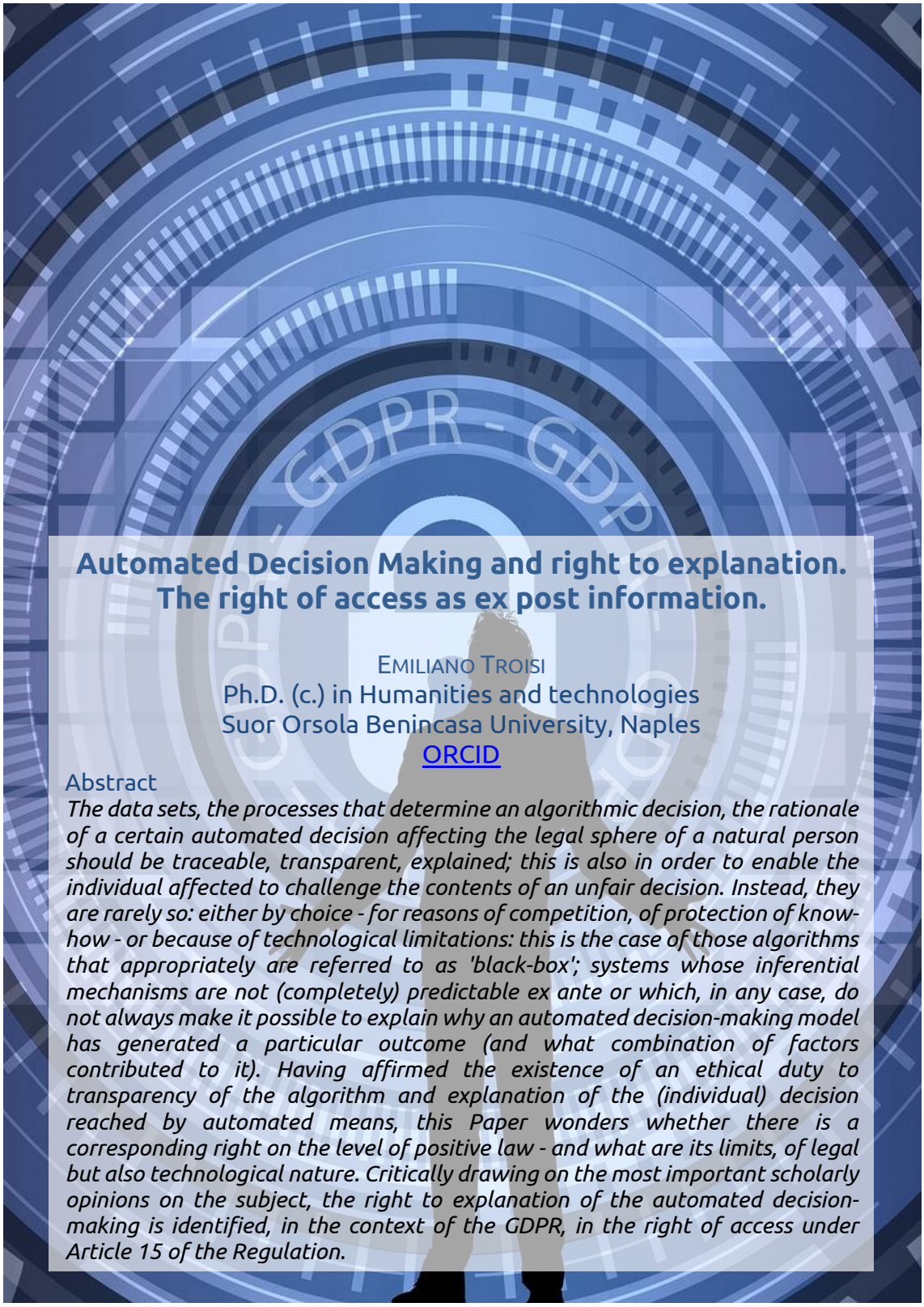
Troisi E, 'Automated Decision Making and right to explanation.The right of access as ex post information.' (2022) 1 EJPLT, Available at: www.ejplt.tatodpr.eu

EJPLT
ISSN 2704-8012



# Automated Decision Making and right to explanation. The right of access as ex post information.

EMILIANO TROISI
Ph.D. (c.) in Humanities and technologies
Suor Orsola Benincasa University, Naples
ORCID

Abstract

*The data sets, the processes that determine an algorithmic decision, the rationale of a certain automated decision affecting the legal sphere of a natural person should be traceable, transparent, explained; this is also in order to enable the individual affected to challenge the contents of an unfair decision. Instead, they are rarely so: either by choice - for reasons of competition, of protection of know-how - or because of technological limitations: this is the case of those algorithms that appropriately are referred to as 'black-box'; systems whose inferential mechanisms are not (completely) predictable ex ante or which, in any case, do not always make it possible to explain why an automated decision-making model has generated a particular outcome (and what combination of factors contributed to it). Having affirmed the existence of an ethical duty to transparency of the algorithm and explanation of the (individual) decision reached by automated means, this Paper wonders whether there is a corresponding right on the level of positive law - and what are its limits, of legal but also technological nature. Critically drawing on the most important scholarly opinions on the subject, the right to explanation of the automated decision-making is identified, in the context of the GDPR, in the right of access under Article 15 of the Regulation.*

Keywords: Automated Decision Making; algorithm; opacity; explanation; right of access; Gdpr.

Summary: 1. Automated Decision Making and the problem of algorithm opacity. – 2. GDPR: 'meaningful information' between 'generic algorithm functionality' and 'right to explanation'. – 3. The right of access as ex post information. – 4. Right to explanation: the functional implication between information and contestation of the automated decision. – 5. Conclusions.

## 1. Automated Decision Making and the problem of algorithm opacity.

The expression Automated Decision Making (acronym: A.D.M.), refers - broadly - to any process that enables, through the use of technological means, decision making without, or at least with irrelevant, human involvement. Such a definition, therefore, does not imply - being broader - but clearly includes, the use of Artificial Intelligence technologies, as more generally any computer technique that, relying on algorithms - namely, a sequence of operations executable by a processor[1] - enables repetitive tasks to be performed with data without the need for constant human guidance.[2]

Among these, AI systems - according to one of the most widely accepted definitions - are distinguished as, essentially, rational systems[3] capable of *'acting and thinking humanly*';[4] in other words, capable of solving problems by imitating what human behavior would be under similar circumstances. An intelligent system is capable of collecting data from a certain data base or environment, interpreting it,[5] and - in light of the goal to be achieved - deciding what the best action or decision is,[6] then acting accordingly in an almost automatic way.[7] This decision-making process is conducted by the machine - according to the AI technique implemented - by applying static reasoning schemes or by resorting to machine learning techniques (e.g. *machine learning, deep learning, neural networks, decision trees* and others). Without going into

[1] Treccani online, entry: algoritmo, https://www.treccani.it/vocabolario/algoritmo/

[2] Within the *State-of-the-Art Report on Algorithmic decision-making* by *Algo:aware*, a December 2018 study commissioned by DG Connect (Directorate General for Communications Networks, Content and Technology), Department of the European Commission, decision making algorithm is more broadly defined as "*A software system – including its testing, training and input data, as well as associated governance processes – that, autonomously or with human involvement, takes decisions or applies measures relating to social or physical systems on the basis of personal or non-personal data, with impacts either at the individual or collective level.*" Consequently, "*the definition of algorithmic decision-making is to be interpreted as a decision taken by a decision-making algorithm*".

[3] S Russell, P Norvig, *Artificial Intelligence: A Modern Approach*, (Prentice Hall, 2009). For further information, as to the more strictly legal sphere, it seems useful to refer also to M Durante, entry *Intelligenza artificiale. Applicazioni giuridiche*, *Digesto Italiano. (Terza appendice di aggiornamento della IV edizione. Discipline Privatistiche*, Vol. II, II Agg.), 714.

[4] M Somalvico, F Amigoni, V Schiaffonati, *Intelligenza Artificiale*, in S Petruccioli (ed.), *Storia della scienza*, (vol. IX, Istituto della Enciclopedia Italiana, Roma, 2003), 615-624; VV. AA., *Artificial Intelligence and life in 2030*, *One hundred year study on Artificial Intelligence* (Stanford University Press, 2016) 5.

[5] Reasoning about them or processing the information inferred from the data.

[6] Possibly changing the environment (natural or virtual) in which it operates or otherwise proposing some output, the solution to a specific problem.

[7] Cf. the definition developed by the Expert Group established by the European Commission: *A definition of AI: Main capabilities and scientific disciplines*, High-Level Expert Group on Artificial Intelligence (AI HLEG), (Brussels, December 18, 2018).

detail, it suffices to know - for present purposes - that in the latter case the machine, instead of executing pre-defined behavioral patterns, processes 'on its own' and dynamically - in application of self-learning and adaptive algorithms - the decision rule; in some cases resulting, therefore, also in being able to better respond and adapt to changes in the environment or refine, with use experience, the ability to generate an appropriate *output*.[8]

These techniques make AI an incredibly useful tool, capable of making decisions and predictions much earlier and more accurately than humans would, moreover in contexts that require the analysis of huge amounts of data - often unstructured[9] - that would otherwise be prohibitive for human agents.

AI, however, can make mistakes.[10] The use of Automated Decision Making systems can lead to decisions that are biased, harmful to human beings, discriminatory, and detrimental to fundamental human rights; jeopardizes the human right to free and equal access to goods and services and information; not to be unknowingly manipulated by unfair, aggressive and intrusive forms of marketing; it threatens - as this Paper will show - the legitimate claim to challenge harmful decisions made by the machine or, more importantly, to have them not made in a non-transparent manner and on the basis, perhaps, of incorrect and incomplete data.[11]

The initial myth of algorithmic neutrality has now been shattered; [12] intelligent machines are created, programmed and trained by human operators and therefore, like humans, are fallible; they can 'reason', or 'learn to reason' in a biased way.[13] The inferential rules given to the machine by its programmer - by which it then interprets knowledge and operates in its target environment - may reflect cultural biases, even well-meaning and unintended ones.[14] In data-driven systems, such as machine learning systems, moreover, biases can also arise from data collection, which may be insufficient, incomplete, or flawed; from training, due to biases induced by human interpretation of the data, in the case of supervised-learning algorithms, or - when un-supervised - as a result of

---

[8] *Amplius*, s. Y Bathaee, *The Artificial Intelligence Black Blox And The Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, (2018) Vol. 31, n.2, 890.

[9] This is the case for using self-learning algorithms

[10] So opens the European Commission's *White Paper on Artificial Intelligence* - COM(2020) 65 final, February 19, 2020: "Art*ificial Intelligence* […] *will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine. At the same time, Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes.*"

[11] Among others, see B Casey, *Title 2.0: Discrimination Law in a Data-Driven Society*, [2019] *J. L. & MOB.*, 36, https://doi.org/10.36635/jlm.2019.title; K Crawford, *The Hidden Biases in Big Data*, [2013] *HARV. BUS. REV.*, https://perma.cc/E95C-TUQU; see also, E Troisi, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, in *European Journal of Privacy Law & Technologies*, (2019)1, at: https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1027/276

[12] See, M Airoldi, D Gambetta, *Sul mito della neutralità algoritmica*, *The Lab's Quarterly*, (2018) XX, 4, 29

[13] J Burrell, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, *Big Data & Society*, (2016)1, 1-12, for whom the claim that algorithms classify information, and thus make decisions, in a more "objective" way cannot be taken literally given the degree of human judgment involved in the design of the algorithms themselves, particularly from the standpoint of defining clustering criteria, pre-classifying training data, and adjusting thresholds and decision-making parameters.

[14] On how classification systems can be and are concretely influenced, even with far-reaching consequences, by the 'point of view' of those who construct them, see the interesting work by GC Bowker, SL Star, *Sorting Things Out: Classification and Its Consequences*, (Cambridge, MA, The MIT Press, 1999).

online learning and self-adaptation through user interaction (think of common rank algorithms).

These biases can lead to unfair decisions.[15] And this situation is further aggravated by the opacity that often characterizes these systems, which is reflected in the possibility of contesting the decision made by automatic means.[16]

The data sets, the processes that determine the decision of the algorithms, the reason for a certain decision affecting the legal sphere of a natural person should be traceable, transparent, explained; this is also in order to enable the individual affected by it to contest its content. Instead, they are rarely so, at least adequately: either by choice[17] - for reasons of competition, protection of know-how - or because of technological limitations: this is the case of those algorithms that are properly referred to as 'black-box', systems whose inferential mechanisms are not (completely) predictable *ex ante*[18] or which, in any case, do not always make it possible to explain why a model has generated a particular result or decision (and what combination of factors contributed to it).[19]

The emergence of these problems and the growing concern, partly due to the fear that mistrust of new technological tools might limit their diffusion in the market, led to the adoption of a series of Ethical Charters - public or even private - at all levels: international, European, national.[20]

---

[15] See, on the point *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, (8 April 2019); *Statuto Etico e Giuridico dell'IA*, Fondazione Leonardo, (2019).

[16] See, among others, J Burrell, op. cit., for whom "*opacity seems to be at the very heart of new concerns about 'algorithms'* (operating on data) *among legal scholars and social scientists*".

[17] The fortunate appellation of 'Black Box Society' is due to Frank Pasquale, who masterfully outlines its features with evocative expressions such as '*the Secret Judgments of Software*' and '*the Secrecy of Business and the Business of Secrecy*'; cf. F Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, (Harvard University Press, 2015).

[18] But it is sometimes possible - it should be noted - to investigate its behavior by analyzing the responses the system produces in response to the stimuli it receives. So-called 'explanatory tools' [or *post-hoc* explanation techniques, c.f. J Zhong, E Negre, *AI: To interpret or to explain?*, (INFORSID, 2021)] are capable of *ex-post* reconstructing the functioning of certain 'opaque' decision-making models; in particular results of the examined model would be explained by finding the links between input data characteristics and results, or by constructing a simpler model to approximate the original model (Ibid., 6); the accuracy and reliability of these 'explanations' is challenged, for example, by C Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, (2019) Vol. 1, No. 5, 206– 215.

[19] J Burrel, op. cit, 3, with whom it is easy to agree, distinguishes 3 types of opacity in decision-making algorithms, which he describes as follows: "*Three distinct forms of opacity include: (1) opacity as intentional corporate or institutional self-protection and concealment and, along with it, the possibility for knowing deception; (2) opacity stemming from the current state of affairs where writing (and reading) code is a specialist skill and; (3) an opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human- scale reasoning and styles of semantic interpretation".* S. also, Y Bathaee, op. cit.; D Castelvecchi, *Can We Open the Black Box of AI*, in *Nature*, (2016) Vol. 538, 20.

[20] A successful but already incomplete attempt to map the various Ethical Charters, Declarations of Principles, or Guidelines, classified by geo-political context and analyzed by content, is credited to A Jobin, M Ienca, E Vayena, *Artificial Intelligence: the global landscape of ethics guidelines*, [2019] Nat. Mach. Intell. To name a few of the most relevant ones: *First Draft Of The Recommendation On The Ethics Of Artificial Intelligence*, SHS/BIO/AHEG-AI/2020/4 REV.2, Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, United Nations Educational, Scientific and Cultural Organization (UNESCO), (7 September 2020); *Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, Council of Europe, CM/Rec(2020)1; more dated and limited to the Justice field: *European ethical Charter on the use of Artificial Intelligence in judicial*

Based on the idea that - as digital technology becomes an increasingly central part of all aspects of human life - people should be able to trust such technology[21] and it should be developed to serve humans, be ethical, and respect fundamental rights, a set of principles and requirements have been identified to which intelligent systems, their applications, producers, programmers, and users (i.e., all stakeholders) should adhere. Leverage, or one of the main levers, of this strategy for 'human-centric' and 'trustworthy' AI[22] - to use the words of the Expert Group on Artificial Intelligence appointed by the European Commission[23] - is precisely transparency.[24] Often, a prerequisite for ensuring that basic human rights and ethical principles are respected, protected, and promoted.[25]

If Artificial Intelligence is to be developed at the service of human beings, humans must be able to make use of it consciously:[26] they have the right to always be aware of the fact that they are interacting with an A.I. system;[27] they must be able to understand its purpose, capabilities, and operating mode ('explicability' ) and the decisions, to the extent possible, must be explainable to those directly or indirectly affected by them[28] ('explainability') so that they can challenge its modalities and content ('contestability'[29]), being able to resort to human intervention to that end.

---

*systems and their environment*, European Commission For The Efficiency Of Justice (CEPEJ), (December 2018); In European Union context: *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence,(8 April 2019); in Italy is to be noted the paper edited by the Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale, *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, (2018). Famous among the privately authored declarations of principles is the one drafted within the *Future of Life Institute* and signed by some 1,800 researchers and nearly 4,000 other endorsers, some as famous as Stephen Hawking or digital market giants like Elon Musk and Jaan Tallinn: *The Asilomar AI Principles*, 2017, https://futureoflife.org/ai-principles/

[21] In this sense, *White Paper on Artificial Intelligence. A European approach to excellence and trust*, European Commission, COM(2020) 65, (February 19, 2020).

[22] See, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions, *Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168 final, 8 april 2019

[23] See the above-mentioned High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8 April 2019

[24] See, *First Draft of The Recommendation on the Ethics of Artificial Intelligence*, Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, United Nations Educational, Scientific and Cultural Organization (UNESCO), SHS/BIO/AHEG-AI/2020/4 REV.2, (7 September 2020), III (2) §39, according to which *"transparency may contribute to trust from humans for AI systems"*. See also the document compiled by the US National Institute of Standards and Technology (NIST): Vv. Aa., *Four Principles of Explainable Artificial Intelligence*, available online, in draft version, doi: https://doi.org/10.6028/NIST.IR.8312-draft

[25] Cf., Ad Hoc Expert Group (AHEG) UNESCO, Ibid., III (2) §37. On the relationship between transparency and trust, see, H Felzmann, E Fosch-Villaronga, C Lutz, A Tamó-Larrieux, *Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns*, Big Data & Society, (2019)1, 1-14; J Schoeffera, Y Machowskia, N Kuehla, *A Study on Fairness and Trust Perceptions in Automated Decision Making*, (2021), online in arXiv: arXiv:2103.04757v1

[26] See, Vv. Aa., *Paper sui Principi etici*, in *Statuto Etico e Giuridico dell'IA*, Fondazione Leonardo, (2019)

[27] See, *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, (8 April 2019), §78.

[28] Ibid., §53

[29] To explainability refers Section III (2) §40 of the *First Draft of The Recommendation on the Ethics of Artificial Intelligence*, Ad Hoc Expert Group (AHEG) – UNESCO, cit., according to which: *"Explainability refers to making intelligible and providing insight into the outcome of AI systems. The explainability of AI systems also refers to the understandability of the input, output and behaviour of each algorithmic building block and how it contributes to the outcome of the systems. Thus, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should be understandable and traceable, appropriate to the use context".*

Therefore, if there exists an undisputed ethical duty of transparency[30] of the algorithm and explanation of the (specific) decision reached by automated means (ADM), it need to be addressed the question of whether there is - and to what extent - a corresponding right on the level of positive law, and what are its limits, of legal but also technological nature.

General legal framework of reference (to be investigated)[31], at least in cases where automated decision-making involves personal data of natural persons is given by Regulation (EU) 2016/679, the General Data Protection Regulation.[32]

[30] According to the findings of the study conducted in 2019 by A Jobin et al. (s., A Jobin, M Ienca, E Vayena, *Artificial Intelligence: the global landscape of ethics guidelines*, op. cit., 7 ff.) '*transparency*' is the most prevalent principle in the current literature, present in 73 of the 84 documents analyzed, albeit with different shades of meaning, summarized by the locutions: *'transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing'*.

[31] Specific regulations may be found, however, for example, in the area of consumer protection legislation; for an overview of the situation and specific problems, see, among others, M Grochowski, A Jablonowska, F Lagioa, G Sartor, *Algorithmic transparency and explainability for EU Consumer Protection: unwrapping the regulatory premises*, *Critical Analysis of Law*, (2021) Vol. 8, No.1, 43-63.

[32] On April 21, 2021, the EU Commission published a proposal for a Regulation for Artificial Intelligence, the so-called *AI Act*, which aims to define the regulatory framework for the development and deployment of artificial intelligence systems. According to a progressive regulatory approach based on risk - which can be unacceptable (whereby the system is banned), high, or low or minimal - AI systems considered to be high risk will be allowed on the European market under certain mandatory requirements and subject to certification of compliance. Among these requirements is transparency. Indeed, Article 13 of the Proposal requires that high-risk systems be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately, and be provided with precise instructions for appropriate use. For other systems, deemed to be of lower risk, only very limited transparency requirements are imposed, e.g., in terms of providing information to signal the use of an AI system in interactions with human beings; moreover, with broad exceptions that have already been met with several criticisms (think, for example, of the non-applicability of the transparency requirement to biometric recognition systems used for crime prevention and detection purposes). This is not the appropriate place to extend the discussion to an analysis of the AI Act's regulatory framework; however, a few brief points cannot be ignored. First, given that Artificial Intelligence requires massive processing of data, very often personal data, it is abundantly clear that the future AI Regulation will have to function in a complementary way to the GDPR, harmonizing with its provisions and obligations, possibly specifying and/or integrating it with reference to the peculiarities of 'intelligent' systems; however - precisely on the very point of a clear connection between the two disciplines - the Proposal seems lacking, with the risk of engendering regulatory gaps (even serious) or uncertainties, even with regard to the roles and competences of the Notifying Authorities involved. In addition, however, to the concerns related to the lack of adequate harmonization with the General Data Protection Regulation (for censure see Joint Opinion No. 5 /2021 by EDPB and EDPS on the Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence; online here: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en), what is worrying about the proposal under comment is the inadequacy, in terms of the provision of protection mechanisms (including for the many exceptions) and redress-that is, mechanisms that would allow users who suffer detrimental automated decisions to obtain a direct remedy-(absent), to go beyond the discipline of Art. 22 GDPR, which does not cover the full scope of potentially harmful automated decision-making systems and is-as will be seen-in many ways precarious, thus the subject of long-standing criticism in doctrine. Even from the specific point of view of transparency, since it is not always possible - also due to technological limitations we will discuss below - to obtain a prior comprehensible explanation of the results produced by the machine, the EDPB and the EDPS do not fail to point out (see Joint Opinion No. 5/2021, cit, in particular: §72) how the Artificial Intelligence Regulation should promote new, more proactive, and timely ways of informing users of AI systems about the decision-making status the system is in at any given time, providing early warnings about potentially harmful outcomes so that people whose rights and freedoms might be compromised by the machine's autonomous decisions are able to react to or challenge those decisions.

## 2. GDPR: 'meaningful information' between 'generic algorithm functionality' and 'right to explanation'.

Article 22 GDPR expressly states[33] the right of the data subject 'not to be subject to a decision based solely on automated processing, including profiling[34], which produces legal effects concerning him or her or similarly significantly affects him or her.' The prohibition, under paragraph 2 of the same article, does not apply only if, and to the extent that, automated decision-making processing is necessary for the conclusion or performance of a contract between the data subject and a data controller or is based on the data subject's consent.[35]

Overlooking the scope of this exemption - also, not the only one[36] - which allows, in essence, Automated Decision Making under the "consensual" bases

[33] This is not, however, a legislative innovation, since the problem of automated decision-making was already addressed, with a less extensive prohibitive provision than the current one, by Article 15 and Recital 41 of Directive 95/46/EC. *Amplius*, among others, see LA Bygrave, *Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, in Comput. Law Secur. Rev., (2001)17, 17-24.

[34] The Regulation does not give a precise definition of automated processing; however, Article 22(1) expressly includes profiling in its scope. Under Art. 4, profiling, in the context of the GDPR, means 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." Profiling is thus a species of automated processing that consists in the exercise, by a software, of an inferential operation capable of deriving, from a certain amount of data entered in a data-set - analyzed and identified among them, automatically, statistical correlations - other data related to characteristics or behavioral patterns attributable to a given individual or groups of them in order to classify him/her in precise groups or categories and/or predict probable future behavior.

[35] It should be emphasized that automated processing and profiling involving sensitive data as referred to in Article 9 of the GDPR are prohibited in any case, unless the data subject has given his or her explicit consent or this is authorised under special legal basis by Union or Member States' laws and provided that, the processing is necessary for reasons of public interest and proportionate to the goal pursued. The data processed then, pursuant to Art. 5 of the Regulation, for (automated) processing to be lawful, must be collected for specified, explicit and legitimate purposes, and subsequently processed in a way that is not incompatible with those purposes. However, as has been aptly noted, the machine learning underlying ML systems involves existing datasets and other real-time data streams in complex dynamic processes with sometimes not entirely predictable outcomes, which are difficult to reconcile with the need for full prior identification of the purposes of the processing to be consented to ex ante, to the point that some go so far as to cast doubt on the very possibility of giving truly informed consent to processing entrusted to black-box systems (see C Kuner, DJB Svantesson et al, *Machine learning with personal data: is data protection law smart enough to meet the challenge?*, *International Data Privacy Law*, (2017) vol. 7, no. 1, 1.

[36] the GDPR also admits, in fact, that it may be permissible - pursuant to paragraph 2(b) of the article under consideration - to take decisions on the basis of automated processing in other cases contemplated by Union or Member State law to which the data controller is subject; this, exceptionally, even without the consent of the data subject and provided that appropriate measures are taken to protect the natural persons involved. According to Recital 71 of the Regulation, these hypotheses may include, for example, the use of computerized processing for the purposes of monitoring or preventing fraud or tax evasion, but in any case it must be considered necessary - in the author's opinion - that the use of such mechanisms proves to be justified in the light of the pursuit of purposes of significant public interest and that the related compression of the rights and freedoms of the individuals involved be proportionate. In these cases, the Regulation only requires the Legislator to provide for appropriate measures to protect individuals, but the right to request human intervention and to possibly challenge the decision is not ensured by paragraph 3 of Article 22 (as it is in the other cases examined), although nothing, of course, precludes that the rules authorizing such processing, in providing the required safeguards, may provide for even broader forms of protection. *Amplius*, s. G Malgieri, *Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations*, in *Computer Law & Security Review*, (2019) 35(5). http://dx.doi.org/10.1016/j.clsr.2019.05.002

for processing[37] - considered to be suitable, at least in theory, to ensure a broader awareness of the data subject[38] - it is provided that in cases where automated decision-making is permitted, the data controller has a duty - pursuant to Art. 22(3)GDPR - to implement appropriate measures to protect the rights, freedoms and legitimate interests of the data subject, including, in particular, measures that guarantee the data subject's right to obtain human intervention by the data controller, to express their opinion and to contest the decision.

Further requirements for the case of automated decision-making as defined in Article 22 are laid down in Articles 13 (2)(f),14 (2)(g) and 15 (1)(h) of the Regulation. These provisions, with identical wording, prescribe - respectively in the context of the information (so for Articles 13 and 14), and as an extension of the right of access (Article 15) - that the data subject has the right to be informed about "the existence of automated decision-making, including profiling, as referred to in Article 22(1) and (4), and, at least in such cases, [to receive, *ed.*] meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".

The data controller (of ADM processing) is therefore under a specific duty to provide information, and thus transparency (involving: existence of an automated decision-making process; meaningful information on the logic of the algorithm; significance and envisaged consequences of such processing for the data subject); and if it is true - as has been argued - that in the structure of the GDPR, the objective of the security of the movement of personal data turns out to be entrusted to a strategy focused, first and foremost, on the transparency of data processing[39], it is evident that the duty of transparency in question must be interpreted, in any case, more broadly than the 'minimum content' of information already due for any other processing. Any other 'reductive' interpretation, would, in fact, make the specification of the European legislator completely superfluous; not to mention that a more protective regulation - thus a wider notion of transparency - would be well justified because of the greater (potential) harmfulness of automated

[37] See, E Troisi, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, in *European Journal of Privacy Law & Technologies*, (2019)1, online: https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1027/276

[38] With the consequence that the processing should be considered unlawful in all cases in which the data subject has not expressly consented to it through his or her own conscious manifestation of will, either directly, or as part of a more complex contractual relationship, on the assumption, however, in the latter case, that the automated processing is to be considered necessary for the conclusion or execution of the agreement. For further analysis of the effectiveness of the informed consent mechanism in the digital environment to ensure awareness, and thus protection of the data subject, see IA Caggiano, *Il consenso al trattamento dei dati personali nel nuovo Regolamento europeo. Analisi giuridica e studi comportamentali,* [2018] *Osservatorio del diritto civile e commerciale,* 67-106*;* and the recent L Gatt, IA Caggiano, R Montanari (eds.), *Privacy and Consent. A Legal and UX&HMI Approach for Data Protection,* (Suor Orsola University Press, 2021), which address the topic with an empirical approach and statistical dignity, accompanying the legal reading with behavioral analysis.

[39] C.f. R Messinetti, *La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, *Contratto e Impresa*, (2019) 3, 861.

individual decision-making compared to other data processing [40]; this is precisely the reason why the legislator decided to regulate ADM with a prohibition rule (albeit with many exceptions).

Following this guideline, the information referred to in Articles 13(2)(f),14(2)(g) and 15(1)(h) of the Regulation (i.e., the 'logic' of the algorithm and the 'envisaged consequences') must certainly consist of more than the indication in concise, transparent, intelligible, and easily accessible form[41] of the categories of data processed and the purpose of the processing, as well as its automated mode.[42] The question is whether such a duty of transparency, however, extends to the point of being able to consider legally enshrined in positive law a right of the natural person subject to the processing to the algorithm's 'disclosure'-that is, at least to something more than the general functional logic implied by the system, a right to 'open the black box'[43] and look inside it-and, more than that, a right to the 'explanation' of the specific decision affecting him or her and affecting his or her legal sphere: the reasons, the individual circumstances why a specific decision, with a certain content, was taken by the algorithm with respect to him or her.[44]

As has been acutely observed,[45] Articles 13(2)(f) and 14(2)(g), which impose a duty of transparency to be performed, precisely, in the context of the privacy information - at a time, therefore, logically preceding the decision-making processing - can only refer to the general functionality of the algorithm and to the pattern of decisions normally expected from its operation through an abstract and *ex ante* evaluation ('*envisaged* consequences' indeed says the Regulation). The content of the data subject's informational right - referring to these legal bases - is therefore necessarily limited and certainly unsuitable to substantiate a duty to 'explain' the automatic decision. However, the reference to "meaningful information," "significance," and "envisaged consequences" for

[40] When automated processing is also likely to produce legal effects or otherwise significantly similarly affect the natural person whose data are being processed, it is no longer just the data-subject's privacy that is at stake but his or her right not to be discriminated against, to have free and equal access to goods and services not to be manipulated without his or her knowledge by unfair, aggressive, and intrusive forms of marketing, but also the legitimate claim to challenge detrimental decisions by a machine or, more importantly, that these are not taken in a non-transparent manner and on the basis, perhaps, of incorrect and incomplete data.

[41] C.f. Recital 39 GDPR.

[42] According to B Goodman, S Flaxman, *European Union Regulations on algorithmic decision-making and a "right to explanation"*, *AI Magazine*, (2017) vol. 38, n. 3, 50-57or available online at: https://arxiv.org/abs/1606.08813v3, "*It is reasonable to suppose that any adequate explanation would, at a minimum, provide an account of how input features relate to predictions, allowing one to answer questions such as: Is the model more or less likely to recommend a loan if the applicant is a minority? Which features play the largest role in prediction?*"

[43] The espression is from R Messinetti, q.v., op. cit.

[44] It should be noted that Recital 71 of the GDPR expressly mentions the data subject's right to "explanation," in these terms: *"[…] 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.*" However, as Watcher et al. point out, after a careful analysis of all the preparatory work, this wording, originally included in the proposed Regulation, is later expunged from the text finally submitted for legislative approval; c.f., S Wachter, B Mittelstadt, L Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, *International Data Privacy Law*, (2017) vol. 7, no. 2, (Oxford University Press, 2017), 76–99. It also reconstructs the genesis of the provision in question, C Djeffal, *The normative potential of the European rule on Automated Decisions: a new reading for Art. 22 GDPR*, ZaöRV (2020) 80, 847-879.

[45] C.f., S Wachter, B Mittelstadt, L Floridi, op.cit.

the data subject requires the data controller, if not certain a disclosure of the algorithm functioning[46], at least to illustrate, in an easily accessible form, the main criteria for its operation[47], while also providing "information about the intended or future processing, and how the automated decision-making might affect the data subject." [48] The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision[49] and express valid consent under Article 4 GDPR.[50] For example, consider a loan file that is fully processed online whereby a bank relies on profiling algorithms to identify the applicant's creditworthiness score (credit scoring) and then decide whether to grant or reject the loan and what interest rate to propose.[51] Where the decision is based solely (c.f. Art. 22 GDPR) or otherwise substantially on such a score, the data controller will have to explain to the data subject that he or she will be evaluated by an algorithm and how this works - not the mathematical formula, but rather how the system actually 'reasons' - thus, what information is taken into account (e.g., employment and income situation; degree of indebtedness; credit history; etc.); what are the sources (e.g., what are the data provided by the data subject in the application form or obtained from third parties; data already collected from previous reports; information from public records; etc.); what circumstances mainly affect the decision and how (e.g., by identifying a possible case history; by resorting to simulations); how users are ranked (how many and what ranks are considered, based on score); how and how much the score affects the credit

[46] In the sense of the non-necessity of disclosure of the algorithm, in order to comply with the minimum level of transparency required by the Regulation, see Article 29 Working Party's *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, No. 251, (Oct. 3, 2017), for which "*The GDPR requires the controller to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm*".

[47] Complexity is not an excuse for failing to provide information to the data subject. Recital 58 states that the principle of transparency is "*of particular relevance in situations where the proliferation of actors and the technological complexity of practice makes it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him are being collected, such as in the case of online advertising*". In the light of an interpretation that looks at the context of the regulatory provisions - in particular the requirements of clarity, intelligibility, accessibility, and plain language that preside over the information the data subject is generally entitled to under Art. 12 (1) GDPR - it must be excluded that the European legislator wants to refer to a duty of mere 'disclosure' of the algorithm, especially if intended in its mathematical guise as i.e. system architecture, having rather to consider that what is pertinent here, and it is up to the data controller to explicate - perhaps in addition to the mathematical function that substantiates the algorithm in question - is its concrete performance: the context in which it is applied, the purposes pursued, the techniques applied and perhaps even examples about its concrete functioning. Only by including such circumstances in the content of the disclosure should the "explanatory" duty that the regulation requires be considered fulfilled.

[48] In this sense Article 29 Working Party, op. cit., p. 26, also adding that "*in order to make this information meaningful and understandable, real, tangible examples of the type of possible effects should be given*."

[49] Thus, verbatim, Article 29 Working Party, op. cit., p. 25.

[50] On this point, see, for example, the recent ruling of the Italian Supreme Court (Cass., sez. 1 Civ, 25/05/2021 no. 14381, Ord.) on the validity of the consent given by the user who did not have prior knowledge of the performing scheme of an algorithm (in this case a rating algorithm, used by the defendant to process users' reputational profiles); speaking precisely of "lack of transparency of the algorithm," the Court in fact affirmed that for the processing of personal data, consent is validly given only if it is freely and specifically expressed with reference to a clearly identified processing, with the consequence that "in the case of a web platform (with associated computer archive) designed to process the reputational profiles of individuals, based on a computational system with at its base an algorithm aimed at establishing reputational scores, the requirement of awareness of the data subject cannot be considered met where the algorithm's concrete functioning scheme and the elements of which it is composed remain unknown or not knowable by the data subjects.

[51] The case is evoked in Recital 71 of the Regulation.

application outcome. To make this information (cf. *significance* and *envisaged consequences* of such processing) *meaningful* and understandable, real and concrete examples should be given of the kind of possible decisions and effects for the data subject.[52] The data controller should also briefly explain to the data subject what measures are taken to avoid, for example, algorithm malfunctions and biases, and inform him or her that he or she has the right to request the review of the decision by invoking human intervention ( cf. Art. 22(3) GDPR). In addition to this,[53] the controller should also make it clear that

[52] Italics refer to Articles 13(2)(f) and 14(2)(g) GDPR.

[53] When it comes to algorithmic procedural activity of the public administration, that is, when automated decision-making systems are used in public decision-making, the minimum content of the duty of transparency imposed on the data controller (in this case, a public entity in the exercise of an authoritative power) is much more defined, thanks to the already intense interpretative contribution of case law. The Council of State (Consiglio di Stato, the Italian administrative judge), in particular, in ruling no. 8472/2019 [see comments by A Mascolo, *Gli algoritmi amministrativi: la sfida della comprensibilità*, *Giornale di diritto amministrativ*o, (2020) 3, 366 ff; M Timo, *Algoritmo. Il Procedimento di assunzione del personale scolastico al vaglio del Consiglio Di Stato,* Giur. It., (2020) 5, 1190] - in part recalling but overruling precedent No. 2270/2019 - affirming a general principle, states (para 15.1), in interpreting what it calls the principle of knowability, i.e., the principle - introduced by the GDPR - whereby everyone has the right to know about the existence of automated decision-making processes concerning him or her and, in this case, to receive meaningful information about the logic involved, that "*such a right* [...] *must be accompanied by mechanisms capable of deciphering its logic. In this view, the principle of knowability is complemented by the principle of comprehensibility* [...]." More specifically, it clarifies how (para 13.1) "*the mechanism through which the automated decision (i.e., the algorithm) is implemented must be 'knowable,' according to a strengthened declination of the principle of transparency, which also implies that of the full knowability of a rule expressed in a language different from the legal language. Such knowability of the algorithm must be ensured in all aspects: from its authors to the procedure used for its elaboration, to the decision mechanism, including the priorities assigned in the evaluation and decision-making procedure and the data selected as relevan*t"; so, the "*technical formula*" must be accompanied "*by explanations that translate it into the 'legal rule' underlying it and make it readable and understandable*." This is "*in order to be able to verify that the criteria, prerequisites and outcomes of the robotized procedure comply with the requirements and purposes established by the law or by the administration itself upstream of that procedure, and so that the methods and rules under which it was set are clear - and consequently open to review.*" [*author's translation*]. The absence of such a minimum level of comprehensibility (could be said, explanation) of the algorithm constitutes, in the opinion of the Council of State, such a vice as to invalidate the entire procedure (and consequent measures) for being its entire method censurable for lack of transparency. It is clear, however, and the writer of the cited ruling makes this explicit, that such an interpretation of the rules of the Data Protection Regulation, which are general in nature and therefore applicable both to decisions taken by private parties and, as in this case, by public bodies, is affected, however, in the latter case, by the concurrent effectiveness of the principles already governing public action: in particular, the principle of knowability set forth in the GDPR, interpreted in the aforementioned terms, constitutes in this specific case, more than anything else, "*direct specific application of Art. 42 of the European Charter of Fundamental Rights ("Right to a good administration"), where it states that when the Public Administration intends to take a decision that may have adverse effects on a person, it has an obligation to hear the individual before acting, to allow him or her access to its archives and documents, and, finally,* [...] *to 'give the reasons for its decision*.'"

More extensively, on transparency and the necessary 'comprehensibility of the decision-making algorithm used in the exercise of administrative activity, see G Fasano, *L'intelligenza artificiale nella cura dell'interesse generale*, *Giornale di diritto amministrativo*, (2020) 6, 715.

In France, it is the legislator who by LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte) amended the Code des relations entre le public et l'administration providing for the introduction of Art. L. 311-3-1 for which "*Sous réserve de l'application du 2° de l'article L. 311-5, une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande*". Rule further specified with the Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique (https://www.legifrance.gouv.fr/eli/decret/2017/3/14/PRMJ1632786D/jo/texte) providing for the introduction of Art. R. 311-3-1-1 ("*La mention explicite prévue à l'article L. 311-3-1 indique la finalité poursuivie par le traitement algorithmique. Elle rappelle le droit, garanti par cet article, d'obtenir la*

the user has the right to know (have *access* to) the score assigned to him by the algorithm and the rationale - i.e., the logical premises - for this ranking; this point will be addressed in the next section, after a necessary introduction.

## 3. The right of access as *ex post* information.

As mentioned above, it can be easily agreed that the transparency obligations in Articles 13(2)(f) and 14(2)(g) GDPR do not seem suitable to ground an actual right of the data subject to an *ex post* explanation of the automated decision; nevertheless, an equally restrictive interpretation of Art. 15(1)(h) of the Regulation, variously justified, in the scholarly literature, on the wording of the regulatory provision[54] - identical to that of the above articles[55] - or on the assumption that the right of access has the same scope and purpose as the right to information.[56]

Within the framework of the analyzed rules - which the GDPR provides for the protection of the natural person in the case of Automated Decision Making - the oversight power attributed to the data subject is embodied not only in the right to information, but also in the right to obtain human intervention and especially to challenge the decision generated by the algorithm.[57] In this system of safeguards for the data subject, which bases the lawfulness of automated decision-making processing on his or her consent (in a broad sense), the information required by Articles 13 and 14 GDPR - as it has been here outlined - fulfills the important function of enabling the data subject to exercise, with awareness, his or her right to manage (i.e., give, withdraw,

---

*communication des règles définissant ce traitement et des principales caractéristiques de sa mise en œuvre, ainsi que les modalités d'exercice de ce droit à communication et de saisine, le cas échéant, de la commission d'accès aux documents administratifs, définies par le présent livre*") and Art. R. 311-3-1-2 ("*L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes : 1°) Le degré et le mode de contribution du traitement algorithmique à la prise de décision ; 2°) Les données traitées et leurs sources; 3°) Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé; 4°) Les opérations effectuées par le traitement*). Moreover, the principles are the object of Décision n° 2018-765 DC du 12 juin 2018 del Conseil Constitutionnel which appears to give them a narrow reading. Available at the link: https://www.conseil-constitutionnel.fr/decision/2018/2018765DC.htm.
On algorithmic decision-making in the French legal system, s., G Malgieri, *Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations*, Computer Law & Security Review, (2019) 35(5), 13. http://dx.doi.org/10.1016/j.clsr.2019.05.002.

[54] See, S Wachter, B Mittelstadt, L Floridi, op. cit.; G Finocchiaro., *Intelligenza artificiale e diritto. Intelligenza artificiale e protezione dei dati personali*, Giur. It., (2019) 7, 1657, according to which there would not be granted a full right to explanation under Art. 15 GDPR, the wording of the provision precluding it, assuming a "lack of harmony" between it and Recital 71 of the Regulation.

[55] For S Wachter, B Mittelstadt, L Floridi, op. cit., this provision, mentioning the "envisaged consequences" of the decision-making processing would unequivocally refer, in this case as in Articles 13 and 14, to a necessarily *ex ante*, and therefore limited, 'explanation': in essence, no explanation at all.

[56] Along these lines, Article 29 Working Party, Op. cit., p. 27, concludes as follows: "*Article 15(1)(h) says that the controller should provide the data subject with information about the envisaged consequences of the processing, rather than an explanation of a particular decision".*

[57] See, R Messinetti, op. cit.; C Malgieri, G Comandè, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, *International Data Privacy Law*, (2017) Vol. 7, No. 4; ME Kaminski, *The Right to Explanation, Explained*, *Berkeley Technology Law Journal*, (2019) 34, 189–218.

restrict, etc.) consents[58], while the right of access would add to this same function the additional one (in this specific context) of enabling him or her to effectively challenge - because duly informed - the harmful decision.

In spite of the almost identical wording of Articles 13 (2)(f) and14 (2)(g), on one side, and Art. 15 (1)(h), on the other, it is not possible to make a symmetrical interpretation: while in the context of the information notice the information due can only refer to the generic functionality of the algorithm and to the sorts of decisions normally expected from its operation, in the case of the right of access it is more appropriate to interpret the rule extensively, deriving from it a duty of information, for the data controller, that is more significant and detailed and that pertains to the specific decision - possibly already taken by the system towards the data subject - and to the inferential stages that led (the machine) to that given output[59]: an *ex post* explanation.

Going back to the example of the online loan, it is reasonable to think - given these premises - that the right of access accorded by the GDPR to the data subject - the applicant, in the example - extends to the score, i.e., the profile, assigned to him by the rating algorithm that would define his creditworthiness; this is in fact - on closer inspection – an applicant's personal data, while also, at the same time, the product of a profiling process, thus decisive for the purposes of the automatic decision of the loan application in which he is interested. Knowing this data, in addition to being a right of the data subject (Art. 15 GDPR), is also essential so that he or she can, for example, assess its accuracy, perhaps request its rectification (or in general exercise the further rights over his or her own data set forth in Art. 16 to 20 of the Regulation);[60] properly challenge the automatic decision (e.g., the non-granting of the loan or its undesirable conditions). Admittedly, the profile - the rank assigned to the

[58] Of a role of the Information Notice as a tool meant to enable the data subject to exercise control over data processing, speaks IA Caggiano, *Il consenso al trattamento dei dati personali nel nuovo Regolamento europeo. Analisi giuridica e studi comportamentali*, op. cit, who refers to S Mazzamuto, *Il principio del consenso e il potere della revoca* in R Panetta (ed.) *Libera circolazione e protezione dei dati personali*, (Milano, Giuffrè, 2006), 1004.

[59] Of this opinion, among others: C Malgieri, G Comandè, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit.; TW Kim, BR Routledge, *Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach,* in *Business Ethics Quarterly*, (2022) Vol. 32, Issue 1, 75 – 102. https://doi.org/10.1017/beq.2021.3.

[60] If this were not the case, the rights of rectification, erasure and restriction of processing (recognized under Articles 16, 17 and 18 GDPR, respectively) would, paradoxically, inevitably be undermined precisely in those cases where they would be most needed: profiling and, more generally, data analytics techniques, involving elements of 'prediction' and 'extrapolation', greatly increase the inherent risk of data inaccuracy (output) and expand the range of data processed, making it more difficult for the data subject to control them. In the interpretation aimed at here, the information referred to in Article 15 GDPR, would be considered to be functional to the exercise of these rights (as well as that of challenging the decision based solely on automated data processing, in the case of ADM). Cf. on the point, with reference to profiling, Article 29 Working Party's *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, No. 251 of October 3, 2017, WP 251 rev01, English version, p.18, according to which, verbatim, '*The rights to rectification and erasure apply to both the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person)*'. On the (broad) content of the right of access in the case of inference, s. C Malgieri, G Comandè., Op. ult. cit., p. 246, who also enhances in this sense the scope of Article 20 GDPR, which recognizes the data subject's right to data portability in a readable format. Article 20, on this point, however, is recognized as having a far lesser scope by Article 29 Working Party, which excludes data produced by algorithmic activity of the data controller from the scope of the obligation: see in particular, *Guidelines on the right to data portability*, Article 29 Working Party, No. 242, (December 13, 2016), WP 242 rev.01, p.10 ff. [recently referred to and confirmed by *Guidelines on the targeting of social media users*, EDPB, No. 8, (September 2, 2020)].

score determined by the algorithm - expresses behaviors or characteristics that are probable, expected, but which as legally relevant (to the granting or terms of the loan) must nonetheless be reasonably inferred from the data of the data subject, who must therefore be enabled to assess such reasonableness and "express his or her point of view" (cf. Art. 22(3)). In order for this to happen, the information the data controller is required to provide (as per Art. 15) cannot have the same content as the *ex-ante* Information; it cannot simply explain in an understandable way how the algorithm supposedly works, rather it should - sticking to the example above - indicate the score assigned; identify the concrete reasons for it (i.e., at least the specific circumstances that weighed in the judgment); explain (or at least report in an 'intelligible' way) why that score resulted in the rejection of the application and perhaps what different score or characteristic would have led to a different outcome.[61]

## 4. Right to explanation: the functional implication between information and contestation of the automated decision.

The interpretation of the right of access as a right to an *ex post* information (or rather, at this point, it should be said, explanation) of the ADM-decision is, in this writer's opinion, the natural consequence of a systematic and consistent reading of paragraph 1(h) of Article 15 of the General Data Protection Regulation. As seen, in fact, Article 22 GDPR, in paragraph 3, expressly provides for the right of the data subject affected by an ADM process to question the decision based solely on automated processing through the possibility of obtaining human intervention, namely, to interact in a 'dialectical' way to the data controller, expressing his or her opinion, requesting a verification of the decision and also being able, subsequently, to contest its assumptions. This is a right, on closer inspection, which, unless its scope being frustrated, must necessarily imply, for the data subject (and therefore owed by the data controller), a specific information on the functioning of the decision-making algorithm; an information that is not abstract and prognostic, but concrete and *ex-post*,[62] referring to the specific application that concerns him/her; only in this way, in fact, would the data subject be put in a position to fully exercise his/her oppositional right, having the appropriate instruments to formulate a specific and reasoned contestation.[63] It is arguably not coincidental that Recital 71 of

[61] The last reference is to the so-called counterfactual explanations technique; see, S Watcher, B Mittelstadtd, C Russell, *Counterfactual explanations without opening the Black box: automated decisions and the GDPR*, *Harvard Journal of Law & Technology*, (2018) Vol. 31, No. 2.

[62] In the sense intended by S Wachter, B Mittelstadt, L Floridi, op.cit., p. 6, according to which "*an ex post explanation occurs after an automated decision has taken place*" and "[…] *can address both system functionality and the rationale of a specific decision*". See also, TW Kim, BR Routledge, op. cit.

[63] Among others, s. AD Selbst, J Powles, *Meaningful information and the right to explanation*, International Data Privacy Law, (2017) Vol. 7, Issue 4, 233-242. https://doi.org/10. 1093/idpl/ipx022. Kaminski speaks of a *'qualified'* transparency, to be defined precisely because of its functional nexus with the exercise of the other rights accorded to the data subject, not least the right to contest the automatic decision: ME Kaminski, op. cit., according to which "*there is a clear relationship between the other individual rights the GDPR establishes—contestation, correction, and erasure—and the kind of individualized transparency it requires. This suggests something interesting about transparency: the substance of other underlying legal rights often determines transparency's substance. If one has a right of correction, one needs to see errors. If one has a right against*

the Regulation[64] - which, unlike the regulatory text, expressly mentions the right of the data subject to "obtain an explanation" - explicitly relates this safeguard to the right to "obtain human intervention" and to be able to "express his or her opinion" about the automatic decision. The functional implication link between information and contestation is, it would seem, also reinforced by the Explanatory Report accompanying the so-called Convention 108+ of the Council of Europe - i.e., the Protocol of Amendment, signed by Italy, which, in updating the 1981 Strasbourg Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data to the new technological context, expressly adds in the new Article 9 ( titled: Rights of the Data Subject), subpar. a, the right of every individual "*not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration*" - where, in § 75a, it emphasizes how the right of access is actually functional to the possibility of effective review of the automatic decision: "*It is essential that an individual who may be subject to a purely automated decision has the right to challenge such a decision by putting forward, in a meaningful manner, his or her point of view and arguments*. […]".[65]

The right of access, furthermore, as can also be seen from Recital 63 GDPR, gives the data subject the right to obtain all the information referred to in Art. 15 of the Regulation, not only at a time prior to or coinciding with the beginning of the processing, but also during the course of the processing, and necessarily extends even to those data (and processes) in the meantime 'produced' by the processing, i.e., to the data inferred from the processed data (input) as a result (even if only intermediate) of the automated process (of extrapolation,

*discrimination, one needs to see what factors are used in a decision. Otherwise, information asymmetries render underlying rights effectively void.*" (Ibid. p. 213). It should be noted how also in the ruling no. 8472/2019 of the Council of State (Consiglio di Stato, the Italian administrative judge) – see *supra* footnote 53 - in justifying a stronger concept of transparency, functional to the judicial review of the algorithmic decision in the administrative field, it is pointed out that (para 13. 3) Article 15 GDPR, unlike Articles 13 and 14, has the merit of providing for a right enforceable by the data subject and not an obligation to the data controller, and also allows to overcome the time limits set by Articles 13 and 14, allowing the subject to acquire information even if the processing [...] has even already generated a decision. Also recognizing the existence of a right to ex post explanation of the individual automatic decision are KIM T.W. and ROUTLEDGE B. R., op. cit., who, however, motivate by valuing the informed consent in cases of algorithmic decision-making processing, which they conside based on a duty of information that is not instantaneous (and ex ante) but continuous and pivotal, a "right to an updating explanation."

[64] It should be noted that although not binding, the Recitals are a guide to the interpretation of norms. See, in the case law, CJEU, Judgment of the Court (Third Chamber) of 13 July 1989, *Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung*, c. 215/88. ECLI:EU:C:1989:331. See also, European Commission, *Guide to the Approximation of EU Environmental Legislation ANNEX* (Environment, 2015), according to which: "*Recitals explain the background to the legislation and the aims and objectives of the legislation. They are, therefore, important to an understanding of the legislation which follows.*"

[65] May it be useful to also refer here to the High Level Expert Group on Artificial Intelligence (AI HLEG), which in the paper published on April 8, 2019, and significantly titled *Ethics Guidelines for Trustworthy AI*, after identifying a number of fundamental principles, including fairness, which the new technological framework should be inspired by to be "trustworthy," significantly states significantly states that the procedural dimension of fairness implies the ability to challenge decisions developed by AI systems and the possibility of effective contestation against them. To this end, decision-making processes must be explainable. Furthermore, specifically on explicability, it is added (p.13): "*This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.*"

clustering, etc.),[66] which, if referable to the data subject in the terms of Art. 4 (1), fall within the objective scope of the GDPR.[67] These inferred data, in fact, in addition to being the 'basis' of a possible (successive) automated decision able to affect the legal sphere of the data subject, are themselves already the 'output' of an automated processing (e.g. data analytics, profiling etc.)[68] with respect to which the right of access stands out, therefore, as a legitimate request for *ex post* information. At least in this circumstance - moreover, not necessarily coinciding with the scope of Art. 22 GDPR, but referable to a much broader number of cases contemplating profiling or in any case a data processing (including non-automated) that involves predictive or, broadly speaking, inferential elements[69] - the right of access must be considered, in fact, certainly referring to a much broader information content than that covered by Articles 13 and 14 of the Regulation, embracing not only detailed information on the personal data used as input to the processing[70] (i.e., in the case of profiling, the data used to create the profile) but also all possible information on the output data (e. g. the profile or score assigned to the data subject; all data 'derived' from the analysis of his or her personal data and the circumstances inferred). [71] When, furthermore, the processing (i.e., the

[66] Recital 63's reference to data such as "diagnoses, examination results, assessments by treating physicians..." are in fact *ante-litteram* examples of inferred data; technological innovation and artificial intelligence applied to data analytics models make far greater and more invasive 'inferences' possible. See, B Mittelstadt, S Wachter, *A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI*, Columbia Business Law Review, (2019)1.

[67] Cf. Article 29 Working Party's *Opinion 4/2007 on the concept of personal data*, available online here: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm; on the personal data nature of inferences, s. also, B Mittelstadt, S Wachter, Op. ult. cit., p. 25 ff.

[68] Data mining and profiling activities, even if not preordained to a decision-making moment, are in themselves already potentially detrimental to individual interests deserving of protection, as the European Court of Human Rights has noted on several occasions (see, ECHR, Guide to the Case-Law of the of the European Court of Human Rights. Data protection., 30 April 2021, online at: https://www.echr.coe.int/Documents/Guide_Data_protection_ENG.pdf), in addition to raising specific problems regarding the protection of 'produced' data, particularly when they have the nature of sensitive data. See, B Mittelstadt, S Wachter, Op. ult. cit. The problem has been particularly prominent regarding health data; *amplius*, W Schäfke-Zell, *Revisiting the definition of health data in the age of digitalized health care*, [2021] International Data Privacy Law; G Malgieri, G Comandè, S*ensitive-by-distance: quasi-health data in the algorithmic era*, Information & Communication Technology Law, (2017) Vol. 26, No. 3, 229-249.

[69] *Amplius*, B Mittelstadt, S Wachter, Op. ult. cit.

[70] their origin, the purpose for which they are processed, and any other information required under Article 15 GDPR.

[71] If this were not the case, the rights of rectification, erasure and restriction of processing (recognized under Articles 16, 17 and 18 GDPR, respectively) would, paradoxically, inevitably be undermined precisely in those cases where they would be most needed: profiling and, more generally, data analytics techniques, involving elements of 'prediction' and 'extrapolation', greatly increase the inherent risk of data inaccuracy (output) and expand the range of data processed, making it more difficult for the data subject to control them. In the interpretation aimed at here, the information referred to in Article 15 GDPR, would be considered to be functional to the exercise of these rights (as well as that of challenging the decision based solely on automated data processing, in the case of ADM). Cf. on the point, with reference to profiling, Article 29 Working Party's *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, No. 251 of October 3, 2017, WP 251 rev01, English version, p.18, according to which, verbatim, '*The rights to rectification and erasure apply to both the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person)*'. On the (broad) content of the right of access in the case of inference, s. G Malgieri, G Comandè., Op. ult. cit., p. 246, who also enhances in this sense the scope of Article 20 GDPR, which recognizes the data subject's right to data portability in a readable format. Article 20, on this point, however, is recognized as having a far lesser scope by Article 29 Working Party, which excludes data produced by algorithmic activity of the data controller from the scope of the obligation: see in particular, *Guidelines on the right to data portability*, No. 242, Article

inference) is automated and is functional to the adoption of a decision having legal effects for the data subject (based solely on the quality of that inference) - and at least in this case[72] - all (further) 'meaningful information about the logic used' by the algorithm that such output has (already) generated must also be considered to be included in the information due; and this is not, evidently, prognostic and abstract information about the functioning of the algorithm in general, but more realistically the reasons for the decision (the output produced) in concrete terms; the logic used in that specific case, the how and why:[73] an explanation. The only limit, as to the contents of this explanation, is given by the protection of the trade secret and intellectual property of the data controller and pertains to the (*full disclosure* of the) formula of the decision-making algorithm, which Recital 63 GDPR significantly contemplates precisely with reference to the right of access; moreover, explicitly underlining its nature of not an absolute limitation, but rather subject to judicious balancing, case-by-case, with the opposing interest (to the explanation and contestability of the decision) of those subject to the automated process (and never suitable, in any case, to justify the refusal to provide information to the data subject).[74]

It is the essential need for effective protection of the human person against digital power[75] that demands for the meaning of Art. 15(1)(h) GDPR to be reconstructed in such a way as to ensure that the data subject has communication of all data concerning him or her: including those 'inferred'; can assess its accuracy, completeness, relevance (to the extent that the Regulation and the purposes of processing allow); and, where that data forms the basis of an (automated) decision that affects (or is likely to affect) his or her legal interests, has access to at least all the information necessary to assess, and possibly sustain an effective contestation of it (by recourse to a human interlocutor or a courtroom).[76]

If the digital world has become a 'social', living space, within which a person's

29 Working Party, (December 13, 2016), WP 242 rev.01, p.10 ff. (recently referred to and confirmed by EDPB's *Guidelines on the targeting of social media users*, No. 8, September 2, 2020).

[72] The reference is to the wording of Art. 15(1)(h) GDPR, intended as alluding to something more than the minimum content of transparency and safeguard that is generally required under Articles 13 and 14, and that is in this case, since in an *ex-post* scenario, possible.

[73] Thus the expression 'significance [...] of such processing' in Art. 15(1)(h) GDPR is intended to be interpreted in this perspective and context (i.e., of information intended to reach the data subject after the decision).

[74] See, G Noto La Diega, *Against the Dehumanisation of Decision-Making. Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, in *Journal of Intellectual Property,* Information Technology and Electronic Commerce Law (JIPITEC), (2018)9, Vol. 3, No.1, Available at SSRN: https://ssrn.com/abstract=3188080 (s. in particular p. 23, para 69).

[75] In addition to the general principle of fair processing, to which certainly can be said to be contrary the refusal, when unjustified, to provide information about the derived data referable to the data subject or about the rationale for a certain automatic decision.

[76] For R Messinetti, op. cit, p. 867, the legal-technological system must enable the person to understand the machine's understanding of that person in the context of decision-making processes directed to affect his or her legal and vital sphere; it would therefore be the need to retain control over one's personal identity (and its formation) that would justify an interpretation of the right of access as a right to understand the logic and justifying reasons for the automated decision (having as its object precisely that identity) and one's personal data "derived" from the original inputs, which constitute the intermediate and/or conclusive outputs of the processing.

'onlife'[77] existence takes place and is delineated, i.e., his or her identity is determined, in a way that is (also) legally relevant (and otherwise uncontrolled), then the protection of human dignity - anchor of all data protection legislation - individual freedom, and the fundamental right of each person to affect his or her individual and collective dimension, goes first and foremost through awareness, i.e. transparency meant as the justification of digital power – namely, its explanation -and, only then, through an active power of control over one's own data - particularly automatically inferred data - and the possibility of contain (and contesting) the impact of technologies on and from them, which without a (prior) right to explanation would be voided. Denying *this* transparency, which ultimately means not only the right to know the rationale for a certain automatic decision affecting a natural person but also the relevant (inferred) data behind it (i.e., who that person *is* online) is to somehow dehumanize the individual.[78]

## Conclusions.

Having postulated the possible legal basis of a right to an explanation of the algorithmic decision, it is necessary, in conclusion, to dedicate at least a mention to its limitations, although this is neither the object nor the goal of this Paper.[79]

On a legal level, a first limitation would derive from the rather limited scope of Art. 22 GDPR,[80] at least basing on the literal wording of the article. Automated processing to which Article 15(1)(h) of the GDPR is applicable are in fact only those defined by Article 22 of the regulation, i.e., expressly, only processing that are "*decision-making*",[81] in other words, lead to decisions affecting the data subject, and are "*based solely on automated processing*".[82]

[77] The evocative term is from Luciano Floridi, who chose it to represent man's experience in hyper-historical societies in which he "no longer distinguishes between online or offline" and even becomes increasingly "unreasonable to ask whether one is online or offline." Cf. L Floridi, *The Fourth Revolution. How the Infosphere is reshaping human reality,* (Oxford University Press, 2014); ID. (ed.), *The Onlife Manifesto: Being Human in a Hyperconnected Era,* (Springer, 2015).

[78] Such an understanding of transparency can be found, for example, in R Waelen, *Why AI Ethics Is a Critical Theory*, Philosophy & Technology, (2022) 35:9. https://doi.org/10.1007/s13347-022-00507-5

[79] Please refer for further to the larger work of this same author, E Troisi, (Forthcoming) *Decisione algoritmica, Black-box e AI etica: il diritto di accesso come diritto a ottenere una spiegazione*, Jus Civ., (2022)4

[80] In this sense, see, among others, S Wachter, B Mittelstadt, L Floridi, op. cit.

[81] This implies that all those data processing that is not typically 'inferential,' i.e., where the use of computer techniques is limited, perhaps, to the storage and/or organization of personal data without involving any analytical and/or evaluative passage of them through the application to the data-set of deductive rules algorithmically defined in the model, are excluded from the scope of the Article.

[82] The literal interpretation of the provision would therefore require that only fully automated decisions fall within the scope of application of the discipline under examination, while excluding all those in which it is possible to detect even the slightest human involvement that can variously 'interfere' with the automated decision-making process, being able to verify or modify the decision but also, for example, merely ratify it. Such an interpretation would risk excluding from the scope of application of the strict protection regulation under consideration a good portion of decisions that are in essence the outcome of automated processing and only formally applied by human, merely passive, intervention. Therefore, a broad interpretation of Article 22 GDPR should be more appropriately preferred, considering included in the scope all decisions that are 'automated in substance,' even those, then, that involve some human intervention in the decision-making process, but this intervention is irrelevant to the content of the decision: in this sense, see, for

These decisions (of which the individual affected would be entitled to an explanation ), according to Article 22 (1) of the Regulation, in addition to this, must also "*produce legal effects* [concerning the data subject] *or* [at least] *similarly significantly affect him or her*." The effect of such decisions should therefore affect, by denying or limiting it, the free exercise of a right of the individual recognized by the legal system.

However, of this last clause,[83] as of the entire §1 of Article 22 GDPR, it is also possible to make a broad interpretation, as many scholarly opinions have already convincingly argued; this first limitation on the right to explanation therefore seems to have most likely negligible impact.[84]

A more important limitation on the right to explanation of automated decision-making outcome, on the other hand, lies in the need to balance the right of the individual subject to ADM to be adequately informed about how the algorithm works and the right - opposed - of those who developed and use the algorithm for economic purposes to keep its design secret in order to avoid dispersing its commercial value to the benefit of competitors. According to Recital 4 of the GDPR, "*the right to the protection of personal data is not an absolute right; it must be […] balanced against other fundamental rights, in accordance with the principle of proportionality*" and the Regulation "*respects all fundamental rights and observes the freedoms and principles recognised in the Charter as enshrined in the Treaties* […]" including "[…] *freedom to conduct a*

example, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, No. 251, Article 29 Working Party, (October 3, 2017), 21, where, verbatim: "*The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing. To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data*". And before that, the British Data Protection Authority, the Information Commissioner's Office (ICO), also observed a similar orientation, for which: *the interpretation of the word solely" in the context of Article 22(1) [...]is intended to cover those automated decision-making processes in which humans exercise no real influence on the outcome of the decision, for example where the result of the profiling or process is not assessed by a person before being formalized as a decision*" C.f.r., *Feedback request – Profiling and automated decision-making*, Information Commissioner s Office, (2017), 19, also referenced in G Malgieri, G Comandè, *Right to legibility of automated decision-making*, op. cit.

[83] In fact, the catalog of rights potentially affected by automated decision processing widens considerably where all those fundamental personal rights of constitutional rank are also taken into account; these are expressed in a broad form, such as freedom of expression, freedom and secrecy of correspondence, freedom of movement, freedom of worship, the right of defense and, most of all, the principle of non-discrimination: naturally 'elastic' rights and freedoms, which lend themselves to covering a whole open range of individual legal situations that may be exposed to concrete risk, for those who hold them, in the event of conflicting, and unjust, decisions by an Artificial Intelligence. This broad protection of individual rights, is then further expanded by the very following of the provision; §1 of Art. 22 in fact extends its scope also to those ADM cases which, while not affecting legal situations overtly qualified as rights, "affect [however, ed.] in a similar and significant way the person," i.e. - in the interpretation that seems preferable - affect, and seriously impact (this is the meaning to be attributed to the expression 'significantly') legitimate interests or otherwise legitimate claims of the human person as long as they are not merely factual, i.e., legal situations that the legal system, although it does not recognize as rights immediately enforceable by the individual, considers equally worthy of protection either by granting them some measure of safeguards or at least by demonstrating an overall judgment of reprehensibility toward practices and behaviors that threaten their free and full individual fruition. See, on this point, among others, G Malgieri, G Comandè., *Right to legibility of automated decision-making*, op. cit.; see also E Troisi, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, op. cit.

[84] See E Troisi, (Forthcoming) *Decisione algoritmica, Black-box e AI etica: il diritto di accesso come diritto a ottenere una spiegazione*, Jus Civ., (2022)4, and footnotes no. 81, 82, 83.

*business* […]". More specifically, Recital 63, on the right of access, textually states that "*that right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software*", with the clarification that "*however, the result of those considerations should not be a refusal to provide all information to the data subject*". In this, echoing, among other things, the recent Council of Europe Recommendatio [85] according to which "*States should establish appropriate levels of transparency with regard to* […] *use, design and basic processing criteria and methods of algorithmic systems* […]" and "*the legislative frameworks for intellectual property or trade secrets should not preclude such transparency, nor should States or private parties seek to exploit them for this purpose*"[86], which, moreover, undoubtedly gives priority, among the protected situations, to the right of individuals to be properly informed.

The issue is far from irrelevant, considering that most of the automatic decision-making algorithms, even very invasive and used - including those of Google and Facebook, to be clear - are proprietary and kept confidential as to their mechanism. In general, it can be said - but this is especially with regard to the area of Competition, for which more case law is given - that the case law of the Court of Justice[87] allows limitations or exceptions to copyright and trade secret[88] in "exceptional circumstances",[89] that is, when conflicting and equally important general interests or fundamental rights are at stake (within the limits of what is proportionate to the purpose).[90]

In any case, an assessment among all the interests at stake should be required on a case-by-case basis, bearing in mind that when decision-making processes affect individuals, the degree of necessary 'explainability' of the algorithm (and its decision), based on the context, should increase as the severity of the consequences (on human rights and the interests at stake) increases if such output is incorrect or otherwise inaccurate. And that anyway, in the author's opinion, when serious human rights impacts are to be expected, transparency may also require the sharing of specific algorithms or data sets.[91]

The last limitation to be addressed is the technological one: the question is whether the data subject's right to explanation of the algorithmic decision that affects him or her - namely, to receive, ex post, 'meaningful information' about the logic concretely used in that specific automated decision-making process

[85] The reference is to Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.
[86] Article 4.1
[87] See: CJEU, Judgment of the Court of April 6, 1995, Joined Cases C-241/91 P and C-242/91 P, ECLI:EU:C:1995:98 (so-called Magill); CJEU, Judgment of the Court (Fifth Chamber) of April 29, 2004, C-418/01, ECLI:EU:C:2004:257 (so-called IMS Health); CJEU, Judgment of the Court of First Instance (Grand Chamber) of September 17, 2007, T-201/04, ECLI:EU:T:2007:289 (Microsoft v. Commission).
[88] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.
[89] See the just-mentioned Court of Justice of the European Union rulings Magill, IMS Health and Microsoft v. Commission.
[90] For a more detailed analysis, please refer to E Troisi, (Forthcoming) *Decisione algoritmica, Black-box e AI etica: il diritto di accesso come diritto a ottenere una spiegazione*, Jus Civ., (2022)4
[91] For more on this issue, see among others, G Malgieri, *Trade Secrets v Personal Data: a possible solution for balancing rights*, International Data Privacy Law, (2016), 6(2):ipv030. Available online at: http://dx.doi.org/10.1093/idpl/ipv030; G Noto La Diega, Op. cit.; ME Kaminski, Op. cit.

that concerns him or her - does not encounter, once overcome the legal limits, a technological limitation capable of hindering its actual enforcement.

According to what has been widely argued in the literature,[92] the inferential rule that animates the automated processes of Machine Learning systems would not (always) be knowable,[93] at least with respect to the reconstruction of the logic that is actually behind the processing of a specific output.[94]

While this is not always the case[95] - and while awaiting promising new elaborations of technology[96] - this would nevertheless - as things stand - represent a limitation, of technological nature, capable of making, in these scenarios, the right to explanation 'useless'; and this is because it would be impossible for the data controller who relies on the use of self-learning "black box" systems to retrace (and thus share) the decision-making logic followed by the machine. Certainly, *ex ante*; in many cases *ex post*.[97]

Given this, it is necessary to wonder how the assumption of the practical unfeasibility of the right to explanation reacts in the system of the GDPR. As

---

[92] See among others, Y Bathaee, Op. cit.; D Castelvecchi, Op. cit.

[93] Among others, s,. PJ Lisboa, *Interpretability in Machine Learning. Principles and Practice*, (Springer, 2013). On the difference between explainability and interpretability of decision-making models and the relationship between them, see J Zhong, E Negre, op. cit.; see also DA Broniatowski, *Psychological Foundations of Explainability and Interpretability in Artificial Intelligence*, NISTIR 8367, (2021). Online: https://doi.org/10.6028/NIST.IR.8367.

[94] It seems appropriate to note, however, how there is no shortage of different approaches in the literature to solving the mediated problem of the decision-making 'fairness' of automated systems, which, having abandoned the perspective of the necessary (reconstructability of an) 'explanation' of the machine's 'opaque' decision, provide, for the protection of individuals, for the implementation of systems suitable for detecting and preventing treatment the discriminatory within the automated decision-making process. See in this sense, C Dwork, M Hardt, et al., *Fairness Through Awareness*, ITCS 2012, (Cambridge, MA, USA, 2012). Also online at https://arxiv.org/abs/1104.3913v2.

[95] Explanatory tools make it possible in some cases, and only ex post, to analyze the reasoning steps of the system by reconstructing its essential decision-making moments so as to return, upon request, the decision logic concretely used by the algorithm in the analyzed case; see supra, note 18. Moreover, research has long been active in this area, with the goal of making ML algorithms increasingly amenable to ex ante and ex post control; see among others, A Datta, S Sen, Y Zick, *Algorithmic transparency via quantitative input influence*, in *37th IEEE Symposium on Security and Privacy*, (2016), online at: https://ieeexplore.ieee.org/document/7546525; S Galhotra, R Pradhan, B Salimi, *Explaining black-box algorithms using probabilistic contrastive counterfactuals*, 2021, online on arXiv:2103.11972v1; S Wachter, B Mittelstadt, C Russell, *Counterfactual explanations without opening the Black Box: automated decisions and the GDPR*, *Harvard Journal of Law & Technology*, (2018) Vol. 31, No. 2, 843-887; M Loi, A Ferrario, E Viganò, *Transparency as design publicity: explaining and justifying inscrutable algorithms*, [October 2020] Ethics and Information Technology, doi: https://doi.org/10.1007/s10676-020-09564-w; on the concept of Algorithmic accountability, and the possible role, as an external audit to the automated decision-making system, of reverse-engineering algorithms, see also N Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, [2014] Tow Center for Digital Journalism, Columbia University, online: https://doi.org/10.7916/D8ZK5TW2.

[96] The reference is to that strand of research that goes by the name of XAI, Explanatory Artificial Intelligence, aimed at defining methods and tools of "deep explanation," that is, capable of shedding light on the workings of algorithms with 'opaque' functioning, i.e., the construction of intelligent systems that are more transparent and interpretable; see HJ Watson, *The Need for Explainable Processes and Algorithms*, in *Business Intelligence Journal*, Vol. 25, No. 2, p. 5; AB Arrieta, N Dìaz-Rodrìguez, et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, (2020), online in ArXiv, abs/1910.10045.

[97] Skeptical that technological solutions truly capable of ensuring high standards of transparency of decision-making processes animated by complex ML algorithms may be found, even in the future, and therefore critical of the possibility and desirability of affecting technological design through regulatory tools, is, for example, Y Bathaee, op. cit., p. 929; on the possibility of different approaches to that of ex ante or ex post (audit) transparency of the algorithm, see, among others, C Dwork, M Hardt, et al., op. cit.; on the advantages of a scalar case-by-case approach, see again Y Bathaee, op. cit., p. 889.

correctly noted,[98] from this perspective, in the hypotheses in which the type of explanation requested is assumed to be precluded, making it impossible to verify the lawfulness and fairness of the processing, the consequence produced on the logical level would not be the inapplicability of the rule - thus its uselessness - but the reinstatement of the prohibition - absolute, lacking the minimum conditions for its derogation[99] - of subjecting the data subject to automated decision-making processing under Article 22 (1) GDPR. Unless the data controller, therefore, can guarantee, including on a technological level, that the decision, having legal effects, based solely on automated processing of the data subject's personal data, is also explainable (i.e., its concrete decision-making logic can be traced), the data controller is obliged to refrain from implementing automated decision-making processing.

[98] R Messinetti, op. cit., p. 889.
[99] That is, the suitable measures referred to in the Article 22(3) GDPR