Athens Journal of Law 2024, 11: 1-17
https://doi.org/10.30958/ajl.X-Y-Z

Artificial Intelligence in Decision-making: A Test of Consistency between the "EU AI Act" and the "General Data Protection Regulation"

*By Claudio Sarra**

The recent Regulation that sets down harmonised rules on Artificial Intelligence in the European Union, known as the "AI Act," includes a significant requirement for human oversight in high-risk AI systems during their use (art. 14). This requirement embodies the "human-in-command" approach, ensuring both legal and ethical compliance. The AI Act is intended to complement the General Data Protection Regulation (hereinafter GDPR), thereby forming a consistent and comprehensive legal framework. This paper focuses on AI systems producing decisions and examines the consistency of the AI Act's mandatory human oversight measures (art. 14) with GDPR's provisions on decisions based solely on automated processing (art. 22). At first glance, the provisions seem mutually exclusive. Mandatory human oversight under the AI Act could render art. 22 of GDPR inapplicable, as it applies only to decisions made by automated processing, implying no human involvement in decision-making. However, art. 22 of GDPR provides crucial safeguards for individuals, such as the right to human intervention, the ability to express opinions, and the right to contest decisions. This raises questions about whether the AI Act will exhaust these safeguards, and if it is capable of providing equivalent protection for decisions made by AI systems. This paper aims to analytically address these questions and arguments for a revision of the ordinary interpretation of art. 22 of GDPR, § 1.

Keywords: *AI Act; Algorithmic decisions; GDPR; Human oversight.*

Introduction

Following a procedure that lasted more than three years, on June 13th, 2024, the new European Regulation (EU) 2024/1689 setting down harmonised rules on artificial intelligence, usually referred to as the "AI Act", was finally approved and formally signed. This was a long-awaited and extremely complex piece of regulation, not only because of the subject matter but also because it stood at the centre of an already existing (and still not completed) regulatory universe about data governance and AI that informs the European Union digital strategy as outlined in 2020 in a formal communication by the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions[1].

*Ph.D., Associate Professor, Department of Private Law and Critique of Law, University of Padova, Padova, Italy.
Email: claudio.sarra@unipd.it
[1]See European Commission (2020).

Although already in force, the complete application of the AI Act was delayed till August 2nd, 2026, letting the stakeholders prepare themselves for a full compliance. In the meantime, some parts of the Act would receive a progressive application, creating all the specific governance institutions and technical tools provided for in the Regulation.

The high complexity of the AI Act was aggravated by the modifications it underwent during the procedure of approval: From April 2021, the time of the first proposal by the Commission, much transpired in AI technology making the Regulation at risk of being severely obsolete from the start[2].

Since the AI Act explicitly acknowledges many other regulatory European acts, and given the fact that AI could in principle be implemented in many specific fields with their own regulations, the legal community will be engaged in quite a lot of difficult work to make unitary sense of such normative universe.

In fact, along with the numerous interpretative issues of such a complex regulation *per se,* the most serious application problems will be determined by the intersection of these many regulations in the new socio-economic dynamics that technology is expected to determine.

Of course, one of the most important Regulations the AI Act is destined to interact with is the General Data Protection Regulation (Reg. UE 2018/679, hereinafter also as "GDPR"), and, as a matter of fact, the AI Act explicitly states that it must be applied "without any prejudice" to its application[3].

Therefore, the AI Act is supposed to complement the GDPR when the AI system is used to process "personal data" (which obviously is not always the case) in building a general legal framework. Consequently, they are meant to be consistent with each other in order to prevent regulatory shortcomings.

One of the major points in the AI Act is the mandatory prescription to design and develop high-risk (HR) AI systems in a way to ensure human oversight during the period of use[4]. This provision aims to realise the continually recommended "human-in-command" approach, and it is a crucial requirement in order to guarantee AI systems not only legal but also ethical compliance[5].

An AI system is defined as a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to

[2]The major innovation has been of course the advent of generative AI which at the moment seems to be the most powerful driving force in AI economic implementation, but that was completely uncovered by the original AI Act Proposal. The final formulation includes a Chapter V entitled "General purpose AI models" which are those that display significant generality and are capable of competently performing a wide range of distinct tasks regardless of the way the models are placed on the market and that can be integrated into a variety of downstream systems or applications (AI Act, art. 3 n. 63). Chapter V provides specific obligations for providers of such models (with a significant distinction between models presenting "systemic risk" and others), that need to be coordinated with those already introduced for providers of HR AI systems in case the model is implemented in that kind of systems.

[3]AI Act, art. 2 § 7.

[4]art.14.

[5]See the 2019 document entitled *Ethics Guidelines for a Trustworthy AI*, approved by the High-Level Expert Group on Artificial Intelligence, set up by the European Commission, at 14 – ss.

generate outputs such as predictions, content, recommendations, or ***decisions*** that can influence physical or virtual environments"[6].

This paper concerns AI systems producing decisions and how the AI Act art. 14 requiring mandatory human oversight measures impacts onto the consolidated interpretation of art. 22 of GDPR about decisions based solely on automatic data processing. I will assume a referent scenario in which both regulations are theoretically applicable, that is in the case of an AI HR system which processes personal data and takes decisions that produce legal effect on a subject or affect him/her significantly (more on this later). The rest of this paper is structured as follows: first it will briefly recall the legal European discipline about automatic decision-making as stated in the GDPR (§ 2); then it will discuss the Human Oversight Principle (HOP) as provided by the AI Act showing that the interaction with the GDPR, as usually interpreted on point, is likely to determine a diminished protection of the data subject, (§§ 3-4); and finally, on the basis of the previous discussion, it will argue for a different and more precise interpretation of art. 22 of GDPR in order for people interacting with AI-based decision-making tools to take advantage of a full protection, thus better realizing the main goal of building an anthropocentric technological development (§§ 5-6)[7].

Automatic Decision making in art. 22 of GDPR

In the last few years, the General Data Protection Regulation has been one of the most studied and discussed pieces of regulation in the European legal field: not only has it impacted almost in every economic and social sector but, as technology has improved more and more, the interpretative challenges have increased accordingly.

As far as fully automatic decision-making (ADM) is concerned, the GDPR offers the main discipline in art. 22, which has been discussed thoroughly in the academic literature. For the sake of the discussion proposed here, it is useful to recall at least the main points of that framework[8].

Art. 22, § 1 of the GDPR, states that The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. Although framed as a right of the data subject, scholars agree that – in accordance with the *privacy by design* principle - the provision should be read as stating a prohibition directed primarily to the data controller. [9]

[6]AI Act, art. 3, italics bold added.

[7]The goal of developing a safe, trustworthy, human centric technological innovation is always explicitly stated in the EU official documents, see, for instance, HLEG (2019), European Commission (2020), AI Act, *Recital 1.*

[8]For a full discussion, see Sarra (2020b); Larus, Hankin, Carson, Christen, Craga, Grau, Kirchnet, Knowles, McGetttick, Tamburri, & Werthner (2018); Brkan (2019); Mendoza & Bygrave (2017); Veale & Edwards (2018); Bygrave (2021); Bygrave (2019); De Hert & Papakonstantinou (2016).

[9]This principle states that "the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and

This restrictive approach does not exclude significant exceptions in which cases ADM can be legitimate, but some extra safeguard measures should be adopted to protect the data subject's rights.[10]

It is worth noting that fully automatic decision-making is not prohibited *per se* but only since it implies a modification of the legal situation concerning the recipient or "similarly significantly" affects him or her.

Let us call this level of application "the threshold": automatic decision-making is prohibited only when it goes above the threshold.

So, as far as the GDPR approach to ADM is concerned, the framework is pretty clear, at least theoretically: in principle, they are prohibited as long as they are a) fully automated; b) produce legal effects or; b1) affect the data subject (the threshold) similarly and significantly.

Interpretative issues are related to each and every one of the points highlighted. Let us briefly discuss them:

a) Fully automated

It is generally acknowledged that for a decision to be not solely based on automatic data treatment, the human intervention should be relevant. In other words, the human being who intervenes should have the competence and the authority to change the decision when this is the case. It is worth noting that this point is not explicitly stated in Article 22, but is the result of the interpretation given to the text since the first discussions about the new GDPR[11].

Sometimes Article 22 has been criticised mainly because it endorses quite a limited vision of the role of humans in complex decision-making systems. In fact, in order to exclude the complete automation of the procedure, this interpretation asks for a human presence in the final stage of the processing, before the decision is actually implemented, with the authority to confirm or change the output reached by the machine. As a consequence, a substantial human intervention in previous stages appears to be irrelevant, which is debatable[12].

Moreover, it was noticed that in those cases in which intelligent systems *outperformed* human abilities[13], the human being involved would more likely be inclined to confirm the decision taken by the machine, either because the judgement may be affected by the so called *automation bias*[14], or simply because it could be

organisational measures, such as pseudonymisation, which are designed to implement data-protection principles" (GDPR, Art 25), see Preece (2018); Wagner (2020).

[10]Art. 22, § 2- 3, GDPR.

[11]ART29WP (2018), at 21.

[12]It is true, however, that in complex systems it may not be clear when an automation actually can be considered started: if you go back enough in a process, chances are that a human acting somewhat emerges here and there. So, making relevant *any* form of human intervention in *any* phase of a complex procedure ending in a decision above "the threshold" could delude the purpose of the prohibition in art. 22 GDPR.

[13]Veale & Edwards (2018) at 400.

[14]The *automation bias* is the systematic tendency to over-rely upon the output of a system, ignoring or underestimating one's experienced assessments. See Tsamados, Aggarwal & Cowls (2021) at 4; Goddard, Roudsari & Wyatt (2012); Mosier & Skitka (1999); Skitka, Mosier & Burdick (2000).

extremely difficult for her/him to justify the specific reasons why the decision taken by so an accurate tool needs, in a certain, specific case, to be changed[15].

b) and b1) *The Threshold*

For the prohibition set forth in art. 22 GDPR to apply, the decision taken by the machine must have a significant impact on the data subject in a way that either produces a formally representable modification of his legal condition (b) or significantly affects the recipient condition, even though there is no legally binding obligation on the subject who decides (b1). Examples of this last situation are online recruiting and automatic refusal of credit application as indicated in *Recital 71*. Of course, these examples are far from offering a clear interpretation of this part of Article 22, § 1 for a secure application without controversy. In fact, if the principle is the prohibition of automatic decisions which "significantly affects" the recipient, there can be cases similar to the ones indicated as examples that could not necessarily be above this threshold. For example, a credit application for buying unnecessary costly gadgets just for entertainment. This paves the way for a case-by-case consideration.

As anticipated, the general prohibition has three exceptions indicated in §2: one being the case of a Regulation by the EU or a member state (lett. b); the other two are situations which are in the hands of the parties, the data subject and the data controller (lett. a, c), namely, the decision is necessary for entering, or performance of, a contract between the data subject and a data controller (a); the decision is authorised with the data subject's consent (c).

As we can see, the exceptions are not all on the same level, and that is the reason why I presented them, separating the cases in which the automation is authorised by a legal act (by the EU or Member State) from the others, emphasizing the fact that the latter ones are under the parties' control. In other words, and this is noteworthy, the GDPR allows the parties to decide by themselves to use ADM tools – even above the threshold - in their relationship.

This is an important acknowledgment of their autonomy, although the complexity of our contemporary society demands a high level of awareness and accountability in taking those kinds of decisions.

c) *The Extra Safeguard Measures*

For all the exceptional cases, some "safeguard measures" to protect the rights and the legitimate interests of the data subject must be implemented.

Notably, in cases a) and c) - those left to the parties' autonomy - this is a specific obligation of the data controller. Whereas, in the case of a statutory authorization, this requirement must be fulfilled by the regulatory act authorizing the automatic decision[16].

---

[15]Sarra (2020a).

[16]It is worth noting that there is a significant difference between art. 22 and *Recital 71.* The former requires the UE or member State to provide for adequate safeguard measures without prescribing anything more on point, thus letting the legislator decide the quality and quantity of those protective

Furthermore, the measures provided for should aim to protect the "rights, freedoms and legitimate interests" of the data subject[17].

As we can see, this is quite a demanding requirement, whose object is also very wide spanning from "rights" to "freedoms", to "legitimate interests". The *ratio legis* seems to lay in the aim to charge the data controller (or the EU/member State Legislator) with the responsibility to take full care of the data subject who is supposed to be the part more exposed to the potential damaging consequences of ADM. On this point we will see a different approach in the AI Act, which will be a reason for concern.

The minimum safeguard measures the data controller should implement are indicated in art. 22, §3, GDPR, and they take the form of some extra rights: to request a human intervention; to express one's opinion; and to contest the decision.

The list of rights is quite longer in *Recital 71*, which offers a first hermeneutical aid to read art. 22. In fact, it includes also the right to "specific information to the interested party", and to obtain "an explanation of the decision". Both are supposed to give the data subject a deeper and more specific understanding of what happened than the *ex-ante* information s/he had received in compliance with the information duties provided for in art. 13 and 14.

The fact that the actual formulation of art. 22, GDPR gives a shorter list of minimal safeguard measures raised a doctrinal dispute about the existence in the GDPR of a *right to* (full) *explanation* of algorithmic decisions[18], as well as on the way in which those extra rights are supposed to work together.[19] This, however is not the issue at hand.

Instead, it is more interesting to reflect upon the comprehensive view endorsed by the GDPR about human intervention.

Therefore, in the framework of art. 22 GDPR, the human participation to a decision taken by a machine is relevant in at least two ways.

If a qualified human intervenes before the decision is taken or implemented, that is *not* a fully automated decision, and it falls outside art. 22. To reiterate: this is *not* stated in the GDPR, it is the way in which art. 22, §1 has been interpreted so far. The main concern that led to this was the fear of an easy circumvention of the prohibition, by simply putting an employee to merely read the output of the machine without being entitled to any modification. But this interpretation has a shortcoming: it relies on the mere *presence* of a qualified human to exclude the full automation and not on a specific activity or participation in the decision-making process. In other words, the decision can *actually* be totally taken by the machine and still not

measure. On the other hand, *Recital 71* gathers all the exceptions together stating that "in any case" some minimal safeguards should be provided for, namely the right to "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". Given the acknowledged hierarchical primacy of EU law on member States, following one or the other formulation can have internal constitutional consequences.

[17]Roig (2018).

[18]Wachter, Mittelstadt & Floridi (2017); ART29WP (2018) at 25; Goodman & Flaxman (2017); Brkan (2019); Sarra (2020a); Malgieri & Comandé (2017); Edwards & Veale (2017).

[19]See for a full discussion at Sarra (2020b)

be qualified as fully automated *just* because a qualified human was somewhat there even though s/he acted as a rubber-stamper[20].

On the other hand, a qualified human may intervene after the decision is taken and implemented if invoked by the data subject. In this case, the human intervention is supposed to be a *safeguard measure*. Thus, even though there are no mandatory prescriptions about what the intervening human is supposed to do, s/he should take care of the rights, freedoms and legitimate interests of the data subject.

As such, this is just one of the minimal rights to be acknowledged to the data subject and it shares its defensive potential with others, in particular with the right to contest the decision. The importance of such a right has been highlighted in the recent literature so that some authors have argued for the existence of a *contestability by design* principle in the GDPR[21].

Now, let us turn to the Human Oversight Principle in the AI Act.

The Human Oversight Principle (HOP)

The HOP is regulated in art. 14 of the AI Act, and is included in Chapter III, Section 2, which provides for the requirements for high-risk artificial intelligence systems. This means that it is strictly mandatory only for those kinds of AI systems[22]. However, the *Ethics Guidelines for Trustworthy AI* released by the High-Level Experts Group on Artificial Intelligence set up by the European Commission in 2019, considers the principle of human oversight as one of the seven main general requirements directly derived from the four principles for a trustworthy AI[23]. As such, from an ethical point of view, it should be implemented in all AI systems.

That document defines three main ways to practically shape the HOP: the Human In The Loop model (HITL), Human On The Loop (HOTL) and Human In Command (HIC)[24].

Apparently, the AI Act approach on the subject seems to endorse a fusion of the HOTL and HIC models: HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system and the ability to decide when and how to use the system in any particular situation[25].

According to art. 14 of the AI Act, AI HR systems should be designed and developed in such a way to be overseen by natural persons in order to prevent or minimise the risk to health, safety or fundamental rights during use or reasonably foreseeable misuse of the AI system; the implementation should be commensurate

[20]Wagner (2019).
[21]Almada (2019); Mulligan, Kluttz & Kohli (2020); Alfrink, Keller, Kortuem & Doorn, (2023) with a review of the relevant literature.
[22]AI systems are classified as "high-risk" if they meet the requirements stated in art. 6. In general, high-risk systems are supposed to pose a significant risk of harm to the health, safety and fundamental rights of natural persons.
[23]The four principles are: respect for human autonomy, prevention of harm, fairness and explicability. See HLEG (2019) at 12.
[24]HLEG (2019) at 16.
[25]*Ibid.*

to the risks, context, and level of autonomy of the system. By means of appropriate measures identified by the provider of the system, the deployer should assign the human oversight to a competent natural person enabled to monitor, understand the capacities and limitations of the system, to remain aware of the danger of "automation bias", to correctly interpret the output, to *decide not to use the AI System or to override or reverse the output,* to intervene in the operation or stop the system. Finally, in case of remote biometric identification, no action should be taken unless the output is confirmed by two competent natural persons[26].

Therefore, whatever the function of the system, whatever the output, there must always be a human monitoring, whose position and purpose seems to be very similar (and, actually, even wider) to the ones the current interpretation of art. 22 GDPR refers to in order to qualify a human intervention as able to exclude the complete automation of the decision. The line in italics above, in particular, seems to hit the precise point of what it takes for a human to be competent enough to avoid the application of art. 22.

To state the point clearer, in case of a decision taken by an AI HR system based solely on personal data, the requirements set forth by art. 14 AI Act seem to systematically prevent the application of art. 22 GDPR.

Therefore, this leads to the conclusion that high-risk AI systems can never be used to produce fully automated decisions in the sense of Article 22 GDPR itself.

At first glance, this conclusion may sound quite soothing. After all, HR systems are those supposed to endanger the safety, health and fundamental rights most, so the continuous human monitoring is a way to assure the recipients that someone is taking care of them.

However, the disapplication of art. 22 GDPR comes also with the impossibility to resort to the whole protection designed in that article, which includes – on the one hand – some peculiar safeguard measures, and - on the other – a much wider set of sensible aspects of the receiver to be taken into account.

Moreover, as we saw earlier, in art. 22 GDPR the possibility to use fully automated decision-making systems may depend on (the European or the member state Legislator or) the parties themselves, leaving to their autonomy the evaluation of the situation in terms of advantages, effectiveness of the safeguard measures as implemented by the data controller and so on.

When nothing of that sort applies anymore, is the recipient enjoying at least the same level of protection? Is the parties' autonomy still empowered?

Comparing the Levels of Protection

Subsequently, when it comes to using ADM systems processing personal data the AI Act endorses the position that high-risk AI systems can take decisions, even so invasive as to be above the threshold, and following the current interpretation of art. 22 GDPR, we should conclude that the HOP is to be considered a sufficient

[26]See art. 14 AI Act.

safeguard measure, since all the others prescribed by art. 22 GDPR, at this point are no longer available.

There are at least two points that need to be highlighted here.

First, what the natural person in charge of the human oversight is responsible for, according to art. 14 of the AI Act, is far less than what is supposed to be protected by the safeguard measures that must be set up following art. 22 GDPR.

In fact, in the first case, the human in command is supposed to be present to prevent or minimise risks to safety, health and fundamental rights in general that may be compromised during the use of the system or its foreseeable misuse[27].

Of course, we may easily presume that this provision, although stated in general terms, includes the consideration of those aspects even with reference to a specific recipient of an automated decision when this is the case.

Now, safety, health are fundamental rights, so the formula used in the AI act can be reduced without restricting the cognitive and normative content, to fundamental rights, while these are indeed basic rights.

But this is quite different from what art. 22 GDPR prescribes: there the aim of the measures is to safeguard "the rights, freedoms and legitimate interests" of the data subject[28].

Thus, the formula used in the GDPR is wider in scope and content and more precisely tailored towards the recipient of the decision.

And since one of the safeguard measures that can be invoked by the data subject is a human intervention, we can conclude that, following the discipline of ar. 22 GDPR, the recipient of an automated decision may, theoretically, enjoy a full assessment of its personal condition as determined by the machine with a natural person specifically devoted to this[29].

Instead, the formulation of art. 14 AI Act seems to indulge in a more formal, legalistic approach, where the human is there to assure a kind of general compliance.

But there is more, and this is the second point.

Art. 22 GDPR, assures also a right to express one's opinion which is not that very protecting but also a right to contest the decision which has been taken quite seriously in the recent literature[30]. The main reason for the re-evaluation of the right to contest is exactly the need to enforce the *ratio legis,* that is the protection of "rights, freedoms and legitimate interests" of the data subject. A contestation is more than a mere opinion, it is a defensive act which includes some requirements in order to be effective. One of those is the right to receive specific information about how it happened that the case was decided in a certain way. One cannot effectively contest

[27]Art. 14, §2, AI Act.

[28]Art. 22, §§2 and 3, GDPR.

[29]Art. 22 GDPR does not prescribe any specific behaviour to the intervening human, and this is something that raised some concerns about the actual protective capacity of the measure, see Sarra (2020a). However, since it is repeatedly asserted that the goal of all the measures is the safeguard of rights, freedoms and legitimate interests of the data subject, it is to be expected, at least theoretically, that this is the range of concerns that the natural person called upon should handle.

[30]Besides the literature already cited on point see also Vaccaro, Xiao, Hamilton & Katahalios (2021); Lyons, Velloso & Miller (2021); Alfrink, Kellet, Yurrita Semperena, Bulgin, Kortuem & Doorn (2024).

something one knows nothing about, and this may raise the stake of the obligations of the data controller quite a bit[31].

The reassessment of the right to contest the decision has been seen as the correct interpretative way to resolve the doubts about the existence in the GDPR of a *right to explanation* of the decision, which although acknowledged in *Recital 71*, is not mentioned in art. 22.

Curiously enough, the AI Act provides for a right to explanation (art. 86) but not a right to contest.

The fact is that after the approval of the GDPR, the attention to the value of explicability of automated decisions has increased considerably, with suggestions coming not only from the academia with the discussions cited, but also from national courts and European institutions along their way to building the current normative framework.

A notable example of a court decision endorsing the need for explicability is the Italian maximum administrative authority, in 2019, recognising that, in the context of assessing the legitimacy of administrative action that has made use of complex algorithmic tools, the "knowability of the algorithm must be guaranteed in all aspects: from its authors to the procedure used for its elaboration, to the decision-making mechanism, including the priorities assigned in the evaluation and decision-making procedure and the data selected as relevant. This is in order to be able to verify that the criteria, prerequisites and outcomes of the robotised procedure conform to the prescriptions and purposes established by law or by the administration itself upstream of that procedure, and so that the modalities and rules on the basis of which it was set up are clear - and consequently open to review"[32].

Conversely, the already cited *Ethics Guidelines for Trustworthy AI* endorses the principle of explicability meaning "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"[33]. Please note that both these documents correctly relate the explicability principle to revision and contestation.

Eventually, the right to explanation was introduced in the AI Act.

So, ironically, none of the safeguard measures explicitly prescribed by art. 22 GDPR are included in the AI Act, while the most dubious one is!

But even in this case, although the wording of art 86 makes clear reference to that of art. 22 GDPR, the protection acknowledged seems to be quite poor if compared to both the formulation of the right to explanation adopted by scholars in

[31]Pagallo (2018).

[32](Italian) C.d.S., 8472/2019.

[33]HLEG (2019) at 13. The document also acknowledges that "an explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate". Overall considered opacity is the product of a complex code, high-dimensionality of data and changing decision logic, see Burrell (2016); Mittelstadt, Allo, Taddeo, Wachte & Floridi (2016) at 6.

their discussion about art. 22, and the explanatory needs included – dynamically – in the right to contest.

First of all, the right to explanation as provided for by art. 86 AI Act does apply to HR systems but not to all of them: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity are excluded.

Secondly, although it refers to decisions which produce legal effects or similarly significantly affect a person (the very same threshold we saw in art. 22 GDPR), it is granted in case of threat to health, safety and fundamental rights. In other words, the right to explanation can be invoked as a protection of the same extension of the HOP, which, as we saw, is narrower than the scope of the safeguard measures in art. 22 GDPR.

Moreover, the supposed explanation seems to have a fixed content and be limited to information about the role of AI and the main elements of the decision[34] which is more restricted than required by the principle of explicability as related to the need of contestability.

Lastly, it is residual as § 3 states that such a right is given only if it is not provided for by other EU laws, which, in the case of the GDPR, is a problem because the existence of a right to explanation is discussed.

In conclusion, even considering the right to explanation as formulated in the AI Act, it seems that in case of HR AI systems taking decision by personal data processing, *if* the Human Oversight Principle is able to prevent the application of art. 22 GDPR, the recipient is left with reduced protection.

A Different Interpretation

To briefly recap the main point discussed so far: Following the usual interpretation of art. 22 GDPR, the HO principle as implemented in the AI Act makes art. 22 GDPR inapplicable for AI HR Systems taking decisions above the "threshold" on the base of personal data processing.

In this scenario, the recipient would enjoy the benefit of a thorough human monitoring of the system and, in case of a decision threatening his/her health safety and fundamental rights, s/he would be entitled to a right to explanation in the terms mentioned above.

However, both the focus of the HOP and the conditions of application of art 86 AI Act are limited as is the content of the explanation s/he would receive.

On the other hand, s/he could not take advantage of the more articulated and demanding discipline of the safeguard measures provided for in art. 22 GDPR.

This conclusion amounts also to a limitation of the parties' autonomy: as a matter of fact, art. 22, §2, a) and c) GDPR allow them to decide whether to waive the prohibition and take advantage of ADM or not. This decision is taken under the awareness that, in case the data controller has to warrant the safeguard measures

[34]art. 86, § 1.

particularly including a right to contest, the decision, in all, is needed to make it effective.

It may be noticeable that, as a consequence, the AI Act seems to modify the approach to the use of ADM: while art. 22 GDPR adopts a sort of *participatory model of rights protection* based around the contestability principle. In the scenario just presented, the receiver is left with a kind of paternalistic top-down approach, presuming that the human oversight – as implemented by the right to explanation set forth in art. 86 AI Act - is sufficient to guarantee the protection of the data subjects.

However, for the reasons just seen, this may not be the case.

But, all things considered, this conclusion is not inevitable. After all, it depends on an interpretation which raised no particular controversy at the time the GDPR was approved, and it is still repeated today although the normative landscape as well as the technological development have reached new levels of complexities.

As legal scholars, we need to think anew about how the single elements in this new socio-legal scenario interact with each other and direct our effort to offer a systematic framework in line with the fundamental principles at stake.

Although the GDPR and the AI Act are formally on the same hierarchical level as source of law, the latter states that the Union law on the protection of personal data should be applied in connection with the rights and obligations it lays down. Moreover, it acknowledges a general prominence to the GDPR by stating that its own application shall not affect it (art. 2, § 7). The only exceptions it makes, are not exceptions at all: the first one, art. 10, §5, gives the possibility to process special kinds of personal data for *bias* detection and correction, but it requires the application of all the conditions set forth in art. 9 GDPR and a lot more indicated in art. 10 itself.

As for art. 59 AI Act, the other apparent exception, it admits the further processing of personal data lawfully collected for other purposes in the context of so called "regulatory sandboxes". These are "controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision" (art 3, n. 55). But, again, since the faculty of more personal data processing is given under severe conditions, the first of which is that the AI system is developed for safeguarding substantial public interest, this can be easily seen as a special case of the legal base provided for in art. 6, § 1, lett. e) GDPR.

Therefore, apparently the AI Act endorses not only a general principle of mutual consistency between itself and the GDPR, but also a prominence of the latter.

As a consequence, in the absence of an explicit derogation, an interpretation that leads to the conclusion of a mutual exclusiveness of the regulations should be taken very cautiously and only when there is no other rational way to give sense to the framework.

Emphatically the current interpretation of art. 22, § 1, GDPR about what makes a decision based "solely on auromated pocessing" is untenable after the approval of the AI Act. In particular, it lacks the precision needed to make it work in accordance with the new Regulation on AI on behalf of the data subject receiving a decision elaborated by a high-risk AI System.

In other words, we need an updated interpretation of art. 22, §1, GDPR in order to reconcile it with the HOP in the AI Act, and to make both Regulations applicable, in the general cases when the decision is substantially determined by the machine.

The new interpretation should explicate the conditions under which the exercise of the human oversight is so focused as to exclude that the decision is "based solely on automatic processing", given that the mere general monitoring of the system and the abstract possession of the competence to do otherwise are not sufficient.

This may be done by interpreting the expression "solely based on automatic processing" in 22, §1, in a way that requires a specific and material involvement of the qualified human in order to avoid the prohibition.

And here is how it may sound:

> *A decision should be considered fully automated whenever it is taken by a machine, despite the presence of a qualified human monitoring the operation unless it is proven that such a qualified human has specifically and materially intervened in the individual decision-making process by significant acts that, if absent, would have led to a different decision.*

Conclusive Remarks

Let us see now the consequences of the proposed interpretation in terms of the complete discipline of ADM in light of both the GDPR and the AI Act.

First of all, in the ordinary course of action the implementation of the HOP would not prevent, by itself, the application also of art. 22 GDPR, assuming the HR AI system is used to process personal data. This would let the data subject enjoy the protection of both the GDPR and the AI Act as well, which is more in line with the principle stated in the AI Act of application without affecting the EU regulations on personal data.

Secondly, in that very same scenario, decisions above the threshold would face the general prohibition set forth in art. 22 GDPR, § 1, unless the data controller can give actual evidence of specific human steps taken in the path that led to the decision. Of course, this makes the burden of proof heavier for the data controller, but it is in line with the accountability principle which is at the core of the GDPR. As a consequence, the relative duties the data controller already has to comply should also include the accurate record of any specific human intervention during the ordinary course of application of the human oversight principle.

Thirdly, since the prohibition can be waived by the parties, this interpretation gives back to them (especially to the data subject) all the autonomy to evaluate and decide if they are to let the automation governing their relationships or not, instead of being forced to over rely on the efficiency of the person in charge of the HOP. This consequence is in line with the ethical principle of respecting human autonomy included in the *Ethical guidelines for trustworthy AI.*

Fourthly, in the ordinary case of a decision above the threshold taken by an HR AI system processing personal data, the HOP would not prevent the data subject from enjoying all the safeguard measures provided for in art. 22, § 3, GDPR.

However, in this case, the specific application of those measures undergoes a slight shift except, perhaps, for the "right to express one's opinion" which remains as defensively insignificant as it has always been. Instead, the "right to a human intervention" would take the form of a right to have the person in charge of the HO – since ex art. 14, AI Act, s/he is in the right position and has the due competence - to take specific care of the individual situation giving specific reasons in case of confirming the decision.

As for the "right to contest the decision", as we saw, it has an intimate link with the "right to explanation". As a matter of fact, this link goes in a double direction. On one side, the only way for such a right to have the sense it is supposed to have, meaning to be a safeguard measure for rights, freedoms and legitimate interests, is to guarantee the data subject access to specific information about how the decision was made. In this sense, the right to explanation is included in the right to contest and cannot be otherwise.

Conversely, since the contestation is a specific claim about why a decision should be changed, it is very focused on the individual case, thus giving a concrete measure to the amount of explanation needed. In other words, the quantity and quality of explanation to be given should not be measured in abstract or with reference to some objective technical standard, but it should vary according to the need to deal with the specific controversy[35].

But here, we are facing a problem:

As we saw the AI Act, while granting a (almost) general right to explanation in case of decision taken by the deployer based on the output of an HR AI system, it also states that this right is residual: it can be invoked to the extent that it is not otherwise provided for under the EU Law[36].

Now, the GDPR provides that *indirectly:* in other words, as we have shown, the right to explanation is given as a pre-requisite for the right to contest to be effectively actionable. But this is, at the moment, a doctrinal reconstruction and has a drawback: the data subject is allowed an explanation as long as s/he exercises the right to contest. The existence of a right to explanation *per se* remains dubious.

In this situation, and in the scenario considered, that is an HR AI system taking a decision by processing personal data, can the data subject invoke the right to explanation ex art. 86 AI Act, without contesting the decision? In other words, does s/he have a right to explanation *quo talis?*

In answering this question, I would remain cautious between the boundaries of the *litera legis*, since we do not have any court specific decision on the subject as yet.

In my opinion, since art. 86, § 3 AI Act, states that the right there provided for is given "to the extent that" it is not otherwise provided for, and since the GDPR *does not* give such a right *per se,* I would conclude that the question should be

[35]See Sarra (2020b). This is the only way to make sense of the different wording between *Recital 71* and art. 22, §3, GDPR, and in the meantime to give proper value to the right to contest which is explicitly stated in both.
[36]Art. 86, § 3.

answered positively, but the content of the explanation s/he would receive is strictly limited to that envisaged in art. 86, § 1.

But, since, because of this limitation, this is not supposed to act as a "safeguard measure" in the wide and deep sense of art. 22 GDPR, the data subject is entitled *also* to a deeper level of explanation in case s/he decides to contest the decision. In this case, the content of the explanation required is not fixed *a priori* but it would depend on the specific claims advanced.

To conclude this paper, it is worth reminding that a human-centric AI is a challenge at the moment. The complexities of the many regulations do not let us rely on some pre-ordered formula, but we need to be prepared for the many judicial controversies that we are about to see as soon as the AI Act will be fully applicable.

This paper is nothing but a drop in the ocean. It may be useful to alert about the interpretative work to be done. Perhaps, it may appear as a low-profile old school dogmatic work, and not the "epochal" warning about a coming revolution or the big announcement of the "age of the machines".

But when the wave comes, perhaps it will be better to rely more on solid levees built by low-profile craftsmen that on self-proclaimed prophets of a new era.

References

Alfrink, K., Kellet, I., Yurrita Semperena, M., Bulgin, D., Kortuem, G. & N. Doorn (2024). 'Envisioning Contestability Loops: Evaluating the Agonistic Arena as a Generative Metaphor for Public AI' in *She Ji: The Journal of Design, Economics, and Innovation*, 10(1): 53–93. https:// doi/org/10/gtzwft

Alfrink, K., Keller, I., Kortuem, G. & N. Doorn (2023). 'Contestable AI by Design: Towards a Framework' in *Minds & Machines* 33: 613–639.

Almada, M. (2019). 'Human intervention in automated decision-making: Toward the construction of contestable systems' in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2–11. ICAIL '19. Montreal, QC, Canada: Association for Computing Machinery.

ART29WP (2018). *Guidelines on Automated individual decision-making and Profiling for the Purposes of Regulation 2016/679*, last Revised and Adopted on 6 February 2018.

Brkan, M. (2019). 'Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond' in *International Journal of Law and Information Technology* 27(2):91–121.

Burrell, J. (2016). 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' in *Big Data & Society* 3(1):1–12.

Bygrave, L.A. (2019). 'Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making' in *SSRN Scholarly Paper.* https://papers.ssrn.com/abstract=3329868.

Bygrave, L.A. (2021). 'Automated Profiling: Minding The Machine: Article 15 Of The EC Data Protection Directive And Automated Profiling' in *Computer Law & Security Review* 17(1):17–24.

De Hert, P. & V. Papakonstantinou (2016). 'The New General Protection Regulation: Still a Sound System for the Protection of Individuals?' in *Computer Law and Security Review* 32:179-194.

Edwards, L., & M. Veale, (2017). 'Slave to the Algorithm? Why a "Right to an Explanation"

Is Probably Not the Remedy You Are Looking For' in *Duke Law and Technology Review* 16 (1):1–65.

European Commission, (2020). *A European Strategy for Data,* Communication by the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, February 22nd, 2020

Friedman, B. and Nissenbaum, H. (1996). 'Bias in Computer Systems' in *ACM Transactions on Information Systems* 14(3):330–347.

Goddard, K., A. Roudsari, & J.C. Wyatt. (2012). 'Automation bias: a systematic review of frequency, effect mediators, and mitigators' in *Journal of the American Medical Informatics Association* 19(1):121–127.

Goodman, B & S. Flaxman. (2017). 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation' in *AI Magazine* 38(3):50–57.

HLEG (High-Level Expert Group on sustainable finance) established by the European Commission in 2016.

High-Level Expert Group on Artificial Intelligence (2019). *Ethics Guidelines for Trustworthy AI.* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en

Larus, J., Hankin, C., Carson, G.S., Christen, M, Craga, S., Grau, O., Kirchnet, C., Knowles, B., McGetttick, A., Tamburri, D.A. & H. Werthner (2018). *When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*. ACM. https://dl.acm.org/doi/book/10.1145/3185595.

Lyons, Henrietta, Eduardo Velloso and Tim Miller. "Conceptualising Contestability." *Proceedings of the ACM on Human-Computer Interaction* 5 (2021):1 - 25. https://doi.org/10.1145/3449180

Malgieri, G. & G. Comandé (2017). 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' in *International Data Privacy Law* 7(4):243–65.

Mendoza, I., & L.A. Bygrave (2017). 'The Right Not to Be Subject to Automated Decisions Based on Profiling' in T.-E. Synodinou, P. Jougleux, C. Markou, & T. Prastitou (eds.) *EU Internet Law: Regulation and Enforcement*, 77–98. Cham: Springer International Publishing.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachte, S. & L. Floridi (2016). 'The Ethics of Algorithms: Mapping the Debate' in *Big Data & Society* 3(2):1-21.

Mosier, K.L. & L.J. Skitka (1999). 'Automation Use and Automation Bias' in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(3):344–348.

Mulligan, K.D., Klutt,z D.N. & N. Kohli (2020). 'Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions' in Werbach, K. (ed.), *After the Digital Tornado. Networks, Algorithm, Humanity*, 137-151. Cambridge: Cambridge University Press.

Pagallo, U. (2018). 'Algo-Rhythms. The Beat of the Legal Drum' in *Philosophy and Technology* in 31(4):507-524.

Preece, R. (2018). 'The GDPR accountability principle and the use of scenario workshops in the digital age' in *Journal of Data Protection & Privacy* 2:34–40.

Roig, A. (2018) 'Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)' in *European Journal of Law and Technology* 8(3):1-17.

Sarra, C. (2020a). 'Defenceless? An Analytical Inquiry into the Right to Contest Fully Automated Decisions in the GDPR', in David A. Frenkel & Anna Chronopoulou (eds.), *An Anthology of Law,* 235 – 252. Athens, ATINER;

Sarra, C. (2020b). 'Put Dialectics into the Machine: Protection against Automatic-decision-making through a Deeper Understanding of Contestability Design in: Global Jurist' in

Global Jurist. https://doi.org/10.1515/hj-2020-0003
Skitka, L.J., Mosier, K. & M.D. Burdick (2000). 'Accountability and automation bias' in *International Journal of Human-Computer Studies* 52(4):701–717.
Sweeney, L. (2013) 'Discrimination in Online Ad Delivery' in *Commun. ACM* 56(5): 44–54.
Tsamados, A., N. Aggarwal & J. Cowls (2022). 'The ethics of algorithms: key problems and solutions' in *AI & Soc* 37:215–230.
Vaccaro, K., Xiao, Z, Hamilton, K. & K. Katahalios (2021). 'Contestability For Content Moderation' in *Proc. ACM Hum.-Comput. Interact* in 5(CSCW2): 318:1-28. https://doi.org/10.1145/3476059.
Veale, M., & L. Edwards (2018). 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling' in *Computer Law & Security Review* 34, n. 2: 398–404.
Wachter, S., Mittelstadt, B. & L. Floridi (2017). 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' in *International Data Privacy Law* 7(2):76–99.
Wagner, B. (2019). 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' in *Policy & Internet* 11(1):104–22.
Wagner, B. (2020). 'Accountability by design in technology research' in *Computer Law & Security Review* 37:105398.