Acta Univ. Sapientiae, Legal Studies, **11**, 2 (2022) 65–80

DOI: 10.47745/AUSLEG.2022.11.2.04

Some Remarks on the 'AI Judge' in the Context of Recent European Union Regulatory Action

János SZÉKELY
PhD, Senior Lecturer
Sapientia Hungarian University of Transylvania, Department of Legal Science
ORCID: 0000-0003-4254-2054
e-mail: szekely.janos@kv.sapientia.ro

Abstract. The utilization of artificial intelligence (AI) as an aid during adjudication is no longer a future prospect but a reality. While current implementations of the technology are as of yet far removed from the future science fiction would have us fear, the prospect of the 'AI judge' must now be seriously considered. In our analysis, we investigate whether such a prospect would be compatible with fundamental rights and proposed EU norms set to govern the use of AI technology. We also examine the ethical requirements for utilizing AI, including in the judicial domain. We find that in lack of a possibility of granting a reasoned decision, in the course of a transparent procedure AI fails to meet the basic requirements that would allow its use under current and predictable future regulatory conditions during adjudication in the European Union. We further find that the shortcomings of the technology and the regulatory environment would hinder the accountability required for implementing the 'AI judge'. We conclude that the specific needs of adjudication have not been duly considered during the preparation of EU instruments in the field of AI, and further regulation as well as research will be necessary.

Keywords: artificial intelligence, ethics, courts, fundamental rights, fair trial, European Union

1. Introductory Remarks

Artificial intelligence (AI) has become something of a catchphrase in almost all disciplines of science and even the arts in recent years. That it holds great promise and equally great risks stands beyond any doubt. Yet, until lately, national legislatures as well as international organizations have been reluctant to propose regulation that would either direct or hinder the development or deployment of

AI to certain tasks; compulsory rules have only been formulated and implemented regarding personal data protection.[1]

This 'silence of the legislator' is rapidly becoming untenable. The development of AI applications now appears to have definitively left behind the era of false starts and sudden stops that have plagued the technology in the past.[2] Practical and economically viable AI solutions are already in use, or they are, at the very least, rapidly emerging. Thus, the period of 'salutary neglect' by regulators is at its end: on both shores of the Atlantic, they now aim to set forth new AI laws as soon as possible.[3] In this drive to legislate, the European Commission has tabled two regulatory drafts to constitute the cornerstones of European AI law: the proposed Artificial Intelligence Act[4] (AIA) and the proposed AI Liability Directive.[5] These proposals, still in the course of development and subject to public debate, aim to channel AI development in EU Member States according to the precautionary principle and implement fault-based liability aided by some presumptions of misconduct and causation in case AI systems should cause damage during their intended or unintended functioning.

One of the possible implementations in which AI presents great potential is that of dispute resolution, either as an instrument for aiding judicial decision-making in one way or another or as an autonomous AI adjudicator.[6] In our study, we aim to analyse the compatibility of these two forms, and specifically the 'AI judge' with the way the European Commission, a body of the European Union, currently envisages AI regulation in order to predict the possible future(s) of AI-based or AI-aided dispute resolution in the European Union.

2. Principles of AI and the 'AI Judge'

During the development of the AIA, the European Commission convened the High-Level Expert Group on Artificial Intelligence (AI HLEG), which would develop the main guidelines for the proposed regulation. The AI HLEG formulated a 'European' approach to AI based on three guiding principles: 1. compliance with the law, 2. fulfilment of ethical principles, and 3. robustness. These principles were then expanded into the following list of assessment criteria: 1. human agency and oversight; 2. technical robustness and safety; 3. privacy and data governance;

1 di Carlo–De Bondt–Evgeniou 2021.
2 Francesconi 2022.
3 Casovan–Shankar 2022.
4 European Commission, Directorate-General for Communications Networks, Content and Technology 2021.
5 European Commission 2022.
6 See Szekely 2019.

4. transparency; 5. diversity, non-discrimination, and fairness; 6. societal and environmental well-being; 7. accountability.[7]

In grounding EU AI law in principles such as these, the AI HLEG did not propose any particular rules for AI-based adjudication. Still, some basic conclusions can already be drawn from the list of principles and requirements regarding the future regulatory landscape of AI-based adjudication. Firstly, the principle of compliance with the law as outlined by the AI HLEG seems to mirror some of the guarantees associated with a fair trial (compliance with a procedure conducted according to the law by a court that is itself established under the law).[8] Secondly, through requirements such as human agency and oversight, as well as transparency, non-discrimination, fairness, and accountability, the AI HLEG set forth the framework with which AI-based adjudication would need to take place. This framework is not all that distant from that found in other international instruments such as the European Convention on Human Rights[9] (specifically articles 6 and 14) or the Charter of Fundamental Rights of the European Union[10] (articles 21 and 47).

We should note here that the AI HLEG, as opposed to the drafters of these other instruments, did not limit itself to only enumerating desiderata without presenting how they should actually be achieved and when they are considered to have been achieved. In fact, it also elaborated a document with the title *Ethics Guidelines for Trustworthy AI*[11] and another one with the title *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*.[12] Both are useful for exploring the notions that underpin the AIA and in part also the AI Liability Act.

2.1. Desiderata of an Ethical and Trustworthy AI in the AI HLEG Preparatory Documents

Observance of ethical principles during the development of AI systems and the coordinates of their trustworthiness according to the AI HLEG must be subject to constant monitoring and later to review during the operation of the AI system. Ethical principles according to the AI HLEG are deemed to have been adequately considered[13]

7 The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2019.

8 Guide on Article 6 of the European Convention on Human Rights. Right to a Fair Trial (Civil Limb) 2022. 58–59. See the case-law of the European Court of Human Rights as cited in the Guide, for example, in cases *Guðmundur Andri Ástráðsson v Iceland* (GC), 2020, §§ 207 and 211 as well as *Pasquini v San Marino*, 2019, §§ 103 and 107 (court established by law, judges appointed according to the law); *Xero Flor w Polsce sp. z o.o. v Poland*, 2021, §§ 245–251 (court operating with impartiality and independence according to the law).

9 European Convention on Human Rights 1950.

10 Charter of Fundamental Rights of the European Union 2012.

11 Independent High-Level Expert Group on Artificial Intelligence 2019.

12 Independent High-Level Expert Group on Artificial Intelligence 2020a.

13 Independent High-Level Expert Group on Artificial Intelligence 2019. 9.

if the 'moral and legal entitlements' (specifically those constituted by fundamental rights enshrined in EU Treaties and the EU Charter) have been considered during its development. Thereby, any AI development must by definition be human-centric in keeping with the higher-order principles of human dignity, freedom of the individual, respect for democracy, justice and the rule of law, equality, non-discrimination, solidarity, as well as respect for citizen's rights.[14] These principles are translated by the AI HLEG into the world of AI by the desiderata of respect for human autonomy, prevention of harm, fairness, and explicability.[15] In this context, human autonomy is understood as being respected when AI is not utilized to manipulate, coerce, or otherwise direct the behaviour of humans to an 'unjustifiable' degree, and it must ensure human oversight of the operation of AI systems. Such systems must be designed not to cause harm to human beings, or to 'exacerbate' harms already caused, and they should be designed so as to prevent possibilities of abuse. Furthermore, AI must be developed with fairness. The AI HLEG distinguishes here between substantive (material) and procedural (formal) meanings of fairness. In the material sense, fairness would demand an equitable distribution of gains and risks, as well as the desiderata of proportionality between means and ends, while ensuring non-discrimination (both as bias and as stigmatization). Importantly for AI adjudication, procedural fairness would entail a right to an effective remedy against adverse AI outputs, even if they have been authorized by a human operator. This latter desideratum strongly associates procedural fairness with explicability. It is here that we find the Achilles heel of the ethics-based approach by the AI HLEG, as on the question of explicability a major compromise takes place. The desideratum of explicability is defined thus:

> (…) 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. 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.[16]

14 Independent High-Level Expert Group on Artificial Intelligence 2019. 10–11.
15 Id. 12–13.
16 Id. 13.

Simply put, explicability may be considered as observed when 'the system as a whole respects fundamental rights'. The question may justly be asked: what does the AI HLEG mean by a system which, while not subject to explicability due to the very mechanism of its operation, still respects fundamental rights? To illustrate the problem, we would like to point out that the way things stand, as in the case-law of the European Court of Human Rights, the right to a reasoned (i.e. 'explicable') decision is in and of itself a fundamental right.[17] It is also known that AI may be non-transparent by its very nature, i.e. unable to provide reasons for even correct outputs (a problem referred to by the AI HLEG). This problem is known in the literature as 'opacity'.[18]

The tension between the principle of human autonomy and prevention of harm is emphasized by the AI HLEG;[19] however, the expert group seems to have ignored the fundamental synergy between explicability and fairness (constituting two facets of the right to a fair trial in the judicial context) when considering the possibly opaque way in which AI operates.[20] Even more disturbingly, the fall-back solution proposed by the AI HLEG for cases of opaque AI operation, called 'other explicability measures', also fails to address this issue, as no amount of traceability and auditability of an AI system will result in obtaining a true 'reasoning' from a 'black box' AI. Such measures may help with attaining output-based legitimacy[21] of the AI adjudicator (by e.g. verifying during an audit that a human judge would have reached a similar solution or that the solution is in keeping with the prevailing case-law); however, this will do nothing to ensure a reasoned decision. It seems that while for some other applications development and use of non-transparent AI may be considered ethical, based on the de minimis set of fall-back measures, these are not apt for solving the tension between explicability and fairness in the case of adjudication: fairness in the procedural sense is unattainable without explicability. This results prima faciae in an incompatibility of opaque AI with applications in the field of adjudication.

2.2. The ALTAI Assessment Criteria and Their Potential Impact on AI Used for Adjudication

The ALTAI assessment criteria also emphasize respect for fundamental rights (even if the right to a fair trial is not mentioned).[22] The first assessment criterion (Requirement #1) of the list, which verifies conditions of human agency and

17 Guide on Article 6 of the European Convention on Human Rights. Right to a Fair Trial (Civil Limb) 2022. 96–97; Fink 2021.

18 See Wischmeyer 2020.

19 Independent High-Level Expert Group on Artificial Intelligence 2019. 13.

20 For a detailed presentation of the most significant problems posed by lack of explicability in AI adjudication, and the requirement of explainable AI (xAI), see Deeks 2019. For a critique of explicability as understood by the European legislator (i.e. in senses other than a fully human-readable, clearly reasoned decision), see Edwards–Veale 2017. 65 et seq.

21 Chesterman 2021. 275.

22 Independent High-Level Expert Group on Artificial Intelligence 2020a. 5–6.

oversight, asks, inter alia, whether the AI system may generate the overreliance of human operators (a risk known as automation bias).[23] This criterion is based on the foregone conclusion that automation bias exists as a phenomenon and may constitute a problem in the realm of AI use.[24] Already in this stage of assessment, the possible 'black box' operational model, the inability of AI to produce the reasoning for its output poses problems. Evaluation of the risk of automation bias is strongly linked to the output-based legitimacy of the AI system, whereas, as recent research has pointed out, human operators tend to be biased by automated systems (just as they tend to do in case of human advice) selectively,[25] i.e. when the advice given is in line with their pre-existing biases. This manifestation of the automation bias is more difficult to guard against when no reasoning for AI output is present.

As part of Requirement #1 of the ALTAI assessment criteria, human oversight of the AI system also must be evaluated.[26] Four situations of such oversight are considered as possible by the AI HLEG: fully autonomous systems, human-in-the-loop (HITL), human-on-the-loop (HOTL), and human-in-command systems (HIC).[27] By projecting these operational models onto possible AI implementations in adjudication, we may find that full automation is not an option (as the human factor must be relied upon at least during the enforcement phase of some judicial decisions), and HOTL solutions would make it impossible for the human factor to be effectively involved in the adjudication activity undertaken by the AI outside the design and operational monitoring phases of implementation. HITL and HIC solutions are the most likely compliance options with Requirement #1.

The main ALTAI criterion that must be considered when contemplating AI adjudication is Requirement #4, which refers to transparency.[28] This is actually constituted of a subset of several coordinates for evaluation, namely: traceability, 'explainability' (the exact wording used in ALTAI Requirement #4, for which we shall use the notion of 'explicability' in line with the terminology in the AI HLEG Ethics Guidelines for Trustworthy AI), and communication. Traceability refers to the ability to document the source of the data, the content of the procedures (models)

23 For a discussion on automation bias, see Skitka–Mosier–Burdick 1999.

24 For a recent discussion on automation bias during use of AI applications, see Zuiderveen Borgesius 2020.

25 Alon-Barkat–Busuioc 2022.

26 Independent High-Level Expert Group on Artificial Intelligence 2020a. 8.

27 'Human-in-the-loop refers to the capability for human intervention in every decision cycle of the system. Human-on-the-loop refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. Human-in-command refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal, and ethical impact) and the ability to decide when and how to use the AI system in any particular situation. The latter can include the decision not to use an AI system in a particular situation to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by an AI system.' Independent High-Level Expert Group on Artificial Intelligence 2020a. 8.

28 See Independent High-Level Expert Group on Artificial Intelligence 2020a. 14–15.

used in generating a given output by the AI system, and the quality assessment of the output, as well as the logging of outputs in the form of AI decisions or recommendations. The AI HLEG ALTAI criteria define explicability in a manner identical to that found in the *AI HLEG Ethics Guidelines for Trustworthy AI*, which have been presented above, but for its assessment they only provide for the AI operator to state whether it explained (whenever possible) the reasons for the AI output to the party these have later affected. Finally, the AI operator must communicate to the party subject to the activity of an AI that the party's interlocutor is an AI entity. This last criterion in the case of adjudication should involve a prior warning to parties that their case will be subject to automated adjudication or an AI entity will provide feedback to the adjudicator regarding the solution that is to be given.

Requirement #5 of the ALTAI criteria emphasizes diversity, non-discrimination, and fairness. Under this criterion, avoidance of unfair bias must be achieved through taking the necessary measures both in algorithm design and when compiling the dataset used for 'training' the AI, which must be representative of the persons subjected to the operation of the AI. Avoiding bias also involves monitoring outputs for detecting potential bias. When developing and monitoring AI, fairness must be assessed after considering several definitions of fairness and consulting the 'impacted communities' regarding the operation of the AI (an action which must be assessed not just as an element of fairness but also when verifying that stakeholder participation in AI development took place). The requirement for considering several definitions of fairness is problematic when AI is to be employed as an autonomous adjudicator or (more likely) a guide for a human judge. As aptly observed by John Rawls when developing his theory of justice,[29] fairness is by no means a unitary concept and may even depart from the 'contractarian' view of equality between the parties. Utilitarian and intuitionist views of justice (mainly manifested in the prioritization – based on various criteria – of interests to be conserved, before those that must be disregarded when faced with the need to resolve a problem such as a judicial dispute) may even collide.[30] A recent meta-analysis[31] of studies regarding measures taken to ensure the fairness of AI systems has also shown that this is more difficult to attain than simply filtering protected attributes (e.g. ethnicity, individual, social, and economic attributes, etc.) from the data and would require contrasting individual aspects of fairness with the currently engrained notion of collective fairness, something which cannot be done without a yet non-existent definition drawn from international human rights (case-)law. The ALTAI criteria impose considering several definitions of fairness, whereas they ignore the possibility that for some AI applications, such as adjudication,

29 See Rawls 1999. 15 et seq.
30 See Rawls 1999. 25–45.
31 Varona–Lizama-Mue–Suárez 2021.

a unitary definition (along the lines of the ideas present in the AI HLEG Ethics Guidelines for Trustworthy AI, perhaps defined in human rights case-law) should be utilized and, in fact, also standardized.

Requirement #7 of the ALTAI criteria imposes accountability on AI systems and their operators.[32] Auditability as the first component of accountability is considered as conceptually equivalent to traceability (ensuring sufficient documentation of the AI in its design and implementation to prevent and detect unintended functions), which is complemented by the possibility of independent third-party review in the form of an audit. It is here that we must consider the way in which 'auditability' during the judicial process is implemented: by means of the appeals process available to the parties. This type of specific 'auditability' is inexorably linked to the explicability of the output, not simply the functional oversight of the AI design and operation, as without a human-readable reasoning, no 'audit' of the legality of judicial decisions can be undertaken. Risk management is also paramount in the ALTAI criteria. This requires monitoring and reporting on potentially hazardous consequences of AI operation, with the involvement of third parties (including members of civil society). This again seems to contrast with the basic characteristic of the administration of justice as a highly specialized activity undertaken by highly specialized personnel engaged in providing a public service while also exercising public authority (imperium).

Keeping the problems identified above in view, it is in no way surprising that the AI HLEG when analysing regulatory requirements for AI in the law enforcement and justice sectors in another document, titled *Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence*,[33] concluded that even though there is a marked need for such applications, including during the organization of case material and the supervision of judicial outcomes to detect bias…

> deployment at greater scale generates risks and opportunities that are not yet fully understood. More research, scrutiny and deliberation are needed prior to formulating legal, ethical or policy guidance. It would be, therefore, important to launch a wide-spread policy debate in Europe (and beyond) on the development, use and impact of AI-assisted and AI-enabled decision-making systems in justice and law enforcement.[34]

The AI HLEG practically abdicated from considering specific policies for such systems at the current moment. This procrastination did not prevent the European Commission from proposing regulation for such situations in the AIA (even if the regulation – as we shall see – is rather scant).

32 Independent High-Level Expert Group on Artificial Intelligence 2020a. 21–22.
33 Independent High-Level Expert Group on Artificial Intelligence 2020b. 10–11.
34 Id. 11.

3. Principles Governing the 'AI Judge' as Reflected in the Language of the AIA

The European Commission recognized in the very text of the AIA, in Recital (3), that AI may contribute greatly to the administration of justice inter alia. In Recital (40) of the AIA, however, the Commission proposed that such systems should be considered as high-risk AI…

> considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is appropriate to qualify as high-risk AI systems intended to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

In the very same recital, the Commission attached a caveat to this proposal, stating: 'Such qualification should not extend, however, to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks or allocation of resources.'

Thus, the Commission instituted a two-tiered approach to AI regulation when administration of justice is concerned. AI used during adjudication (finding facts, interpreting and applying the law to them) or which assists such activities must be deemed high-risk, while AI used only in organizing the administration of justice is to be considered low-risk. We tend to find this approach imperfect. The activity of (civil) courts during adjudication far exceeds finding and interpreting facts, and then applying the law, as has been shown[35] in the literature, also involves considering the parties' submissions, organizing admissible evidence, or documenting the case based on legal literature, prevailing doctrine, and case-law in a fundamentally adversarial procedure. These activities all influence the outcome of adjudication, and any AI involved in them, even in a purely 'ancillary administrative' capacity, should be considered as high-risk.

Annex III of the AIA (which lists high-risk AI implementations according to Article 6(3) of the regulation) at point 8(a) specifically refers to 'AI systems intended to be used by a judicial authority or on their behalf to interpret facts or the law and to apply the law to a concrete set of facts'. The same critique that may be levelled against Recital (40) is also applicable in this case. Article 6(3) of the Regulation imposes considering AI implementations listed in Annex III as high-

35 See Gerber 2002.

risk 'unless the output of the system is purely accessory in respect of the relevant action or decision to be taken and is not therefore likely to lead to a significant risk to the health, safety or fundamental rights'. Whether any given output is accessory is set to be decided by the European Commission through the adoption of implementing instruments to the Regulation. Therefore, some clarity of whether any given AI applied to adjudication is or is not high-risk shall only be achieved once these norms have also been adopted.

Both Recital (40) and the context of Annex III of the AIA seem to hint that AI participation during adjudication was considered by the Commission as a possibility subject to regulation, even though the AI HLEG in its Sectoral Considerations on the *Policy and Investment Recommendations for Trustworthy Artificial Intelligence* specifically stated that further research is necessary before guidance on the topic (including its regulation) may be offered.

Chapter 2 (articles 8–15) of the AIA sets forth the specific rules for utilizing high-risk AI implementations. Analysing those rules contained in this chapter that are of interest to our inquiry, we may quickly ascertain that their content is only partly reminiscent of those contained in the AI HLEG preparatory documents.

In order to ensure fairness, for example, Article 10(3) of the AIA imposes that datasets used must be statistically representative. Article 10(4) furthermore requires that these consider the purpose of the AI as well as the location and the 'behavioural or functional setting' in which it will be deployed. Article 10(5) makes bias monitoring compulsory for providers of the AI implementation in the measure required to assure intended functioning, a purpose for which processing of data that would be otherwise prohibited by the GDPR is allowed.

Article 13 (with the marginal title *Transparency and Provision of Information to Users*) of the AIA markedly departs from the exigence of transparency as proposed by the AI HLEG. The text only requires that during the design and operation of the AI system interested parties be informed of the major characteristics of this system, of its abilities and limitations as well as the identity of the operator. No requirement of explicability is set forth at all. This solution, called by some authors in its version present in the GDPR as a 'transparency fallacy',[36] is in no way compatible with the notion of a reasoned decision, as it only provides possible technical information on 'how' but not on 'why' the AI has reached a given decision.

Human oversight (Article 14) was regulated in the AIA; however, the human 'overseer's' abilities to monitor and influence the AI system are permitted to be limited by the nature of the system (by the 'as appropriate to the circumstances' clause in Article 14(4) first sentence of the AIA). Human oversight is thus mainly relegated to monitoring, constant awareness of possible automation bias, and interpreting outputs (Article 14(4)(a) to Article 14(4)(c)). The human 'overseer' must also have the ability not to stop utilizing the AI system at any time 'or otherwise

36 Edwards–Veale 2017. 65–67.

disregard, override or reverse the output of the high-risk AI system' (Article 14(4)(d)) and 'to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure' (Article 14(4)(e)). These two requirements impose the HIC model of human oversight to AI. As such, they seem to exclude any AI with a function in adjudication from operating independently of a highly human-qualified factor (such as a judge) who oversees its activity, as the AI's output must remain under human control. We should keep in mind here that this requirement, just as the rules cited above, is enacted subject to the 'as appropriate to the circumstances' clause and may as such be dispensable.

Record keeping (Article 12), as well as the existence of an appropriate risk management system (Article 9), based on ex ante analysis of risks and on monitoring operation, are also provided for; these rules are, however, not central to our inquiry.

4. Accountability for Malfunctions of the 'AI Judge'

Accountability for the malfunctions of an AI (manifested as erroneous output or lack of output) is underscored several times by the AI HLEG documents and the vast majority[37] of similar works. Yet these documents fail to elaborate on the practical ways in which accountability should be achieved, apart from emphasizing pre-emptive measures and monitoring, with little to no mention of ex post facto liability issues. In fact, a recent meta-analysis of concepts used in most, if not all, international policy documents on AI to date (2022) showed that while the notion of 'accountability' is quasi-ubiquitous,[38] this is not correlated with 'liability'. The latter is in fact conspicuously absent from among the regulatory priorities in the field of AI.

This holds all the truer for AI employed during or for the purposes of adjudication. A good example for this phenomenon is the *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* developed by the European Commission for the Efficiency of Justice (CEPEJ). This evidently partisan policy document, which aims to represent exclusively the interest of the judiciary in AI development, mentions the notion of 'liability' only twice: once in the context of European Court of Justice case-law, which requires that Member States compensate damage resulting from the egregious breach of Community law by their courts, and another time, when discussing the risks of liability presented by AI in the case of judges who chose to decide against predictive algorithms.[39] The main form in which accountability is manifested under civil law would, of

37 Lupo 2022. 628.
38 Lupo 2022. 627 et seq.
39 European Commission for the Efficiency of Justice 2019. 24, 56.

course, be civil (i.e. pecuniary) liability, yet this concept seems somehow alien to most AI policy documents.

The proposal for an AI Liability Directive that was tabled by the European Commission aims to treat non-contractual civil liability issues arising out of the operation of AI systems. The scope of the proposed directive (Article 1) is not limited to liability between subjects of private law, and therefore its use, if adopted, will be conceivable in cases when liability for the erroneous results of AI-aided or AI-generated outcomes during adjudication are concerned. The AI Liability Directive primarily imposes obligations on the 'provider' of an AI system (Article 2(3) of the directive). The provider is defined in Article 3(2) of the AIA as being 'a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge'.

In some cases, the AI Liability Directive may also be employed against the 'user' of an AI system (Article 2(3) of the directive), as defined in Article 3(4) of the AIA: 'any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity'. It is this latter situation in which most AI systems in the administration of justice will be utilized.

The European Commission, when drafting the AI Liability Directive, could have opted for a model of strict, or 'no-fault', liability or for imposing compulsory insurance on AI system providers and users for cases when damage is caused in a non-contractual setting. These policy options, though considered during the impact assessment of the Directive by the Commission (and explicitly referred to in the Explanatory Memorandum published as the introductory part of the proposed Directive), were discarded in favour of maintaining the 'fault-based' liability model, which constitutes the general rule of non-contractual liability in most Member States.

The difficulty posed by this model is that it requires that both proof of fault and proof of causation between the fault and any damage be provided by the aggrieved party. In the case of AI systems, such proof is near-impossible due to the substantial number of participants during the development of the systems, the proprietary nature of some, if not most, components, and its characteristic autonomy that makes some AI outputs (lack of output) less than perfectly predictable. To ease the evidentiary requirements and facilitate compensation for damage caused by AI systems, the AI Liability Directive institutes two presumptions and a requirement to provide evidence for parties providing or operating high-risk AI systems.

As a first measure, Member States are required to ensure that their courts may order disclosure of evidence by the defendant if a high-risk AI is suspected of having caused damage, such evidence was already requested by the aggrieved party, that party has undertaken all reasonable measures to gather the evidence

of its own effort and has demonstrated to a sufficient degree that the defendant is likely to be in possession of the relevant evidence (Article 3(1) and Article 3(2)). Any evidence disclosed must be proportionate to the claim, and the court must preserve the confidentiality of proprietary information (Article 3(4)).

In case the defendant does not comply with the court order for disclosure of evidence, it may be presumed not to have complied with its duty of care imposed by the AIA (as per Article 3(5), Article 4(2) and Article 4(3) of the AI Liability Directive, as well as Article 10(2), Article 10(4), Article 13 to Article 16(a), Article 16(g), Article 21 and Article 29 of the AIA).

Demonstrated or presumed non-compliance with the duty of care in the form of the breach of AIA provisions enumerated above also results in a presumption of causation between the fault of the defendant and the damaging output (lack of output) by the AI system if this causality link is otherwise 'reasonably likely', and it has been demonstrated that the AI's output (lack of output) was the cause of the damage suffered (Article 4(1) of the AI Liability Directive). Thus, in a way relevant to our inquiry, in cases when the AI system was not developed and trained in a way that insures non-biased output, or that do not comply with the transparency or human oversight, requirements laid down in the AIA presumption of causation between these factors and the damaging output (lack of output) may be presumed.

We may observe here that due to the view enshrined in the AIA regarding transparency, the lack of sufficient reasons for an AI output may not be invoked by the claimant to benefit from the presumption of causation between the fault of the provider (or the user) and the AI output regulated by the AI Liability Directive, even if the right of the aggrieved party to receive such a reasoned decision constitutes a fundamental right. Also, in lack of a reasoned decision, as we have shown, no proper human oversight of the 'AI judge' may be achieved. Only this latter situation may be invoked as a basis for employing the presumption of causation laid down in the AI Liability Directive. Therefore, whereas other fundamental rights, such as non-discrimination, are better positioned to be protected by the AI Liability Directive, such protection is less clear in the case of the right to a reasoned decision.

5. Concluding Remarks

In our study, we have attempted to analyse the currently proposed regulatory framework in the European Union for the field of AI implementations in the light of the prospect of AI-aided or AI-generated adjudication, the so-called 'Ai judge'. We have found that preparatory documents, such as those drafted by the AI HLEG, have identified the risks posed by AI systems in a general manner and have not duly concentrated on ensuring a fair trial in case such systems would be employed for adjudication. Specifically, the basic right to a reasoned decision, something that as of

yet seems to exceed the abilities of AI systems, has not been adequately considered among the various exigences of AI transparency even if the lack of such a decision (even of an administrative, not just judicial nature) may make exercise of judicial remedies only an illusory possibility, also affecting human oversight of AI outputs. We have also observed that the regulatory proposals by the European Commission in the form of the AIA and the AI Liability Directive do not tend to provide solutions for the problem of a lack of sufficient reasoning. We, therefore, consider that along with future research in this field, regulation is also necessary that specifically deals with the implications of AI use during the administration of justice.

References

ALON-BARKAT, S.–BUSUIOC, M. 2022. Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice. *Journal of Public Administration Research and Theory* 2022/February: https://doi.org/10.1093/jopart/muac007.

CARLO, C. Di–DE BONDT, M.–EVGENIOU, T. 2021. AI Regulation Is Coming. *Harvard Business Review* 2021/September–October. https://hbr.org/2021/09/ai-regulation-is-coming.

CASOVAN, A.–SHANKAR, V. 2022. A Framework to Navigate the Emerging Regulatory Landscape for AI. *OECD. AI Policy Observatory*. https://oecd.ai/en/wonk/emerging-regulatory-landscape-ai.

CHESTERMAN, S. 2021. Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity. *The American Journal of Comparative Law* 69(2): 271–294. https://doi.org/10.1093/ajcl/avab012.

DEEKS, A. 2019. The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review* 119(7): 1829–1850.

EDWARDS, L.–VEALE, M. 2017. Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16(18): 18–84.

EUROPEAN COMMISSION. 2022. *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive) COM(2022) 496 Final*. https://ec.europa.eu/info/sites/default/files/1_1_197605_prop_dir_ai_en.pdf.

EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY. 2021. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 Final)*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206.

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE. 2019. *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*. Strasbourg: Council of Europe. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

FINK, M. 2021. The EU Artificial Intelligence Act and Access to Justice. *EU Law Live* 2021/May. https://eulawlive.com/op-ed-the-eu-artificial-intelligence-act-and-access-to-justice-by-melanie-fink/.

FRANCESCONI, E. 2022. The Winter, the Summer and the Summer Dream of Artificial Intelligence in Law. *Artificial Intelligence and Law* 30(2): 147–161. https://doi.org/10.1007/s10506-022-09309-8.

GERBER, D. J. 2002. Comparing Procedural Systems: Toward an Analytical Framework. In: *Law and Justice in a Multistate World: Essays in Honor of Arthur T. Von Mehren*. Ardsley.

INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

2020a. *Assessment List for Trustworthy Artificial Intelligence*. Brussels. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

2020b. *Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence*. Brussels. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

LUPO, G. 2022. The Ethics of Artificial Intelligence: An Analysis of Ethical Frameworks Disciplining AI in Justice and Other Contexts of Application. *Oñati Socio-Legal Series* 12(3): 614–653.

RAWLS, J. 1999. *A Theory of Justice. Revised Edition*. Cambridge (Massachusetts, USA).

SKITKA, L. J.–MOSIER, K. L.–BURDICK, M. 1999. Does Automation Bias Decision-Making? *International Journal of Human-Computer Studies* 51(5): 991–1006. https://doi.org/10.1006/ijhc.1999.0252.

SZEKELY, J. 2019. Lawyers and the Machine. Contemplating the Future of Litigation in the Age of AI. *Acta Universitatis Sapientiae, Legal Studies* 8(2): 231–244.

THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. 2019. *Building Trust in Human Centric Artificial Intelligence*, (COM(2019)168). https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.

VARONA, D.–LIZAMA-MUE, Y.–SUÁREZ, J. L. 2021. Machine Learning's Limitations in Avoiding Automation of Bias. *AI & Society* 36(1): 197–203. https://doi.org/10.1007/s00146-020-00996-y.

WISCHMEYER, T. 2020. Artificial Intelligence and Transparency: Opening the Black Box. In: *Regulating Artificial Intelligence*. Cham. 75–101. https://doi.org/10.1007/978-3-030-32361-5_4.

ZUIDERVEEN BORGESIUS, F. J. 2020. Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence. *The International Journal of Human Rights* 24(10): 1572–1593. https://doi.org/10.1080/13642987.2020.1743976.

*** Charter of Fundamental Rights of the European Union. 2012. Official Journal C 326 26.10.2012: 391–407.

*** European Convention on Human Rights. 1950. https://www.echr.coe.int/documents/convention_eng.pdf.

*** *Guide on Article 6 of the European Convention on Human Rights. Right to a Fair Trial (Civil Limb)*. 2022. Council of Europe – European Court of Human Rights. https://www.echr.coe.int/documents/guide_art_6_eng.pdf.