A criterion-based approach to GDPR's explanation requirements for automated individual decision-making

by **Lea Katharina Kumkar and David Roth-Isigkeit***

Abstract: Automation of decision-making processes represents an essential element of the digital transformation. However, automated data processing based on machine learning methods poses increased threats to the fundamental rights of data subjects. One main reason for this is the fact that tracing and explaining the solution path responsible for a certain machine output requires high technical effort. The new European data protection law provides a framework for explanation requirements that apply to users of the new – automated – technologies. This article outlines the current state of discussion on explanation requirements for automated decisions and advocates a restrictive interpretation of the corresponding provisions in the GDPR.

Keywords: GDPR; artificial intelligence; automated individual decision-making; right to explanation

Recommended citation: Lea Katharina Kumkar and David Roth-Isigkeit, A criterion-based approach to GDPR's explanation requirements for automated individual decision-making, 12 (2021) JIPITEC 289 para 1

A. GDPR and the „right to explanation"

1 Methods of automated data processing are on the rise in public and private spheres. In particular, the interest in machine learning applications has exponentially grown in the last years. Main drivers for this are increased availability of large amounts of data and better computing power. Yet, since the European data protection legislator did not consider the developments towards (full) automation in detail, the relationship between data protection law and new possibilities of automated data processing applications remains largely unclear. The resulting conflicts on existence and possible scope of concrete rights and duties are unfortunate not only for the data controller, but also for data subjects, as legal uncertainty could deter data subjects from asserting their rights.

2 One crucial aspect for the protection of data subjects that the General Data Protection Regulation (GDPR) brought up, yet does not resolve, is the question in how far the controller of automated individual decision-making applications has to fulfil certain explanation requirements.[1] Here, the buzzword of the "right to

* The author Kumkar is an assistant professor of civil law, business law and legal aspects of digitalization at Trier University and an affiliated member of the Institute for Digital Law Trier (IRDT). The author Roth-Isigkeit leads a junior research group and the interdisciplinary SOCAI centre for social implications of artificial intelligence, both at Würzburg University. This piece further develops the argument of a previous article of the authors, published in Juristenzeitung 6/2020, pp. 277-286.

1 Under the umbrella term of "explanation requirements", we include here both the duties to inform under Art. 13(2)(f), 14(2)(g) GDPR and the right to information under Art. 15(1)(h) GDPR as well as a possible right to explanation under Art. 22(3) in conjunction with Recital 71 (4) GDPR. For extensive references to the German commentary literature, see L. Kumkar and D. Roth-Isigkeit, 'Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DSGVO'

explanation" has received particular attention in the literature.[2] Underlying this discussion is the so far unanswered question to what extent users of automated decision-making and recommendation systems must be able to disclose the functioning of the system and, if necessary, also the specific decision-making path for individual cases. This question is key especially for advanced automation applications (such as artificial neural networks or complex decision trees). Here, the outcome may lead to so-called black box constellations.[3] In these, the process that leads the machine to dispense a certain output is – if at all – only traceable with a high level of technical effort.

3 The GDPR provides for special explanation requirements of the data controller for certain cases of automated data processing. According to Art. 13(2)(f), 14(2)(g),[4] when collecting personal data, the data controller shall provide the data subject with information on "the existence of automated decision-making pursuant to Art. 22(1) and (4) and – at least in these cases – meaningful information about the logic involved and the scope and intended effects of such processing for the data subject". Art. 15(1)(h) provides for a corresponding right of access. Furthermore, Recital 71(4) mentions the "explanation of the decision reached" as part of the "suitable safeguards" for automated processing. This leads parts of the literature to accepting the existence of a case-by-case requirement to provide detailed and specific explanations for processing operations.[5]

4 Such an obligation would entail quite dramatic consequences for the way in which data controllers will be able to use automated processing in the future. It would imply considerable financial risks for data controllers, in particular due to the strict penalty provisions in Art. 83(5)(b). In addition, disclosure of the decision path may be not possible due to technical difficulties or (legitimate) interests of the processor in preserving business and trade secrets.[6]

5 Referring to the current discussion on explanation requirements, this paper advocates a restrictive interpretation of the relevant provisions of the GDPR. In contrast to the procedural approaches suggested by large parts of the literature, we propose a criterion-based approach, which requires the *ex ante* disclosure of possible decision criteria, but not the *ex post* disclosure of the detailed process of decision-making and weighing in the individual case. For a better understanding of the relevant disputes, we first outline the general principles on automated individual decision-making pursuant to Art. 22 (B.). This is followed by a description of potential points of reference to derive a "right to explanation" for automated decisions (C.). Building on considerations on function and technical limits (D.), we discuss potential implications of the existence of a "right to explanation" (E.).

(2020) 75 (6) Juristenzeitung 277-286.

2 See, for an introduction to this debate with further references B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143, 189.

3 For the social problem of black box constellations, see e.g. F. Pasquale, 'The Black Box Society – The Secret Algorithms that Control Money and Information' (2015), Harvard University Press, 1-18.

4 The following article denominations refer to the General Data Protection Regulation (GDPR) if not otherwise stated.

5 Notably B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143; M. Brkan, 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond' (2019) 27 Int J Law Info Tech 91; M. Brkan and G. Bonnet, Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas (2020) 11 European Journal of Risk Regulation 18; M. Kaminski, The right to explanation, explained (2019) 34 Berkeley Tech. L.J. 189; T. Kim and B. Routledge, Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach (2021) Business Ethics Quarterly, Fist View 1. The term presumably derives from an initially unpublished conference paper by Goodman/Flaxman, EU Regulations on Algorithmic Decision Making and "a Right to an Explanation," available at https://arxiv.org/pdf/1606.08813.pdf (last accessed June 29, 2021). The paper focused mainly on technical issues and was primarily intended to draw attention to the difficulties of explaining complex algorithmic processes.

6 In this sense, the "qualified transparency" called for by e.g. F. Pasquale, 'The Black Box Society – The Secret Algorithms that Control Money and Information' (2015), Harvard University Press, 140 ff. should also be understood as a balancing between different interest groups. See further on the challenges of trade secret protection in the data-driven economy, A. Wiebe and N. Schur, 'Protection of trade secrets in a data-driven, networked environment – Is the update already out-dated?' (2019) 14 (10) Journal of Intellectual Property Law & Practice 814–821.

B. The legal framework for automated individual decision-making pursuant to Art. 22 GDPR

6 Art. 22 provides a framework for automated individual decision-making and profiling measures. A similar provision was already provided for in Art. 15 Data Protection Directive (DPD).[7] Pursuant to Art. 22(1), the data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. Paras. 2 and 3 provide for exceptions to this principle, in particular in situations where the data subject has given consent. Para. 4 sets out specific requirements for particularly sensitive data (cf. Art. 9(1)).

I. Content and Meaning

7 Art. 22 does not establish a separate basis of permission for the processing of personal data, but establishes an additional prerequisite that must be observed during processing. Despite its systematic localization among the data subjects' rights, the provision – at least indirectly – has the character of a prohibition. Processing that does not comply with the requirements of Art. 22 is prohibited even without explicit statement by the data subject.

8 An even more comprehensive prohibition is provided for in Art. 22(4) for special categories of personal data, which according to Art. 9(1) include in particular data revealing racial or ethnic origin, political opinions, religious beliefs, trade union membership, genetic predispositions or health status and sexual orientation. Deviating from other requirements of Art. 22, the inclusion of data pursuant to Art. 9(1) in automated decisions is generally only permissible if the data subject has expressly consented (Art. 22(4) in conjunction with Art. 9(2)(a)), or if the processing is both necessary due to substantial public interest and is proportionate (Art. 22(4) in conjunction with Art. 9(2)(g)).

II. Profiling and automated decision-making

9 The wording of the official title of Art. 22 ("automated individual decision-making, including profiling") is misleading. "Profiling" is not a specific application of automated decision-making, but a special data processing operation.[8] According to Art. 4 (4) profiling includes "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

10 "Automated decisions" within the meaning of Art. 22 build on a data processing operation by linking automated processing with consequences for the data subject. This understanding is also supported by the fact that Art. 22(1) mentions "profiling" as an example following the term "automated processing" – and not the subsequent "decision". Profiling is therefore only one possible manifestation of the processing covered by Art. 22.[9] Regarding the concrete scope of the prohibition contained in Art. 22, there is agreement that neither the profiling process nor the automated processing as such is covered, but only the decision based *solely* on this automated processing – at least if and to the extent that this has legal or similarly significant adverse effects.

11 The GDPR does not define the term "automated individual decision-making." Yet, it can be assumed that it makes a distinction between "decisions" in a narrow sense as opposed to "processing". Not every type of data processing automatically qualifies as a decision within the meaning of Art. 22(1). Rather, a minimum degree of complexity must be inherent, since otherwise even simple if-then connections such as dispensing money at an ATM would fall under the regulation.[10] That such is not intended, also becomes clear from the comparison with profiling, as mentioned in Art. 22(1). Profiling aims

7 Directive 95/46/EC. For the development of the requirements of Art. 22 GDPR from Art. 15 DPD, see also I. Mendoza and L. Bygrave, in: T.E. Sydodinou et al. (eds.), *EU Internet Law* (2017), Springer International Publishing, 77 ff.

8 Expressing its favor for the independence of the two terms: Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 8: "Automated decisions can be made with or without profiling; profiling can take place without making automated decisions."

9 Differently I. Mendoza and L. Bygrave, in: T.E. Sydodinou et al. (eds.), *EU Internet Law* (2017), Springer International Publishing, 77 ff., 90 f., identifying a drafting error and stating the provision should be understood as referring only to profiling.

10 S. Schulz in P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 21; B. Buchner in J. Kühling and B. Buchner (eds.), *DSGVO* (2020), C.H. Beck, Art. 22 para 18.

at the evaluation of personality traits and implies a certain materiality of the processing.[11]

12 Consequently, a decision only exists in the case of an act that selects from (at least) two variants and has a final impact on the external world, which can be attributed to a (natural or legal) person. According to *Bygrave*, a decision means that a "particular attitude or stance is taken towards a person and this attitude/stance has a degree of binding effect in the sense that it must— or, at the very least, is likely to— be acted upon."[12]

13 According to the wording of Art. 22(1), the decision must be based on "solely automated" processing, which means that the decision is taken "without any human intervention", as also clarified in Recital 71.[13] This means that Art. 22 does not apply if the knowledge gained in the course of automated processing is only used as basis or for the preparation of a decision to be taken by a natural person. Here, the natural person involved has to apply a margin of discretion.

14 If an actual review of the content takes place by a human employee with the corresponding decision-making powers to change the processing result, the automated data processing only becomes the (working) basis for the decision of a natural person, and is thus no longer "solely" automated.[14] The situation is different if the result found by the machine is merely accepted by a human administrator without any examination of the content. Also, random checks or interventions in neural networks to improve decisions, such as in supervised learning, do not constitute sufficient human intervention. Since the content of the decision remains unchanged, the situation merely resembles a "maintenance" of the system.[15]

III. Legal effect

15 Art. 22 covers only automated decisions with legal effects or the ones that "similarly significantly affect" the data subject. While the GDPR does not explicitly specify when a decision is to be considered as having "legal effects," it can be assumed that this implies that the legal status of the data subject is altered in any way.[16] Assessing this in more detail, however, much remains unclear. For example, the question arises as to whether this includes only adverse decisions so that (purely) favorable legal consequences remain outside the scope of the provision.[17] A general definition of "similarly significantly affected" has neither yet emerged. However, there is wide agreement that the threshold of mere nuisance must be exceeded.[18]

11 An interim definition of profiling is contained in Recital 24 (2) GDPR where it reads "In order to determine whether a processing activity can be considered to monitor the behaviour of data subjects, it should be ascertained whether natural persons are tracked on the internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes."

12 For a detailed discussion, see L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532.

13 However, the occasionally expressed demand that this implies the absence of human intervention from the collection of the data to the issuing of the decision must be rejected. Here, only the decision-making process in the narrower sense is decisive. The dangers emanating from automated decisions are not less serious for the person concerned if the preceding data collection is (still) manual or only partially automated. Only this kind of understanding does satisfy the comprehensive protective purpose of Art. 22 GDPR.

14 See L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532-533.

15 T. Hoeren and M. Niehoff, 'KI und Datenschutz – Begründungserfordernisse automatisierter Entscheidungen' (2018) 9 *Rechtswissenschaft* 47, 53.

16 L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532.

17 Against this, it could be argued that the wording of Art. 22(1) GDPR does not contain a corresponding restriction. On the other hand, the protective purpose of Art. 22 GDPR contradicts the inclusion of favorable legal consequences, as the data subject does not need to be protected from (purely) favorable decisions.

18 Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 21: "For data processing to significantly affect someone the effects of the processing must be sufficiently great or important to be worthy of attention." Another subject of discussion is the extent a legal effect can be assumed for automated decisions in contractual relationships. While the legal effects in the case of termination and acceptance of contractual offers are predominantly affirmed, opinions are divided in the case of refusal to conclude a contract (under certain conditions). It is convincingly argued against the existence of a legal effect within the meaning of Art. 22(1) GDPR in the cases of a refusal to conclude a contract or a refusal to accept certain conditions that legal effects do not "unfold" as intended but, on the contrary, do not occur at all.

IV. Exceptions and suitable measures to safeguard

16 Art. 22(2) provides for three exceptions to the prohibition of Art. 22(1). The norm does not apply if a decision is necessary for entering into, or performance of, contract between the data subject and the data controller (lit. a), if a decision is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests (lit. b) or if the decision is based on the data subject's explicit consent (lit. c).[19]

17 If the data controller intends to base the data processing on one of the exceptions under Art. 22(2)(a) or 22(2)(c), the controller shall, according to Art. 22(3), implement suitable measures to ensure that the data subject is provided with adequate safeguards. This includes at least that the data subject has (1) the right to obtain personal intervention on the part of the controller, (2) the opportunity to put forward his or her own point of view, and (3) the right to contest the decision (so-called minimum safeguards). As can be seen from the wording of Art. 22(3), the aforementioned list is not exhaustive, i.e. the "suitable measures to safeguard" to be taken by the controller may also require further measures.

C. Possible starting points for explanation requirements

18 Some argue that Art. 22(3), in conjunction with Recital 71(4), provides a "right to explanation" for the data subject, which is intended to apply comprehensively and retrospectively to the entire individual decision-making process (I.) Other authors assume that the data controller's duty to explain can be derived solely from the general information rights of Arts. 13 to 15 (II.). While the two approaches differ significantly in terms of scope and timing of the explanation requirement, they both largely leave open the required content and depth of the explanation (III.).

19 According to Art. 4 No. 11 GDPR consent means "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."

I. A „Right to explanation" pursuant to Art. 22 (3) in combination with Recital 71 of the GDPR

19 In the context of the rights of data subjects pursuant to Art. 22(3), it is being discussed whether a "right to explanation" against the data controller in the form of a case-by-case requirement to justify is being established.[20] Indications for the existence of such a right or – correspondingly – an equivalent requirement to explain on the part of the controller are not to be found directly in Art. 22. The provision only mentions the right to intervention of a person, explanation of one's own position and contestation of the decision. Yet, an indication could be found in Recital 71, which states:

> *"However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, [...], or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."*

20 In this context, the wording of Recital 71(4) suggests that the controller is obliged to justify the specific decision *ex post* and on a case-by-case basis, when it refers to "obtaining an explanation of the decision reached after such assessment". The explanation should therefore not only include the abstract functionality of the device used,[21] but also a justification of the concrete decision in the individual case.[22]

21 However, the question arises in which cases the requirement could be applied at all. This appears problematic because the explanation requirement is *only* contained in the recitals and does not find a counterpart in the wording of Art. 22(3). In particular, referring to the lack of binding effect of the recitals, *Wachter et al.* took the view that a right to explanation is currently not legally imposed by

20 See references in note 5.

21 This is the case with the explanation requirements pursuant to Art. 12 to 15 GDPR, cf. below. C. II.

22 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 81.

Art. 22.[23] They see this supported by the fact that the requirement to explain specific individual decisions was omitted from the normative text of Art. 22 during the deliberations on the drafting of the GDPR. This would suggest that the legislator did not intend to make such a right binding. Even though *Wachter et al.* consider it possible that case law will establish a right to explanation in the future as part of the interpretation of the "adequate safeguards"[24] they do not see it currently imposed by the GDPR.[25]

1. A „Right to explanation" as minimum guarantee?

22 This view is convincing – at least against the backdrop of the current practice of the European Court of Justice (ECJ) on the status of the recitals in the interpretation of substantive guarantees. In European legal acts, recitals are placed before the determining norms as an anticipated statement of reasons (cf. the usual introductory wording "considering the following reasons"). The ECJ has consistently held that "[...] the preamble to a Community act has no binding legal force and cannot be relied on either as a ground for derogating from the actual provisions of the act in question or for interpreting those provisions in a manner clearly contrary to their wording".[26] Starting point and limitation of a teleological interpretation based on the recitals is thus always the wording of the norm in question. In a decision from 1989, the ECJ clarified – albeit with regard to the recitals of a regulation – that a recital "may cast light on the interpretation to be given to a legal rule, it cannot in itself constitute such a rule."[27]

23 For a "right to explanation", the wording of Art. 22(3) is the decisive limit. The enumeration of the "minimum safeguards" to be guaranteed by the controller in Art. 22(3) is exhaustive. Here, if the European legislator had wanted to make the minimum safeguards open-ended, it would have expressed this – following good custom – by adding words such as "for instance", "for example" or "in particular". A non-exclusive understanding would ultimately also undermine the goal of creating binding minimum standards for data controllers and data subjects (cf. Recital 10(1)). There is thus no room for a broadening teleological interpretation. According to this understanding, Art. 22(3), in conjunction with Recital 71(4), does not generally provide for a "right to explanation" in the form of a minimum guarantee.

2. A „Right to explanation" as suitable measure?

24 However, the fact that a "right to explanation" is not mentioned in the enacting terms of Art. 22(3) does not necessarily suggest that a requirement to explain could not exist in any conceivable case.[28] This is because the rights of the data subject are not exhausted by the (minimum) rights explicitly mentioned in Art. 22(3) – as the statutory use of the word "at least" suggests. Rather, according to Art. 22, the data controller must take all reasonable measures necessary to safeguard the rights and freedoms as well as the legitimate interests of the data subject. This does not exclude, at least not systematically, that the "reasonable measures" could in some cases also include an *ex post* and case-by-case explanation of the decision.

25 Art. 22 suggests that the legislator did not intend to include the explanation requirements among the measures to be taken in *every case* to protect the data subject. Rather, they belong to the group of "suitable measures" that go beyond the minimum guarantees. This means whether the explanation requirements under Art. 22(3) apply in an individual case depends on the broader question in which cases the explanation is considered necessary to protect the rights and freedoms as well as the legitimate interests of the data subject. This depends very significantly, on which function can be attributed to the "right to explanation" in the overall structure of the legal protection of data subjects.[29]

23 Ibid., 79 ff.

24 Ibid., 81.

25 Ibid., 80.

26 Case C-345/13 *Karen Millen Fashions* [2014] ECLI:EU:C:2014:2013 para 31; see also Case C-136/04, *Deutsches Milch-Kontor* [2005] EU:C:2005:716 para 32.

27 Case C-215/88, *Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung* [1989] ECLI:EU:C:1989:331 para 31; see also T. Klimas and J. Vaiciukaite, 'The Law of Recitals In European Community Legislation' (2008) 15 ILSA Journal of International & Comparative Law 92 f. and H. Rösler in J. Basedow, K. Hopt and R. Zimmermann (eds.) *Max Planck Encyclopedia of European Private Law* (2012), Oxford University Press, 979 ff.

28 Likewise M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 204.

29 *Infra*, D. I.

II. Information rights according to Arts. 13 to 15 GDPR

26 While the "right to explanation" according to Art. 22(3) in conjunction with Recital 71(4) is in dogmatically uncertain territory, the mandatory nature of the information rights in Arts. 13 to 15 is (at least) *ipso iure* beyond question. According to Art. 13(2)(f) and 14(2)(g), the data controller shall provide the data subject with information on "the existence of automated decision-making, including profiling, referred to in Art. 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.". Art. 15(1)(h) grants the data subject a corresponding right of access against the data controller. On closer examination, however, several details remain unclear.

1. Scope of the Provisions

27 Since the material preconditions of Art. 13(2)(f), 14(2)(g) and 15(1)(h) explicitly refer to Art. 22(1) and (4), the question arises whether the requirements apply solely in the (narrow) cases of automated individual decision-making which are also covered by the preconditions of Art. 22(1); i.e. whether a decision with legal effects or a similarly significant impairment is always required. The wording "at least in these cases" in Art. 13(2)(f), 14(2)(g) and 15(1)(h) could suggest the provisions cover (automated) processing below the threshold of "decision". However, it remains completely undefined according to which criteria such further cases are to be determined. Against the backdrop of the strict penalty for the information duties (Art. 83(5) (b)), it seems unconvincing to extend the information duties to other processing operations.

28 Rather, it must be assumed that the wording was simply copied from the preceding provision in Art. 12 lit. a 2nd Alt. DPD. Yet, while the wording in the DPD was intended to give the Member States room for manoeuvre in implementation, it cannot fulfil this function in the (directly applicable) GDPR. This means with regard to the rights and obligations under Art. 12 to 22, Member States only keep the competence to restrict, but not the right to extend. [30] The fact that only Art. 22 (1) and (4) are explicitly referred to (but not Art. 22(3)) also indicates that the requirement to provide information does not extend to the minimum guarantees contained in Art. 22(3).[31]

2. Relevant timing

29 The relevant timing of the information differs between the various provisions. Pursuant to Art. 13(2)(f), the information must be provided "at the time when personal data are obtained." In light of Art. 14(2)(g), the information must be provided pursuant to Art. 14(3), namely "within a reasonable period after obtaining the personal data, but at the latest within one month" (lit. a). However, if use of the data for communication with the data subject (lit. b) or disclosure to another recipient (lit. c) is intended beforehand, this triggers an immediate obligation to provide information from the time of first communication or disclosure. The right to information pursuant to Art. 15(1)(h), on the other hand, is not limited to the moment of data collection, but can also be exercised after the conclusion of the data processing or the automated decision resulting therefrom.

3. Implications for the content of information requirements

30 In the case of Art. 13(2)(f), relevant inference on the content of the information to be provided can be drawn from the time at which the information is provided. Since the information must be provided at the time of the data collection, i.e. before the actual processing operation, the obligation to provide information in Art. 13(2)(f) cannot be directed at a (subsequent) explanation of the processing operation, but is exhausted in the mere announcement of the forthcoming automated decision. [32] It can be further concluded from this that the declaration of the data controller in the sense of the logic of the norm must also only take into account the general functioning of the decision-making and not the (not yet

30 M. Martini, 'Blackbox Algorithmus' (2019), Springer, 182 f. In cases that cannot be subsumed under Art. 22 (1) GDPR, information can nevertheless be provided on a voluntary basis.

31 Ibid., 187, arguing in favor of an addition in this regard. However, some of the German commentaries derive a corresponding obligation to provide an explanation directly from Art. 22(3) GDPR, see S. Schulz, in: P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 41 f.

32 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 82. Cf. M. Martini, 'Blackbox Algorithmus' (2019), Springer, 191.

determined) specific circumstances of the (still imminent) individual decision.[33]

31 For Art. 14(2)(g) and 15(1)(h) the relevant points in time do not allow for such a conclusion on the content of the information requirement. The information can also be provided after processing with a concrete processing result already available. It could be argued that although the wording of the provisions of Art. 14(2)(g) and Art. 15(1)(h) is the same as in the case of Art. 13(2)(f), the "meaningful information" covered varies depending on the point in time at which the information is provided. Thus, especially the right of access in Art. 15(1)(h) could potentially cover information on the specific circumstances of the individual decision.[34]

32 However, upon closer examination this is not convincing. The information requirements pursuant to Art. 14(2)(g) and 15(1)(h) cannot be attributed to a broader content than the obligations in Art. 13(2)(f). This is not only supported by the fact that the wording of the norms is identical, but also by the circumstance that according to Art. 14(2)(g) and 15(1)(h), only information on the "intended" effects must be provided, which suggests a future orientation. [35] This interpretation corresponds to the assumptions made in the guidelines of the Art. 29 Data Protection Working Party.[36] In summary, it can be stated that Art. 13 to 15 – unlike the "right to explanation" derived from Art. 22(3) – require a prior declaration by the data controller, which is directed at the abstract functionality of the data processing.

III. The search for a common ground in explanation requirements

33 In light of the above, there are differences between the two approaches on explanation requirements for automated decisions. Art. 22(3) in combination with Recital 71(4) intends a *subsequent* explanation of the *specific* decision. Art. 13(2)(f), 14(2)(f), 15(1) (h), on the other hand, require a *prior* explanation of the functionality of the data processing and thus provide for *abstract* information rights. From this perspective, there is no connection between the different explanation requirements.

34 For both, the data subject and controller, such a conclusion seems unrealistic from the perspective of practical data protection. Irrespective of their temporal validity and scope, both requirements concern a common basic question: What level of explanation must the controller of automated data processing (be able to) provide? What information about the data processing must (be able to) be shared with the data subjects? The GDPR is silent on the concrete content of these requirements – and yet endows them with the threat of a hefty fine (see Art. 83(5)). It is therefore crucial to develop a pragmatic standard that both users and data subjects can use as a guideline when providing or requesting explanations for automated data processing and that at the same time fulfils the legal demands of both, Art. 22(3) in conjunction with Recital 71(4) as well as Arts. 13(2)f, 14(2) (f) and 15(1)(h).

D. A joint answer to the required explanation depth for automated decision-making

35 The proposal presented here attempts to combine these requirements in order to develop a joint answer to the question of the necessary depth of explanation based on the previous considerations. With respect to the basic functions of the explanation requirements (I.) and the technical limitations of the traceability of automated decisions (II.), it seems reasonable to limit explanation requirements to outlining the decision criteria that form the basis of the (planned) automated processing (III.).

I. A functional view on explanation requirements

36 Looking at the explanation in connection with automated decisions from a functional perspective, similar purposes can be identified in the cases

33 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 78 ff., who distinguish between *system functionality* (ex ante and ex post) and *specific decision* (ex post).

34 This is argued in particular in the German commentary literature, cf. M. Bäcker in Kühling/Buchner DS-GVO BDSG (2020), Art. 15 Rn. 27 with further references.

35 Cf. S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 83. M. Martini, 'Blackbox Algorithmus' (2019), Springer, 192.

36 Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 26.

of Art. 22(3) in conjunction with Recital 71(4) as well as Art. 13(2)(f), 14(2)(f), and 15(1)(h).

1. „Legibility" of the decision

37 First of all, the explanation enables the data subjects to understand the basis of the (automated) decision. This can be derived from the requirement in Art. 12(1) on how the information should be provided, i.e. "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". Even automated decisions within the limits of Art. 22, which are permissible in principle, entail an increased risk of non-transparency for the data subject. Since the data subject usually has no knowledge on how the upcoming decision will be taken, it is difficult for him or her to assess in advance what risks to his or her data will be associated with the planned processing.[37] Without knowledge of the decision-making process, it will be impossible to control whether a decision may be linked to inadmissible criteria, such as a feature of Art. 9 or the non-discrimination criteria of Art. 21 of the European Charter of Fundamental Rights.

38 The wording of Art. 13(2)(f), 14(2)(f), 15(1)(h) ("meaningful information about the logic involved") is indicative. "Meaning" takes the perspective of the understanding data subject, who should be enabled to draw conclusions about the essential decisional factors from the transmitted information.[38] These contexts of meaning must – which does not seem obvious from the formulation – be available in a form that is comprehensible to humans.[39]

39 Following this line of argumentation, *Martini* adopts a narrow understanding of the explanation requirement:[40] According to him, "explanation" means describing the content of the decision in more detail, but not disclosing the reasons for the decision to its full extent. The phrase "an explanation of the decision reached" refers grammatically to the "individual presentation of the case" of the person concerned. This means in consequence that the right to an explanation only exists to the extent that it is necessary in order to explain to an individual how his or her own point of view has been taken into account in the decision and why the result of the assessment has turned in that specific way.

40 Such an understanding of "explanation" requires outlining the essential basis for the decision in a form that is comprehensible to humans, and thus a kind of "legibility".[41] In this way, the information contributes to the data subject's autonomy that had been endangered through the opacity of processing. With *Bygrave*, one could understand this as a requirement of a concept of cognitive sovereignty pervasive in data protection law, "a human being's ability and entitlement to comprehend with a reasonable degree of accuracy their environs and their place therein."[42]

41 Neither disclosure of the raw data nor the technical aspects of the decision-making mechanism would meet this requirement, because the person concerned usually does not have the technical means to put it into a comprehensible form. Examining the explanation requirement from the perspective of the data subject, it soon becomes clear that the literature opinion that asks for a complete breakdown of the decision program or disclosure of the algorithm to fulfill this requirement misses this aspect.[43] From a functional perspective, only those considerations can be covered by the explanation requirement that contribute to a (human) "legibility" of the automated decision.

2. Due process

42 Further, explanation requirements stand in connection with the right to challenge the (automated) decision, which is highlighted as a "minimum guarantee"[44] in Art. 22(3). In this context, the scope of the explanation requirement can be clarified in a similar manner based on the required information

37 See M. Martini, 'Blackbox Algorithmus' (2019), Springer, 176. Likewise M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 211: "They need to be given enough information to be able to understand what they are agreeing to [...]."

38 A. Selbst and J. Powles, 'Meaningful information and the right to explanation' (2017) 7 International Data Privacy Law 233, 239.

39 Ibid., 240.

40 M. Martini, 'Blackbox Algorithmus' (2019), Springer, 191.

41 See also G. Malgieri and G. Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243 ff.

42 L. Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' in Ienca et al. (eds.), Cambridge Handbook of Life Sciences, Information Technology and Human Rights (forthcoming).

43 See e.g. M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 189 ff.

44 Supra, C. I. 1.

for the data subject to effectively make use of this right to challenge.[45] The data subject "must be able to recognize on the basis of this information whether incorrect data has found its way into the procedure or whether the individual particularities of his or her situation have not been sufficiently taken into account."[46] The key aspect here is that the information may be used to raise substantiated objections and to trigger a human review in a second step.[47]

43 The enumeration of rights of the data subject in Art. 22(3) provides further indication on the required depth of explanation. The wording implies a need for suitable measures, "to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."

44 It is then key whether one understands these various aspects as a unit or as separate rights.[48] While the presentation as a list suggests that they are separate, this interpretation is not very plausible, as it would lead to a kind of circle of decisions and challenges. If the data subject's rights under Art. 22(3) could not be advanced uniformly, the data subject would be confronted with a renewed automated decision on the same factual basis after the challenge, against which the challenge would again be admissible.[49] However, it is precisely here that automated decision-making systems are not (yet) capable of automatic self-correction. If the factual basis remains unchanged, the decision will remain unaltered after repeated runs of the system. The "right to challenge" in the common reading of the rights from Art. 22(3) thus only becomes plausible if it demands a human decision *replacing* the automated decision.[50]

45 This argument in turn allows drawing conclusions on the required depth of explanation. In the context of a human re-decision, a subsequent explanation of the original decision path would be superfluous, as the new decision would be taken uninfluenced by the machine output result.[51] For the effective legal protection of the respective person, an explanation of the algorithmic decision-making mechanism is neither necessary nor expedient.[52]

II. Technical limitations regarding the ability to explain automated decision-making

46 Further indications of limited explanation requirements are the technical limitations regarding the ability to explain automated data processing. Particularly in advanced applications of machine learning, the complexity of the system means that it is only possible with the greatest technical difficulty to find a form of explanation that is understandable for humans.

47 Solutions for this problem are discussed under the umbrella topic of "explainable AI".[53] Contemporary advances allow, for example in image recognition by machine intelligence, revealing certain patterns of decision-making, such as determining which pixel patterns were observed for the recognition of

45 S. Schulz, in: P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 42. For a recent application highlighting the goal of public accountability, Talia B. Gillis and Josh Simons, 'Explanation < Justification: GDPR and the Perils of Privacy' (2019) 2 J.L & Innovation 72 (80).

46 Author's translation: P. Scholz, in: Simitis/Hornung/Spiecker gen. Döhmann (eds.), *Datenschutzrecht* (2019), Nomos, Art. 22 para 57.

47 P. Scholz, in: Simitis/Hornung/Spiecker gen. Döhmann (eds.), *Datenschutzrecht* (2019), Nomos, Art. 22 para 57.

48 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 873.

49 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 873.

50 L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 538.

51 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 874.

52 See also L. Edwards and M. Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law and Technology Review 18, 81, who argue that, with respect to the subjective legal protection of data subjects, the traceability of decisions is not the decisive criterion.

53 On the current state of the legal discussion B. Waltl and R. Vogl, 'Explainable Artificial Intelligence–the New Frontier in Legal Informatics' (2018) Jusletter IT (22 February 2018); P. Hackerl et al, 'Explainable AI under contract and tort law: legal incentives and technical challenges' (2020) 28 Artificial Intelligence and Law 415–439; see further A. Deeks 'The judicial Demand for explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829–1850.

certain shapes.[54] In the case of complex deliberation processes, on the other hand, as would be required in the applications discussed here, it is largely unclear to what extent the output result of the work process made visible would be comprehensible to humans. In general, it can be said that we are currently in a state where greater performance of a program corresponds with a reduced comprehensibility of its internal processes. It is therefore not necessarily to be assumed that technical progress will produce explainable data processing, but the opposite of complete opacity is also conceivable, if not likely.

48 The legal value of this technical limitation of the actual comprehensibility of automated data processing is admittedly rather low. It is only suitable to a very limited extent to determine explanation requirements, otherwise one would also fall into a naturalistic fallacy, deriving norms from facts. This principle is also reflected in the GDPR. For example, Recital 58(3) sets particularly high requirements on transparency for situations of high complexity.

49 Nevertheless, technical feasibility can allow conclusions on what the European legislator intended in the context of the explanation requirements. Here it is unlikely – though not impossible – that the GDPR establishes a legal standard that is not technically feasible. In this respect, the above explanations are helpful supplementary information for the interpretation of the standard, which, just like the functional analysis, point to a limited explanation requirement.

III. Consequences for the depth and direction of the explanation

50 Based on these considerations we propose a standard for the depth and direction of the explanation that fulfils two criteria. On the one hand, it ensures the "legibility" of the decision for the data subject and the ability to challenge it. On the other hand, it is technically feasible for the controller. Both conditions indicate that the explanation requirements should be understood in such a limited way that they require an outline of the decision criteria in a form accessible to humans.[55]

51 Decision criteria can help render the mode of operation of a program transparent and traceable. Admittedly, in such a model not all discrimination risks associated with automated data processing may be avoided. Although all decisions can be traced back to the direct or indirect interaction of decision criteria, an all-encompassing control of the decision program in a way that it could be traced how exactly the interaction of individual criteria led to a certain output result is neither technically nor legally feasible.

52 Furthermore, the imposition of a comprehensive requirement to explain the functionality of the data processing and the concrete outcome of the decision would also be questionable from a legal policy perspective. It would create an appearance of controllability of the internal mechanisms of automated data processing and shift burdens of justification onto data subjects.

53 In practical terms, the criterion-based approach advocated here means that the data controller must disclose the (real-world) criteria that the decision program takes into account for its calculations. On the one hand, this imposes a transformational task on the controller to translate the criteria from the digitized form into a linguistic representation. On the other hand, disclosure of the program's concrete mode of operation is not required. Regarding the question of how specific the disclosure of these decision criteria must be, the sanction practice of the data protection supervisory authorities is likely to become a decisive factor for the further development of the law.

E. Implications

54 The view adopted here understands the explanation requirements as a necessary starting point for a human review. If one considers the requirement to present a catalogue of criteria as the basis for this, the term "explanation" (derived from Recital 71 (4)) is misleading, since this represents only the starting point for the intervention directed at a human decision. It is therefore reasonable to assume that the subjective legal asset discussed under the term "right to explanation" actually turns out to be a preparatory *right to justification*.[56]

55 This understanding entails both opportunities and risks.[57] On the one hand, it allows the law to reflect

54 W. Samek, T. Wiegand and K.-R. Müller, 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models' (2018) 1 ITU Journal: ICT Discoveries 39 ff.

55 Differently M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 209 ff., referring to the high value placed on transparency in the GDPR.

56 See, for the conceptual background, R. Forst, 'The Right to Justification' (2007), transl., Columbia University Press.

57 Under certain circumstances, the considerations made here could also gain significance beyond the scope of the GDPR through the so-called Brussels effect. On this point, see B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic

the general opacity of intelligent decision-making systems in order to provide for a practical way of dealing with the limited explicability. It thus offers a possibility for the social integration of technical progress. On the other hand, law thus recognizes the "autonomy" of intelligent decision-making systems to the extent that the procedural and deterministic explanation of decision-making is replaced by the – comparable to legal protection against human decisions – subsequent substantive legality test. Law thus finds its mode of dealing with the non-explicability of machine decisions in converting its procedures to the model of justification adapted to human decisions. Time will show whether this approach will also prove sustainable in practical terms.

Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143, 185.