individual fairness' is safeguarded when similar individuals are treated alike by the processing system and is thus violated when two individuals sharing similar features, except for a certain (discriminatory) criterion, are treated differently.

Conversely, 'group fairness' is safeguarded through statistical parity, which occurs when each group determined by the model bears similar outcome distributions. The groups upon which machine learning processes build their categorisations can be defined not only based on protected features, as those corresponding to the categories of non-discrimination law, but also on the basis of other statistically relevant features, which may lead to unfair results and outcomes. In this way the same structural functioning mechanisms of machine-learning processes may give rise to stigmatisation and stereotyping patterns.[809]

The impairment of either individual or group fairness ultimately results in the exacerbation of marginalisation and social divides. Given that these processes ultimately decide on the allocation of burdens, benefits and opportunities given to citizens, these forms of decision-making sensitively interact with the notion of equality as a principle of justice.

In the last years, there have been attempts to define fairness from a technical standpoint, by including a certain notion of fairness in

[805] G. Comandé, 'Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems and Opportunities of Algorithms', in T. Cerquitelli, D. Quercia and F. Pasquale (eds), *Transparent Data Mining for Small and Big Data*, (2017) New York: Springer, 169–206; P. Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law', *Common Market Law Review* (2018) 55, 4, 1143–85; J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson and H. Yu, 'Accountable Algorithms', *University of Pennsylvania Law Review* (2016–17) 165, 3, 633–705.

[806] B. Lepri, O. Nuria, E.F. Letouzé and P. Vinck, 'Fair, Transparent and Accountable Algorithmic Decision-Making Processes: The Premises, the Proposed Solutions and the Open Challenges', *Philosophy & Technology* (2018) 31, 611–27.

[807] Hacker (n 805) 1148–50.
[808] Ibid., 1149.
[809] Ibid., 1176–8.

algorithmic decision-making processes 'by design'.[810]

Various normative approaches to formalise fairness have been developed, mainly in the form of (1) parity-based systems and (2) preference-based systems. These techniques are, however, based on egalitarian theories and envy-freeness literature, which link algorithmic biases to 'sensitive' attributes, as gender and ethnicity. Accordingly, Lepri and others[811] have defined fairness as 'the lack of discrimination or bias in the algorithmic decisions'.

However, these techniques appear to be rooted in a qualitatively limited version of 'equality'. Only recently, the literature[812] has acknowledged how the notion of fairness should be more substantively rooted in the stereotypes that machine-learning processes themselves create, disregarding sensitive-attribute groups. In this perspective, fairness is not a property to be considered in isolation, but with regard to systems as a whole and in synergy with other communicating systems.

31.2 Fairness and ethics

The European Commission's 'Ethics Guidelines for Trustworthy Artificial Intelligence'[813] highlight how equality in data-driven decision-making processes requires that 'the system's operations' do not 'generate unfairly biased outputs', this implying particular attention towards 'vulnerable persons and groups', which are 'at risk of exclusion'.[814] The same Guidelines consider fairness together with diversity and non-discrimination objectives.

The Commission's High-Level group of experts has acknowledged two dimensions of fairness: a substantial and a procedural one. The substantive dimension of fairness is accomplished by the 'equal and just distribution of both benefits and costs and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation'.[815]

In the Commission's view, fairness is thus related to the absence of the above-recalled unfair biases in automated decision-making processes. This can be achieved especially by respecting 'the principle of proportionality between means and ends' and to the balancing of competing interests and objectives, as inherently required by the proportionality principle.[816] Framed in these terms, thus, fairness is related to equal opportunity to access to services, ensured through inclusive and diverse design processes.

According to the High-Level Expert Group on AI, equal opportunities in services' accessibility is to be achieved especially through the incorporation of universal design principles,[817] as envisaged under Article 42 of the Public Procurement Directive,[818] requiring technical specifications to consider accessibility and 'design for all'. In addition, accessibility standards should also be taken into account.

For the purposes of a fair and equitable design of data-driven technologies, the Commission also encourages the involvement of the different stakeholders that will be most impacted by the technologies employed. By properly addressing the needs and feedback of a wider range of users, resulting devices would also better protect data subjects' freedom of choice.[819]

From this perspective, fairness is inherently bound to human autonomy, which is in turn associated with the rights to dignity and liberty under Articles 1 and 6 of the European Charter of Fundamental Rights. The violation of human autonomy can indeed be seen as a violation of the right to moral and physical integrity, as enshrined in Article 3 of the same Charter.

---

[810] L. Naudts, 'Towards Accountability: The Articulation and Formalization of Fairness in Machine Learning', *FIP Summer School on Privacy and Identity Management 'Fairness, Accountability and Transparency in the Age of Big Data'* (20–24 August 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3298847.

[811] Lepri et al. (n 806).

[812] S. Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising', *Berkely Technology Law Journal* (2020) 35, 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388639.

[813] European Commission, High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[814] Ibid., 2.

[815] Ibid., 12.

[816] Ibid.

[817] Ibid.

[818] Council Directive 2014/24/EU of 26 February 2014 on public procurement and repealing Directive 2004/18/EC, [2014], OJ L 94/65.

[819] European Commission, High-Level Expert Group on Artificial Intelligence (n 813) 15–16.

In order to address these harms, the procedural dimension of fairness assures data subjects' ability to contest and thus to search for an effective remedy for the unfair treatment received by the AI systems employed.[820] In this perspective, the fairness of automated decision-making processes is strictly dependent, from a procedural standpoint, on their accountability and transparency features, and in terms of explainability.[821]

31.3 Fairness and data protection law

The principle of fairness in data processing is mentioned in the EU Charter of Fundamental Rights. Article 8(1) and (2) state that '1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law'.

Fairness has been defined by the European Data Protection Supervisor as 'a core principle of data protection law', alongside the lawfulness and transparency principles.[822] This has been directly acknowledged by Article 5(1) of the General Data Protection Regulation[823] which requires processing activities to be carried out 'lawfully, fairly and in a transparent manner'.

The interconnection between fairness, lawfulness and transparency has its roots in the Data Protection Directive,[824] where Articles 10 and 11 state data subjects' right to receive among others information about the identity of data controllers and the purposes for data processing, thus guaranteeing 'fair processing'

in respect of the data subject. The interconnection between these principles is affirmed under recital 39 GDPR, declaring that 'any processing of personal data should be lawful and fair', and requiring data controllers 'to provide all the relevant information that enable a fair and transparent processing'. It is also acknowledged by Article 29 Working Party in two Opinions on consent[825] and on transparency.[826]

In the first Opinion, the Working Party clarifies that the fairness principle requires a careful identification and evaluation of the lawful basis of the enacted processing activities. In these terms, it would be unfair to ground a processing activity on a basis different from that which is declared to the data subject. Moreover, it would also be unfair to collect data that is not necessary for a specified purpose of the processing.[827]

Although these principles are strictly related to each other, the principle of fairness has an autonomous stand in respect to the other two. Indeed, a specific processing operation may be, on the basis of *ex ante* assessment, lawful or transparent having fully complied with the mandatory legal requirements (such as the informed consent rule) but can result, from an *ex post* perspective, in unfairness.

Although there is no definition as such of the principle of fairness in positive European data protection law, fairness of the processing is related to the balancing of competing interests and more precisely of fundamental rights and freedoms of the subjects involved, that is, on the one side the data subject and on the other the data controllers and processors.[828] More precisely, fairness requires that in pursuing their data-processing objectives, data controllers and processors take into consideration

---

820 Comandé (n 805) 169–206.
821 Kroll et al. (n 805) 633–705.
822 European Data Protection Supervisor, 'Opinion on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (2016) https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf.
823 Council Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), [2016], OJ L 119/1.
824 Council Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, [1995] OJ L 281/31.

825 Article 29 Data Protection Working Party, 'Guidelines on Consent under Regulation 2016/679' (2018). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051.
826 Article 29 Data Protection Working Party, 'Guidelines on Transparency under Regulation 2016/679' (2018). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227.
827 Article 29 Data Protection Working Party (n 825), 23.
828 G. Malgieri, 'The Concept of Fairness in the GDPR – A Linguistic and Contextual Interpretation', *FAT* *20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency January 2020* (2020) 154–66, https://dl.acm.org/doi/pdf/10.1145/3351095.3372868.

the interests and reasonable expectations of data subjects. This has been made clear by the Article 29 Working Party, which has underlined that transparent data processing means data processing that is expectable and foreseeable 'considering the circumstances and the context of each situation'.[829]

Hence, for a 'fair' protection of these interests, data controllers may need to go beyond the minimum legal requirements.[830] This is well reflected by the difference between formal and substantial fairness, the former relating to the communication to data subjects of the information that the GDPR requires to be transparent, as under Articles 13(1) and 14(1) GDPR the categories of the data recipients; the latter requiring the communication of more specific information based on the peculiar circumstances of the case, such as the actual names of the data recipients. The criterion of fairness is particularly relevant, for example, in cases of further processing of personal data: the fairness principle in such cases imposes that the more intrusive the processing is, the earlier the data subject needs to be informed.

According to the Working Party, in cases where a data controller accomplishes only the strict normative requirements, it has to demonstrate why it has not provided more information. In this perspective, the concept of fairness is linked to 'specific processing' situations.[831]

Specific processing situations are outlined in Chapter IX of the GDPR, which identifies particular hypotheses of conflicts of interests in which the protection of personal data could impact on other interests such as freedom of expression and information, public access to official documents, archiving and research purposes, freedom of religious entities to process personal data or on particular legislative areas requiring specific national intervention, as for example employment law. Under Article 6(2) and (3) GDPR the adoption by member states of measures ensuring a lawful and fair processing should take as an example the 'balancing' provisions listed in Chapter IX.

'Fair substantial balancing' is of a theoretical nature and relies on the particular context in which it is applied. From an operational perspective, fair balancing incorporates the principles of proportionality and necessity, which are explicitly recognised in the GDPR and more precisely in Article 6(1) GDPR: the necessity and proportionality criteria substantiate the fair balancing test and enable the assessment of the appropriateness of controllers'/processors' actions, through the evaluation of the peculiar circumstances in which the processing operation has occurred.

Against this backdrop, the principle of fairness carries out an overarching function of rebalancing of the data subject–controller relationship in case the collection and processing of personal data undermines data subjects' interests.[832] In this perspective, the fairness criterion assures the protection of data subjects from controllers'/processors' abuse, by preventing disproportionate harms stemming from the power asymmetries that characterise the current massive processing environment. Hence, the fairness principle ultimately provides adequate protection of the autonomy and self-determination of data subjects.

This is to be concretely achieved through the adoption by the data controller of additional safeguards that can effectively counterweight the unfair balance between data controllers and data subjects. Data controllers can also be law enforcement authorities processing data under a necessity and proportionality principle. Accordingly, the same data protection authorities are required under recital 129 GDPR to exercise their powers 'impartially, fairly and within a reasonable time'.

At a deeper level, thus, the notion of fairness is to be linked with predictability and legal certainty of the consequences of data processing activities.

From a procedural standpoint, the fairness principle guarantees that data subjects are active market participants and not mere recipients of corporations' machine-driven processing operations. This means that data subjects have a proactive role in protecting their legal position vis-à-vis data controllers/processors and in neutralising existing imbalances. For these purposes, the principle of fairness grounds data

[829] Article 29 Data Protection Working Party (n 825) 17–18.

[830] W.J. Maxwell, 'Principle-based Regulation of Personal Data: The Case of "Fair Processing"', *International Data Privacy Law* (2015) 5, 3, 205–16.

[831] Article 29 Data Protection Working Party (n 825) 7.

[832] D. Clifford and J. Ausloos, 'Data Protection and the Role of Fairness', *Yearbook of European Law* (2018) 37, 130–87.

subjects' specific rights, such as the right to erasure under Article 17 GDPR; the right to object under Article 21 GDPR and the right not to be subject to automated decisions under Article 22 GDPR. These rights overall form a system of checks and balances through which data subjects counteract the power asymmetries between controllers or processors and data subjects.

From the controllers'/processors' perspective, the fairness principle demands that processing operations do not infringe data subjects' fundamental rights and freedoms, in particular freedom from unfair discrimination. In these regards, recital 71 GDPR, requires the controller to 'implement technical and organisational measures appropriate to ensure, in particular, that … the risk of errors is minimised' and to 'secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subjects and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect'.

By so stating, recital 71 GDPR associates fairness of algorithmic processing with the non-discrimination principle. This interpretation of fairness has been reaffirmed by the Article 29 Working Party in its Opinion on automated decision-making, where it observed that the practice of profiling 'may be unfair and create discrimination' in the allocation of deals, 'for example by denying people access to employment opportunities, credit or insurance, or targeting them with excessively risky or costly financial products'.[833] The link between fairness and non-discrimination has also been affirmed by the French Data Protection Authority, which has declared that a 'fair algorithm should not end up generating, replicating or aggravating any form of discrimination'.[834]

A direct concretisation of fairness as non-discrimination is given by the data sanitisation requirements[835] under Article 9(1) GDPR, prohibiting the processing of special categories of personal data, as data concerning health, and Article 22(2) GDPR banning decisions 'based solely on automated processing, including profiling … which is based on the special categories of personal data referred to in article 9'. On these bases, procedural fairness is directly related to the more formal and process-oriented data protection requirements, in respect to which data controllers must provide evidence of compliance and thus have the burden of proof to justify that their actions are in compliance with the fairness principle. The external dimension of fairness is thereby assured by the accountability principle.

31.4 Fairness and enforcement between data protection, consumer and competition law

The notion of fairness, as a rebalancing tool with regard to potentially discriminating data processing enquiries, crosses European law and substantively shapes the application not only of data protection law but also of consumer law and – although less apparently – of competition law.

Regarding consumer law, indeed, contract terms and commercial practices that are grounded in the use of personal data need to conform to European consumer protection requirements as established under the Unfair Terms Directive[836] and the Unfair Commercial Practices Directive.[837]

(CNIL) as part of the ethical discussion assignment set by the digital republic bill December 2017 (2017) https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf.

[835] B.W. Goodman, 'A Step towards Accountable Algorithms? Algorithmic Discrimination and the European General Data Protection', 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, (2016) http://www.mlandthelaw.org/papers/goodman1.pdf.

[836] Council Directive (EU) 2019/2161 of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules, [2019], OJ L 328/7.

[837] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning

[833] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the purposes of Regulation 2016/679' (2018) https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

[834] Commission nationale de l'informatique et des libertés, 'How Can Humans Keep the Upper Hand? The ethical matters raised by algorithms and artificial intelligence', Report on the public debate led by the French data protection authority

From a further perspective, the massive collection of users' personal data through machine learning processes is generating new market imbalances between users and economic actors. This in turn engenders the risk of market abuses in the form of the imposition of unfair terms considered under Article 102 TFUE.[838]

In such scenarios, both from a consumer and a competition law standpoint, the notion of fairness structures different dimensions of consumer welfare, as deeply connected to the protection of consumers' freedom to decide and choose among different contractual terms and different products or services offered on the market. The fairness principle laid down by data protection law, both at substantive and procedural levels, ensures that data subjects are effectively informed and thus empowered in order to exercise such a choice.

Recital 42 GDPR states that pre-designed forms of data subject consent should not contain unfair terms. In addition, national consumer protection authorities are applying the Unfair Commercial Practices Directive[839] to assess the fairness of the collection of personal data, in particular with regard to the legitimacy of consumer consent.

Conversely, from the perspective of competition law, the special responsibility borne by a dominant firm under Article 102 TFUE, requiring dominant firms to refrain from impairing undistorted competition on the market, could follow data protection law's notion of fairness, whereby the stronger the market position of the data controller is, the stricter the principles of fairness and accountability must be in order to effectively protect data subjects' interests.

In such cases, as Graef et al. have pointed out, 'the notion of fairness can be seen as an overarching principle connecting the three fields as to the way in which the protection of choice as the desired outcome is to be achieved'.[840]

Although this may be true, it needs to be recalled that the different foundations of each of the considered branches of European law also imply different foundations of the same notion of fairness, which is rooted in the fundamental rights dimension when it comes to data protection law and conversely has a more market-related significance in the context of consumer and competition law.

In view of these different values across the three recalled branches of European law, the principle of fairness, and its effects-based approach going beyond formal normative requirements, works as an 'overarching principle' capable, if adequately handled by enforcement authorities, to close existing enforcement gaps affecting the legal framework applicable to data-driven machine learning processes.

GIULIA SCHNEIDER

Bibliography

Article 29 Data Protection Working Party, 'Guidelines on Consent under Regulation 2016/679' (2018). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051.

D. Clifford and J. Ausloos, 'Data Protection and the Role of Fairness', *Yearbook of European Law* (2018) 37, 130–87.

G. Comandé, 'Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems and Opportunities of Algorithms', in T. Cerquitelli, D. Quercia and F. Pasquale (eds), *Transparent Data Mining for Small and Big Data* (2017) New York: Springer, 169–206.

European Commission, High Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

European Data Protection Supervisor, 'Opinion on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (2016) https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf.

B.W. Goodman, 'A step towards Accountable Algorithms? Algorithmic Discrimination and the European General Data Protection', *29th Conference on Neural Information Processing*

---

unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), [2005], OJ L 149/22.

[838] G. Schneider, 'Testing Article 102 TFUE in the Digital Marketplace: Insights from the Bundeskartellamt's Investigation against Facebook', *Journal of European Competition Law and Practice* (2018) 9, 4, 213–25.

[839] Council Directive 2005/29/EC (n 837).

[840] I. Graef, D. Clifford and P. Valcke, 'Fairness and Enforcement: Bridging Competition, Data Protection and Consumer Law', *International Data Privacy Law* (2018) 8, 3, 200–23.

Systems (NIPS 2016), Barcelona, Spain (2016) http://www.mlandthelaw.org/papers/goodman1.pdf.

I. Graef, D. Clifford and P. Valcke, 'Fairness and Enforcement: Bridging Competition, Data Protection and Consumer Law', *International Data Privacy Law* (2018) 8, 3, 200–23.

P. Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law', *Common Market Law Review* (2018) 55, 4, 1143–85.

J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson and H. Yu, 'Accountable Algorithms', *University of Pennsylvania Law Review* (2016–17) 165, 3, 633–705.

B. Lepri, O. Nuria, E.F. Letouzé, and P. Vinck, 'Fair, Transparent and Accountable Algorithmic Decision-Making Processes: The Premises, the Proposed Solutions and the Open Challenges', *Philosophy & Technology* (2018) 31, 611–27.

G. Malgieri, 'The Concept of Fairness in the GDPR – A Linguistic and Contextual Interpretation', *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency January 2020* (2020) 154–66, https://dl.acm.org/doi/pdf/10.1145/3351095.3372868.

W.J. Maxwell, 'Principle-based Regulation of Personal Data: The Case of "Fair Processing"', *International Data Privacy Law* (2015) 5, 3, 205–16.

L. Naudts, 'Towards Accountability: The Articulation and Formalization of Fairness in Machine Learning', *FIP Summer School on Privacy and Identity Management 'Fairness, Accountability and Transparency in the Age of Big Data'* (20–24 August 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3298847.

G. Schneider, 'Testing Article 102 TFUE in the Digital Marketplace: Insights from the Bundeskartellamt's Investigation Against Facebook', *Journal of European Competition Law and Practice* (2018) 9, 4, 213–25.

S. Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising', *Berkely Technology Law Journal* (2020) 35, 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388639.

32   Forgotten (Right to Be)

32.1 Introduction: Article 17 of the EU GDPR

The right to be forgotten (RTBF) is established in law in Article 17 of EU's General Data Protection Regulation (Regulation 2016/679, the GDPR). Nevertheless, the exact title of Article 17 is 'Right to erasure ("right to be forgotten")'. Several assumptions can be derived from the formulation of this title. First, Article 17 is mainly about the right to erasure and not about the RTBF; second, that an RTBF should not be considered as an unqualified right, in view of the parentheses around it. Third, that the coexistence of 'erasure' and 'forgetting' in the same sentence invites philosophical analysis (and, thus, legal uncertainty).

The above assumptions are vindicated after examination of the actual provisions of Article 17 of the GDPR. First and foremost, Article 17 is indeed about a right to have personal data deleted and not about any 'right to be forgotten': its paragraph 1 expressly states that 'the data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay'. The specificity of the wording leaves little doubt as to the exact scope of Article 17 of the GDPR.

Indeed, ever since the first days of data protection in Europe back in the 1970s one of its basic premises, together with a special set of principles and establishment of a specialised new state agency, was the award to individuals of a special set of rights to assist them in the protection of their personal data: a right to information that personal data are being processed, a right to access these data, a right to object to the processing, and a right to have the data deleted, if appropriate. This list of rights was maintained in the 1995 EU Data Protection Directive (DPD). The GDPR clarified and brought the right to erasure to the fore, assigning it a whole new article, Article 17.

The right to erasure is a critical component of individuals' protection within the GDPR system. In principle, individuals would have had the right to ask controllers to delete their personal data only if controllers breached one of the GDPR's provisions. However, Article 17 goes much further than that. In fact, out of the altogether six grounds listed in the first paragraph of Article 17 upon which an individual may ask for data deletion only one pertains to 'unlawful processing'.

All others expand the options available to individuals, for instance to include cases where 'personal data are no longer necessary' or when 'there are no overriding legitimate grounds for the processing'. In addition, the second paragraph of Article 17 requires that, if the