Privacy and Data Protection Constraints to Automated Decision-Making in the Judiciary

Marco Almada,[1] Maria Dymitruk[2]

Abstract: Competent authorities of many countries around the world are deploying automation tools in decision-making processes in the judiciary. The scope of application of automated judicial systems, often based on artificial intelligence (AI), is broad, ranging from the improvement and acceleration of organisational or office-based court tasks to the automation of substantive judicial decisions. The ongoing or future judicial automation in the European Union (EU) requires an in-depth legal analysis from the perspectives of data protection and privacy law, which raise crucial legal issues in the judicial automation landscape. The paper, therefore, aims to identify the constraints created by the EU data protection and privacy law to the use of automated decision-making within judicial proceedings. It first describes two models of judicial automation, which need to be distinguished on the basis of the degree of automation and the level of human involvement in the decision-making process (fully and partially automated decision-making and two sub-models of the latter). The paper then discusses the legal bases for both models of automated personal data processing in the judiciary, with a particular emphasis on the French national regulation, based on Article 22 of the GDPR, which provides the sharpest restriction to judicial automation within the European Union. Even when there are lawful grounds for judicial automation, legal requirements will provide constraints and limits to the operation of any automated decision-making systems. Acknowledging this fact, the paper concludes with the analysis of three perspectives which introduce limits to automation — data protection regulation, privacy law, and technological constraints —, as well as the safeguards that these viewpoints provide to the rights and interests of those affected by judicial automation.

Contents

[1] Lawgorithm and Law School, University of São Paulo. E-mail: marco.almada@usp.br
[2] Faculty of Law, Administration and Economics, University of Wroclaw. E-mail: maria.dymitruk@uwr.edu.pl

1. Introduction

Automated decision systems,[3] nowadays mostly based on artificial intelligence (AI), grow in importance in public decision-making. Presumably, they will be increasingly applied also in the judiciary. Relevant examples of initial steps taken within judicial systems can be found e.g. in Singapore Supreme Court[4] or in Brazilian *Supremo Tribunal Federal* as part of the VICTOR project.[5] The extent of court automation may take different forms, such as improving task-managing efficiency, assisting in adjudication, and unassisted conducting of substantial judicial activities. As such, the application of automated decision systems has the potential to significantly reshape the judiciary's organisational, institutional and procedural aspects.

The decision-support systems have already been adopted in the justice sectors of several countries. As an example may serve EXPERTIUS, a judicial decision-support system used in Mexico in the field of family law. This support system advises judges and clerks on the eligibility and the amount of pension to be granted to plaintiffs based on the existence of a "feeding obligation" (Cáceres 2008). The 206 System used by the Shanghai No. 2 Intermediate People's Court is another well-debated example of AI supporting human decision-making in the justice sector.[6] The system is an integrated AI assistive model for criminal cases, which, among others, aids judges in finding relevant facts and authenticating evidence.

Scientific experiences could be an inspiration to judicial automation, too. In order not to be groundless, CLAUDETTE project can be given as an illustration. CLAUDETTE ("automated CLAUse DETectEr") is an interdisciplinary research project hosted at the Law Department of the European University Institute in Florence, which aims to automate the reading and the legal assessment of online consumer contracts and privacy policies, to evaluate their compliance with the EU's contractual terms law and the personal data protection law (GDPR[7]), using machine

[3] The word "system" is used with different, but not altogether unrelated, meanings in computer science and law. In this paper, three of those senses are relevant: (i) *legal systems* as systems formed by laws, such as the laws of a country or the European Union; (ii) *judicial systems* as the organisational system of judicial authorities that interpret and apply the law in the name of the state; and (iii) *computational systems* such as those that implement artificial intelligence algorithms. As a rule, unqualified references to *"system" refer to a computational system, unless a different meaning is clear from the context.*

[4] https://www.statecourts.gov.sg/cws/Resources/Documents/ICCE%202016/2(i)%20Court%20of%20the%20Future%20LeeJ_2016070716114720.pdf#search=artificial%20intelligence [access: 11.02.2020].

[5] http://gpam.unb.br/victor/ [access: 11.02.2020].

[6] https://www.chinadaily.com.cn/a/201901/24/WS5c4959f9a3106c65c34e64ea.html [access: 11.02.2020].

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj [access: 11.02.2020].

learning and grammar-based approaches.[8] While this system has been developed as a tool for customer protection, it may as well be used by judges as support in verifying the compliance of terms of service and privacy policies of online platforms with the substantive law in place.

As the automation of non-substantive judicial activities (excluding adjudication) begins to significantly develop, some countries have already begun to look into fully automating some of the judicial decisions (the idea of a "robot judge"). For instance, the Ministry of Justice of Estonia, aiming to clear the case backlog, plans to introduce a "robot judge" that could adjudicate small claims disputes of less than €7,000.[9]

The aforementioned examples demonstrate that the use of automated systems has a chance to become a global trend in providing smart and timely justice. At the same time, it also runs the risk of "dehumanisation" of the judiciary (D'Amato 1977), lack of transparency, algorithmic biases or even violation of human rights (MSI-NET 2018; CEPEJ 2018). This paper seeks to identify and analyse the constraints created by EU data protection and privacy law to the use of automated decision-making within judicial proceedings. As the progressing court automation is likely to have a more profound impact on the administration of justice in the future, it demands thorough consideration of the legal grounds for automated judicial decision-making both in the data protection context and the privacy framework.

2. Automated decision-making in the Judiciary

Understanding the possibilities — and the risks — involved in judicial automation requires that we consider how automation comes into play at the Judiciary. Drawing from studies on the economics of artificial intelligence (Acemoglu and Restrepo 2018), one might describe the impact of automation in terms of displacing human labour in the performance of professional tasks. Some of the tasks that might be automated, such as the statistical analysis of cases similar to the case being judged[10] or the identification of the case law cited within a case,[11] are *auxiliary tasks*: even when those tasks are entirely automated, their output only becomes relevant to judicial proceedings when a human decides to use it for some purpose. For example, a lawyer could rely on statistical analyses before deciding whether or not to file a lawsuit, or a judge might rely on citation networks to ensure that their decision covers all relevant case law, but any effects are still produced as a consequence of a choice made by a human.

Another possibility would be to automate *core tasks* within the Judiciary, that is, those tasks which are directly relevant either to specific judicial proceedings, such as the evaluation of

[8] http://claudette.eui.eu/index.html [access: 11.02.2020].
[9] https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/ [access: 11.02.2020].
[10] Startups such as Predictice (https://predictice.com) and Case Law Analytics (https://www.caselawanalytics.com) provide legal professionals with a quantified assessment of the risks involved in filing a lawsuit.
[11] As an example, Ravel Law (https://www.ravellaw.com)'s Search Visualization tool generates citation graphs for case law.

whether a case should be admitted to trial,[12] or to the functioning of courts in general. Both kinds of tasks could, at least in principle, be subject to partial or total automation, but in either approach the outputs produced by automation could — either immediately or by shaping the decisions made by a human — have a direct impact on judicial outcomes, which is why specific safeguards have been adopted for those cases.

Many of the examples presented above involve technologies that are usually labelled as AI,[13] and, since AI systems are complex technological constructs, that complexity introduces technological and social challenges to automation. But not all forms of automation require this level of sophistication. Just like a thermostat can be used to maintain a stabilised temperature at a home without the need for intelligent behaviour (Brachman and Levesque 2004, 5), some of the more menial tasks within judicial proceedings, such as verifying whether all required documents are present when filing a lawsuit, can be performed through the simple application of *if-else* statements. Therefore, any discussion of the impact of judicial automation must not only consider the variety in use cases, but also the varying levels of complexity which may be covered by the umbrella of judicial automation.

2.1. Models of judicial automation

Based on the degree of automation and the level of human involvement in the decision-making process, two main models of judicial automation can be distinguished: fully automated decision-making and partially automated (semi-automated) decision-making. In the first model, an automated system has a role of a decision-maker and independently conducts tasks needed to reach a decision, including legal reasoning. In this model, the automated system is authorized to make a decision binding to parties to court proceedings (natural and legal persons), without any human authorisation.

In contrast to the – quite simple in the description – model of fully automated decision-making, the second model is much more complex. Semi-automated decision-making consists of two sub-models: one of them is based on the assumption that an automated system can serve as a decision aid for judges and court clerks; the second one assumes that an automated system may also serve as an adversary and challenge human decision-making. In both cases, the decision-making process is conducted by a human, supported by the automated system.

Automation as a decision aid. In this sub-model, the automated system is designed to help the judge's decision-making. This assistance consists of providing a human with a proposal for the final decision or deciding on the main factors that determine the final decision taken by a

[12] As an example, the aforementioned VICTOR project uses machine learning techniques to identify whether an appeal received by the Brazilian Supreme Court falls into the scope of a previous, binding decision of that court. At the moment, all decisions are reviewed by human specialists, which validate (or not) the initial algorithmic suggestion.
https://www.conjur.com.br/2019-mar-14/fux-mostra-beneficios-questionamentos-inteligencia-artificial

[13] Understood broadly as the use of computer systems to perform tasks that would require intellectual work if performed by a human.

judge[14]. This sub-model assumes that the automated system's suggestion is subsequently verified by a human judge. Acknowledging the correctness of an automatically-generated suggestion, the human judge will issue an identical decision. Denying to follow the automated system's advice, the human judge will reject and amend the decision. To ensure meaningful human participation in such a decision process, automated systems may present more than one suggestion, forcing the human decision-maker to perform a deliberate choice between the automatically produced options, deploying their knowledge and expertise (Almada 2019).

Automation as an adversary. In this sub-model of the semi-automated decision-making, the automated system offers its support only after and in response to an initial decision of a human judge. Here, the automated system is used not as an aid for generating a decision, but for a critique of that decision. The goal of the system operation is to confront human arguments and conclusions. It could reveal weaknesses of human decision-making, show potential solutions, point at omitted evidence, or present diverse legislation or case-law. Unlike the previous sub-model whereby it is the human who verifies the outcome of the automated decision-making, in this sub-model it is the automated system that verifies human decision-making. Considering the above, the adversarial sub-model has the potential to enhance the quality of judicial decisions and to reduce the risk of the rubber-stamping of automatically produced suggestions.

In each of the two sub-models, the intervention of the automated system may either be carried out only once — appearing at a single point of a judge's workflow — or repeated at every stage of the human decision-making process (before or after the determination of the legal basis of the decision, the determination of the factual state of a case, and the final decision). And, since those approaches are not mutually excluding, mixed models that incorporate elements of both sub-models are also possible (e.g. initial suggestions, posterior confrontations).

2.2. Legal bases for automated data processing in the judiciary

Judicial automation involves personal data processing to the extent that it is applied to situations concerning natural persons.[15] As such, it generates various implications related to data

[14] To illustrate the second type of assistance, it would be beneficial to mention one particular example – COMPAS system (an acronym for Correctional Offender Management Profiling for Alternative Sanctions) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist. A number of U.S. states, including Wisconsin, Florida and Michigan, use COMPAS to assist judges with sentencing decisions (Kehl *et al.* 2017). Although the result of the system's operation (the recidivism risk assessment) is only one of several factors that must be taken into account while sentencing, its profound significance makes it one of the essential aspects of the determination of the final decision. Therefore, it can be concluded that even though the system does not provide a human judge with a precise proposal for the decision, it still determines one of the main factors in the case, fulfilling the definition of the discussed sub-model of semi-automated decision-making.

[15] In Subsection 2.2 only automated personal data processing is discussed. When the automated judicial decision-making does not involve personal data (within the meaning of Article 4 of the GDPR), some legal and technological constraints still exist (these issues are discussed below).

protection. The legal basis for the automated judicial decision-making in force is provided by the GDPR[16], and the relevance of its adoption is increasingly recognized also in the judicial sector.[17]

In accordance with Recital 20 of the GDPR, the regulation applies to the activities of courts and other judicial authorities; European Union or Member State law could further specify the operations and procedures in relation to the processing of personal data by courts and judicial authorities. As such, the EU and Member States' legal frameworks for data protection are likely to have an impact on automated decision systems. Article 22 of the GDPR is of particular importance. It gives a person a qualified right[18] "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", except in the cases listed in Article 22(2) of the GDPR. It is hard to imagine a judicial decision-making that would not produce such legal effects as an essence of the court trial is to bindingly determine rights and obligations of the parties to the proceedings. Hence, it could be assumed that every act of automated judicial processing of personal data has to have a legal base in Article 22.

Among lawful bases for processing automated decision-making defined in Article 22(2) of the GDPR, potentially relevant for automated judicial decision-making are the following legal grounds[19]:

1. Explicit consent of the data subject - Article 22(2)(c) of the GDPR,
2. Union or Member State law that permits such automation as long as suitable safeguards for the data subject's rights, freedoms and legitimate interests are present – Article 22(2)(b) of the GDPR.

Explicit consent. Explicit consent is one of the exceptions from the prohibition on automated decision-making and profiling defined in Article 22(1) of the GDPR, but it does not seem to be an appropriate legal basis in the context of the use of automated systems in the judiciary.

[16] Provisions of the GDPR repealed Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (hereinafter "DPD", available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A31995L0046 [access: 11.02.2020]). Article 15 of the DPD provided a similar right to that established in Article 22 of the GDPR (see Bygrave 2019 for the comparison between the provisions of the GDPR and the DPD in the context of automated decision-making). The main difference between these two provisions is the scope of application. Decisions caught by Article 15 of the DPD had to be based on profiling, whereas Article 22 of the GDPR applies to any kind of automated processing, including but not limited to profiling (Bygrave 2019).

[17] See e.g. the INFORM project ("INtroduction of the data protection reFORM to the judicial system") - http://informproject.eu/ [access: 12.02.2020].

[18] The term "right" in the provision does not mean that Article 22(1) applies only when actively invoked by the data subject. Article 22(1) establishes a general prohibition for decision-making based solely on automated processing. This prohibition applies whether or not the data subject takes an action regarding the processing of their personal data (Article 29 Working Party 2018).

[19] The third ground listed in Article 22(2) of the GDPR ('necessary for the entering into or the performance of a contract between the data subject and the data controller') is completely irrelevant in the context of judicial use of automated systems.

According to Article 4(11) of the GDPR, the consent means a "freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she (…) signifies agreement to the processing of personal data relating to him or her". Participation in judicial proceedings is not always voluntary, thus the reliance on the "freely given" consent could be troublesome. Moreover, the GDPR itself indicates that the "consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation"[20]. Taking into account the superiority of the justice system over the parties to the court proceedings[21], the consent should not, as a general rule, constitute the legal basis for the automated judicial decision-making. The consent given by the data subject would be an inappropriate legal ground for practical reasons, too. Pursuant to Article 7(3) of the GDPR, the data subject has the right to withdraw his or her consent at any time. Such withdrawal in the course of proceedings would prevent further automated processing and cause detriment to the speed and efficiency in delivering justice. Accordingly, it can be concluded that the Article 22(2)(b) of the GDPR, i.e. Union or Member State law, is the one providing appropriate legal ground for automated judicial decision-making (Cobbe 2019).

EU or Member State Law. The second (and – as mentioned above – the only suitable) legal basis for automated judicial decision-making is an authorisation given by law which must provide suitable safeguards for the data subject's rights, freedoms and legitimate interests. Safeguards in judicial automation will be discussed in Section 4 below. However, it is essential to note that this legal ground is further limited by the GDPR with regard to "special category data".[22] If a judicial decision is based on such data, automated decision-making is generally prohibited, with two exceptions: explicit consent (Article 9(2)(a) of the GDPR) or substantial public interest (Article 9(2)(g) of the GDPR).[23] As already explained, consent cannot be considered as an appropriate basis for judicial automation, thus only the latter can apply.

Interestingly, an essential legal ground for judicial automation of special category data (i.e. Article 9(2)(f) of the GDPR) is not mentioned by Article 22(4) of the GDPR that enables automated decision-making. Typically, when courts are acting in their judicial capacity, the derogation from the general prohibition on processing the special categories of personal data applies (Article 9(2)(f) of the GDPR), so courts are entitled to process such data. In case of automated decision-making, because of the wording of Article 22(4) of the GDPR, it seems that

[20] Recital 43 of the GDPR.

[21] Another difficulty is that the data used in judicial proceedings can concern other people than the party to the proceedings, so the party would not be entitled to give such consent within this scope.

[22] "Special category data" is personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, or the processing of genetic data, biometric data for the purposes of uniquely identifying an individual, data concerning health, or data concerning an individual's sex life or sexual orientation – Article 9(1) of the GDPR.

[23] Decisions referred to in paragraph 2 shall not be based on the special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place – Article 22(4) of the GDPR.

there is no designated legal ground for the automated processing of the special category data in judicial proceedings. This means that every act of automated processing of the special category data by courts must be permitted by the law and grounded in substantial public interest referred to in Article 9(2)(g) of the GDPR. From the comparison between Article 22(4) and Article 9(2)(f) of the GDPR can also be concluded that the full automation of the judicial proceedings (involving the special category data) would be an exception rather than a general rule, as "acting in judicial capacity" is not mentioned by Article 22(4) of the GDPR. On the other hand, as long as no special category data is processed, proving substantial public interest is not required. Cobbe (Cobbe 2019) rightly pointed out that before a public body (e.g. court) starts to process personal data in accordance with Article 22 of the GDPR, it has to determine whether there exist other effective and less intrusive methods of achieving the same result (i.e. whether it is necessary to employ the automated decision-making). In order to automate judicial decision-making, the court will need to demonstrate that there were no alternative or more privacy-preserving means.

The practical effect of Article 22 of the GDPR also depends on the legislation of individual Member States. Because of the wording of Article 22(2)(b) of the GDPR, automated decision-making in the judiciary will be mostly shaped by the national legislation. As Member States have been given relatively broad latitude in this regard, it can be assumed that significant differences will emerge between national regulatory frameworks for automated decision-making (Bygrave 2019).[24] Malgieri (Malgieri 2019) analysed all existing Member States' laws implementing the GDPR, and identified four approaches with regard to the automated decision-making: a negative approach, a neutral approach, a procedural approach and a proactive approach. Most Member States adopted a negative approach, which means that they do not address the issue of automated decision-making in their national data protection laws, so they do not provide any specific case of permitted automated decision-making under Article 22(2)(b) of the GDPR. Only two Member States (France and Hungary) adopted a proactive approach, not only implementing Article 22(2)(b) of the GDPR but also proposing new and more specific safeguards than those indicated by that Article.

Among all Member State laws, only the French law[25] precisely addresses the automated decision-making in the judiciary,[26] and states that "a court decision involving an assessment of a person's behaviour cannot be based on the automated processing of personal data if such processing is intended to evaluate aspects of that person's personality".[27] Malgieri regarded this

[24] Bygrave (2019) also indicates that this will undermine the harmonisation aims of the GDPR.

[25] Act on data processing, files, and individual liberties (*Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*).

[26] It is worth adding that it is not recently passed law. The 1978 Act on data processing, files, and individual liberties prohibited judicial, administrative, or personal decisions involving assessment of human behaviour insofar as these were based solely on automatic data processing which defined the profile or personality of the individual concerned was an inspiration for the EU legislator when determining Article 15 of the DPD (Bygrave 2019).

[27] French version: 'Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne'

provision as a total prohibition of semi- and fully- automated decision-making if automated processing is intended to evaluate aspects of personality (Malgieri 2019). However, as long as the automated decision-making does not involve evaluation of the data subject's personality, the prohibition shall not apply.[28]

Among all Member States' laws, the French regulation provides the sharpest restriction to judicial automation. While it might be instructive, a blanket ban on even semi-automated processing is not a logical consequence of the GDPR protective system, but instead reflects the fact that the French judicial system highly values secrecy and the presentation of a unified institutional voice (Langford and Madsen 2019). Furthermore, even this strict standard still leaves some space for automated decision-making, as long as the relevant automations are not intended for evaluating aspects of their target's personality. It should however be noted that the phrase "intended to evaluate aspects of that person's personality" can be interpreted in different ways: narrowly and widely. When interpreted narrowly, the evaluation of the personality takes place only in selected judicial proceedings (primarily in criminal proceedings that aim to determine and assess the perpetrator's personality and behaviour). Under the wide interpretation, the prohibition would have the broad scope of application, as most of the judicial proceedings are associated with natural persons, their behaviour and relations between them, which can be connected with the assessment of someone's personality. But as the data processing must under the French law "intend to evaluate" (and not just "involve the evaluation"), the strict interpretation seems to be more appropriate.

3. Limits to judicial automation

3.1. Data protection in judicial automation

Article 22 of the GDPR prohibits the decision-making based solely on automated processing of personal data, while at the same time laying down the exceptions from this general prohibition. From the above it can be concluded that when a judicial decision is not based solely on automated processing, the use of automated systems is allowed by the GDPR. It does not mean, however, that in the case of non-solely automated decision-making the provisions of the GDPR shall not apply, on the contrary. Non-solely automated decision-making is permissible as

https://www.legifrance.gouv.fr/affichTexteArticle.do?idArticle=LEGIARTI000037817723&cidTexte=JORFTEXT000000886460&dateTexte=20190601 (access: 16.02.2020)

[28] In this context, it is also worth recalling the French prohibition on information processing destined for evaluating, analysing, comparing or predicting professional practices of magistrates and members of the judiciary introduced by Article 33 of the Justice Reform Act (*LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*, available at: https://www.legifrance.gouv.fr/eli/loi/2019/3/23/2019-222/jo/article_33). This recent French regulation is aimed at putting an end to judicial analytics as long as the data permitting the identification of judges are involved. The ban on predictive analytics tools identifying trends in judges' behaviour in relation to court decisions is not directly relevant to judicial automation, but it reflects a general attitude towards the use of information technologies in the judicial sector, which is different from but related to the prohibition of automated decision-making in the French judiciary.

long as all general requirements set out in the GDPR are met. This is because every automated decision-making involving personal data is subject to the rules of the GDPR.[29] It means that even if automated decision-making is not prohibited by Article 22(1) of the GDPR because the decision is not solely based on automated processing, all data protection principles (most notably those set out in Article 5 and Article 6 of the GDPR) need to be respected, i.e. in particular the judicial automated processing must be:

1. carried out lawfully, fairly and in a transparent manner in relation to the data subject (Article 5(1)(a) of the GDPR),
2. necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (Article 6(1)(e) of the GDPR),

When the special categories of personal data are involved, the automated processing is permissible if courts are acting in their judicial capacity (Article 9(2)(f) of the GDPR) or any other condition established in Article 9(2) of the GDPR is met.

Given the two models of judicial automation discussed in Subsection 2.1 (fully and partially automated decision-making), it is necessary to determine which model is prohibited in accordance with Article 22(1) of the GDPR, and therefore needs a special legal ground in Union or Member State law permitting the automation.

At first sight, it seems that the decision-making that is *based solely on automated processing* equates with the model of fully automated decision-making that involves no human intervention. However, Article 29 of the Data Protection Working Party's Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679 rightly emphasises the fact that the data controller cannot avoid provisions of Article 22 by fabricating human involvement. As a result, Article 22(1) of the GDPR must be interpreted extensively. Decisions based solely on automated processing, therefore, should be understood as any decisions in which the human involvement is not meaningful[30]. The "meaningful human intervention" condition is met when the controller ensures that decision[31] is overseen by someone who has the authority and competence to change it, after considering all the relevant data (Article 29 Working Party 2018). But even the already mentioned example of the automated decision-aiding systems presenting more than one possible solution, therefore forcing human decision-maker to perform a deliberate choice, may not provide sufficient opportunities for

[29] This conclusion can be drawn both from the general analysis of provisions of the GDPR and Recital 72 of the GDPR.

[30] WP29 gave an example of someone who routinely applies automatically generated profiles to individuals without any actual influence on the result (Article 29 Working Party 2018).

[31] The human oversight may be achieved through governance mechanisms such as (a) the human-in-the-loop approach, HITL (the human intervention occurs in every decision cycle of the system), (b) the human-on-the-loop approach, HOTL (the human intervention is present during the design cycle of the system and monitoring the system's operation), or (c) the human-in-command approach, HIC (the human oversees the overall activity of the AI system, including its broader economic, societal, legal and ethical impact, and decides when and how to use the system in any particular situation) (High-Level Expert Group on Artificial Intelligence 2019).

meaningful human involvement. If a human can choose freely between several possible scenarios, and all of those solutions are automatically-generated, the human decision space is still severely constrained as decision-maker would still lack control over the content of the outcomes (Almada 2019).

An insufficient human control creates the risk of "rubber-stamping" of the automatically-generated decisions (referred to as a "token gesture" by WP29).[32] Considering common problems of many judicial systems such as deficient human resources or a caseload pressure, the danger of "rubber-stamping" of algorithmically prepared decisions is high (MSI-NET 2018). The psychological research on the lawyers' trust level towards artificial intelligence systems lends credence to the alarming hypothesis about the human tendency to overtrust the automated decision-making (Dijkstra *et al*. 1998, Dijkstra 1999, Dijkstra 2001).[33]

Recognizing the key role of human intervention, it should be concluded that when the decision is taken without (or with insufficiently rigorous) human control, the decision-making is, in fact, solely automated. Thus, even when a decision-aiding system is used, the prohibition established by Article 22(1) of the GDPR might still apply. This circumstance should be taken into account both in the design stage of the construction of every judicial automated system and organizational schemes in courts, as well as in the data protection impact assessment (DPIA) carried out by the data controller under Article 35 of the GDPR (Article 29 Working Party 2018).[34]

In conclusion, the general prohibition set out in Article 22(1) of the GDPR applies both to the model of fully automated decision-making and to the model of semi-automated decision-making when the automated system aims only at helping the human decision-maker, but there are no (or insufficient) safeguards against the situation where he or she undertakes a cursory analysis or simply follows the automatically-generated suggestion without further consideration (Cobbe 2019). Thus, while it may seem logical to draw a distinction between fully automated decision-making and semi-automated decision-making, in practice the boundaries between these two may be blurred (MSI-NET 2018). On the other hand, the second sub-model of partially automated decision-making, i.e. the adversarial sub-model, eliminates the danger of the overreliance on the decision-aiding system. This sub-model partially reduces the risk that an inappropriately designed human intervention would compromise the quality of the final judicial decision (instead of improving it). Interestingly, this sub-model reverses the natural order of

[32] "To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture" (Article 29 Working Party 2018, 21).

[33] Ignoring this fact would result in 'quasi-automated decision-making' in the judiciary in which the role of the judge would be limited to indiscriminate following the system's suggestions (Dymitruk 2019a).

[34] According to this provision, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data, especially when using new technologies (Article 35(1) of the GDPR). Such assessment is required in particular when the automatically produced decision produces legal effects concerning the natural person or similarly significantly affects him or her (Article 33(3)(a) of the GDPR).

things, letting the automated system intervene in the human decision-making (instead of allowing the human intervening in the automated decision-making).

It is important to bear in mind, though, that the controller not only must ensure a lawful basis for the automated processing, but also properly safeguard the data subject's rights, freedoms and legitimate interests. The suitable measures will be discussed in detail in Section 4.1 below. The safeguards against an inappropriate automated decision-making aim to ensure fair processing and to prevent any undue impact on the data subjects. They are mentioned in Article 22 of the GDPR twice: once in Article 22(2)(b), and again in Article 22(3). The first provision mentions the "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests" in general, whereas the latter details three basic measures: the right to human intervention, the data subject's right to express his/her point of view, and the right to contest the automated decision[35]. Thus, when a European Union or a Member State law authorise the solely automated decision-making (in the case referred to in Article 22(2)(b) of the GDPR), the selection of suitable measures is a responsibility of the European or a national legislator respectively, as the GDPR provisions do not impose specific conditions for such automated decision-making. However, when the legal basis for the automated processing is necessity for the performance of or entering into a contract (Article 22(2)(a) of the GDPR) or the explicit consent of the data subject (Article 22(2)(c) of the GDPR), the safeguards must include at least these three key rights.[36] The list of safeguards laid down by Article 22(3) of the GDPR is, of course, the bare minimum required[37], but it can become a good starting point for further construction of "suitable measures". The Slovenian human rights impact assessment on automated decision-making is a good example of such amplification of the safeguards requirements (Malgieri 2019). Even though the GDPR stipulates that the data protection impact assessment (DPIA) is mandatory in the case of the automated decision-making (Article 35(3)(a) of the GDPR), the Slovenian example additionally introduces an innovative focus on the impact on human rights and fundamental freedoms, in particular with regard to non-discrimination. This additional feature makes the assessment compliant with the Slovenian national legislation (Malgieri 2019). The provisions of the GDPR do not introduce such extensive constraints.

The GDPR sets other basic requirements in Recital 71, indicating that in order to ensure fair and transparent automated processing, the controller is obliged to "use appropriate mathematical or statistical procedures for the automated decision-making; implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised; secure personal

[35] The above list is not exhaustive. Article 22(4) of the GDPR indicates that "the data controller shall implement (...) *at least* the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

[36] Article 22(3) of the GDPR stipulates that "(i)n the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

[37] It is worth noting that Article 22(3) of the GDPR offers in this regard a higher level of protection than under the DPD (Bygrave 2019).

data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect".[38]

Considering the potential impact of the automated decision-making on the data subjects' rights, courts and other judicial authorities, as controllers, should be particularly mindful of the obligation to provide necessary information to ensure fair and transparent processing.[39] When the automated processing of personal data is involved, the controller must: (a) inform the data subject about the existence of automated decision-making, (b) provide a "meaningful information about the logic involved" and give the explanation of the significance and the envisaged consequences of the automated processing for the data subject.[40] All required information should be specific and easily accessible.[41] WP29 also stresses that even if the automated decision-making does not meet the Article 22(1) definition (i.e. the decision-making is not based *solely* on automated processing), it is, nevertheless, good practice to provide the information listed above (Article 29 Working Party 2018).

It is also worth pointing out that the resolution on automated decision-making processes,[42] recently adopted by the European Parliament, underlined that the automated decision-making presents new challenges in light of the varied nature and complexity of artificial intelligence[43], and stressed the need for a risk-based approach.[44] The European Parliament sees the potential of the automated data processing to deliver innovative and improved services, but at the same time notes that people should be properly informed about how the automated systems function, about how to reach a human with decision-making powers, and about how the system's decisions can be checked and corrected.[45] The resolution also emphasises the need for transparency in data governance, noting the imperative of protecting personal data under the GDPR and the importance of using only high-quality and unbiased data sets in order to improve the output of algorithmic systems and boost human trust and acceptance.[46] The resolution does not focus on the public (or judicial) automated decision-making[47], but it rightly highlights the human-in-the-loop requirement, especially in the context of legal services, recalling the

[38] Recital 71 of the GDPR.
[39] Article 13(2) of the GDPR.
[40] Article 13(2)(f) and Article 14(2)(g) of the GDPR.
[41] The principle of transparency, as set out in the Recital 39 of the GDPR, requires that any information and communication relating to the processing of personal data must be easily accessible and easy to understand, and that clear and plain language be used. See also Article 12(1) of the GDPR.
[42] European Parliament resolution of 12 February 2020 on automated decision-making processes: ensuring consumer protection and free movement of goods and services (2019/2915(RSP)), available at: http://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.html#def_1_18 (access: 18.02.2020).
[43] Point 6 of the Resolution.
[44] Point 7 of the Resolution.
[45] Point 1 of the Resolution.
[46] Point 12 of the Resolution.
[47] It indicates specifically only the automated decision-making systems used in alternative dispute resolution mechanisms on digital platforms (Point 5 of the Resolution).

importance of supervision or independent oversight by qualified professionals in cases of automated decision-making where legitimate public interests are at stake.[48]

3.2. Judicial automation and privacy

Concerns with privacy are frequently mentioned as part of the justification for data protection rights.[49] Even so, the latter should not be conflated with the former, as the case laws of the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECtHR) recognise some differences between both concepts (Kokott & Sobotta 2013, 228). Since privacy is itself a fundamental right within both European human rights systems, the lawful application of automated decision-making to judicial activities cannot restrict itself to compliance with data protection law, but must instead consider other forms in which automation might harm the right to privacy.

Privacy can be understood as an affordance, that is, as a set of possibilities afforded to a person by an environment (Hildebrandt 2019, 84). From that perspective, judicial automation would impact privacy primarily by altering the technological and institutional environments in which judicial decisions happen, which might either expand or constrain the possibilities available to that person with regards to their private and family life.[50] Those rights, in turn, can be grouped into two clusters: *informational* and *decisional* privacy, with the former referring to the interest of individuals in controlling access to data about themselves (van den Hoven *et al.* 2019, 4).

In a society where information and communication technologies (ICT) are involved in various aspects of life, control over information is directly connected to data flows, especially those that happen over digital media. Data protection law thus plays a substantial role in promoting privacy, but European case law retains some key differences between the concepts. A substantive difference is that not all personal data refer to aspects of an individual's personal

[48] Point 10 of the Resolution 10 states that the European Parliament "(u)nderlines that while automated decision-making processes can improve the efficiency and accuracy of services, humans must always be ultimately responsible for, and able to overrule, decisions that are taken in the context of professional services such as the medical, legal and accounting professions, and for the banking sector; recalls the importance of supervision or independent oversight by qualified professionals in cases of automated decision-making where legitimate public interests are at stake".

[49] Two examples can be used to highlight this connection. Within the European Union data protection legislation, GDPR Recital 4 mentions "the respect for private and family life" as one of the fundamental rights warranting particular observation within data protection law. In a similar approach, the Council of Europe's Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Convention 108+) states in Article 1 that the protection of individuals with regard to their personal data is meant to contribute "(...) to respect for his or her human rights and fundamental freedoms, and in particular the right to privacy.".

[50] The European Convention on Human Rights (ECHR, article 8) and the Charter of Fundamental Rights of the European Union (CFR, article 7) both establish that "Everyone has the right to respect for his or her private and family life, home and communications."

life,[51] which means that a processing operation might fall afoul of data processing law without violating the data subject's privacy. However, that does not mean that privacy law is a proper subset of data protection law, as the ECtHR has recognised that legal persons can rely on privacy rights,[52] while data protection law explicitly restricts itself to natural persons (CFR, Article 8; GDPR Article 4(1)). Furthermore, data processing operations that fall outside the scope of data protection law, such as those undertaken "by competent authorities for the purposes of the prevention, the investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security" (GDPR Article 2(2)), might still intrude in a data subject's private or family life, being thus subject to privacy law.

The differences in scope between data protection law and privacy law become more salient when one considers that the latter does not restrict itself to informational considerations. The European fundamental rights system[53] interprets "private life" in a broad sense (Kokott & Sobotta 2013, 223), which encompasses not just the disclosure actions that are the object of traditional privacy law (Pałka 2020, 10), but also matters such as the provision of non-personal information needed for health risk assessment,[54] industrial pollution,[55] and access to legal abortion.[56] Despite the variety of subject matters covered by such decisions, they share one key aspect: they are all concerned with a person's right to make decisions about their private lives without undue interference, a right that has been called *decisional privacy*.

Decisional privacy, as manifested in the use of ECHR Article 8 for contexts beyond the control of personal information, has been justified by courts in terms of preserving the personal autonomy of individuals, understood as their right to develop and fulfil their identities (van der Sloot 2017, 192), which is why it is also called *constitutional privacy* (van den Hoven *et al.* 2019, 4). This movement has met some pushback, such as claims that decisional privacy would diminish informational privacy by overextending the concept (Wacks 2015, xiv), or that

[51] As an example, the ECtHR has ruled that information about criminal convictions is not immediately part of a person's private life, only becoming so after some time has passed and the matter has receded from public attention (Kokott and Sobotta 2013, 224).

[52] See, *e. g.*, *Bernh Larsen Holding AS and Others v. Norway*, application 24117/08, paragraph 159.

[53] While the examples were extracted from ECtHR decisions, CJEU case-law explicitly equates the meaning and scope of CFR Article 7 to ECHR Article 8(1), as interpreted through ECtHR case law (see Case C‑450/06 *Varec* [2008], paragraph 48, and Case C‑400/10 PPU [2010], paragraph 53).

[54] In *Vilnes and Others v. Norway* (application no. 52806/09) the ECtHR ruled that Norway had violated Article 8 of the ECHR by failing to provide deep-sea divers with information needed about the risks they incurred from their employer's use of rapid decompression tables, therefore undermining the divers' rights to make informed decisions about their health.

[55] The ECtHR decided, in *Fadeyeva v. Russia* (application no. 55723/00), that Russia had violated ECHR Article 8 by authorising the opening of a polluting enterprise in the middle of a densely populated town without adopting effective measures to either reduce pollution to acceptable levels or help inhabitants of the heavily toxic area around the enterprise to relocate, as had been required by domestic law.

[56] Bart van der Sloot (2017, 196) provides an overview of various ECtHR cases in which states were found to be in violation of ECHR Article 8 by failing to provide information and/or procedural frameworks required for the exercise of the right to abortion within the requirements set by domestic laws.

constitutional privacy is better understood as a right to liberty.[57] Still, even if one prefers to reserve the "privacy" label to the matters of informational privacy described above,[58] the matters currently described as relating to decisional privacy still are protected by existing case law, thus providing constraints that should be considered when analysing the possibility of judicial automation.

In the case of automated judicial decision-making, both aspects of privacy are at stake. Informational privacy can be put at risk by violations of data protection law, as described in Subsection 3.1, but privacy's broader scope means that undue intrusions into the privacy of legal persons,[59] or informational disclosures similar to those described above but justified by the exceptions carved into data protection law — such as the judicial use of data generated through automated decisions in criminal prosecution within the constraints of the Police Directive[60] — might still lead to legally actionable privacy violations.

When it comes to decisional privacy, interference may happen when a person has no mechanisms for seeking the reversal of unfavourable automated decisions.[61] Curiously, this problem is more salient in situations where automation plays a subsidiary role. Whenever a human judge is mistaken or makes an error, their results, at least in principle, be corrected through judicial channels, such as motions. If, however, that mistake is hardcoded into a computational system, addressing it might not be a straightforward process, and might even require changes that will demand substantial software engineering work. As described above, GDPR Article 22(3) provides a bundle of procedural rights — the rights to express one's point of view, to request human intervention and to contest automated decisions — that are applicable against fully-automated decisions, but those find no direct analogues in scenarios of partial automation or non-personal data, which is why Subsection 4.2 of this paper concerns itself with the safeguards against (undesirable effects of) automation that can be found in privacy law.

Another source of potential violations of decisional privacy comes from the information used in the decision-making processes. Judicial automation, in its various technological approaches, will

[57] Judith DeCew (2018, 13) ascribes this position to William Parent and Judith Thomson.

[58] Among other proposals, Raymond Wacks (2015, 34) claims that decisional privacy would be better understood as lying outside the central instances of privacy, being instead privacy-adjacent matters, while Przemysław Pałka (2020, 41) proposes a "data management" framework for treating individual and social interests affected by data processing, of which privacy would be an informational component.

[59] Current ECtHR case law (see *Bernh Larsen v. Norway*, *supra*, paragraph 159) understands that the standards that must be met by acceptable interferences with the privacy of legal persons are less strict than the ones applicable to the private life of natural persons.

[60] *Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA*. That would exclude, for example, "[p]rofiling that results in discrimination against natural persons on the basis of special categories of personal data" (Article 11.3).

[61] That *recourse* may happen either as the sort of formal appeal usually associated with judicial proceedings, or, in a narrower sense, as behavioural adjustment by the interested person (Venkatasubramanian and Alfano 2020, 288) intended to result in changes to machine behaviour.

always rely on limited data, thanks to a multitude of factors, such as the procedural and temporal constraints to obtaining information about a case,[62] technical and conceptual difficulties in identifying relevant case law, and even the fact that some aspects of human personality might be incomputable in themselves.[63] Faced with those challenges, automation software designers will need to adopt heuristics for addressing the general computational problems, but those heuristics, as generalisations, will necessarily fail to account for the particular situations of all persons involved in a case. As a result, the ensuing automated decisions will be based on coarse-grained pictures of judicial situations which might intrude into a person's private life even if the decision-making process is thought to be generally fair.

The problems ensuing from decisional heuristics highlight how the decisional and informational aspects of privacy are closely connected in the context of automated decision-making, requiring a comprehensive approach.[64] After all, the use of information with potential judicial relevance — contained in petitions, case law, news sites, and other legal or non-legal sources — can lead to interferences in the private life of the parties to the proceedings, of the judges, lawyers, and civil servants involved in a case, or even of uninvolved third parties, which might face the disclosure of information about their private lives or even have their private choices interfered with through automation.[65] Privacy law thus offers substantial constraints to the use of automated decision-making systems in the Judiciary.

3.3. Technological constraints to judicial automation

The activities of the Judiciary branch of any modern state involve such a variety of tasks that automation might still be gainfully employed by judges and other legal actors,[66] even if legal standards restrict some relevant possibilities for total or partial automation. For all that, judicial automation that is possible in theory might not be feasible in practice, either because the task itself cannot be performed by a computer or because the standards that would be required for legal performance cannot be met by computer systems, a failure that might result from intrinsic

[62] Such as the right to have one's case heard within a reasonable time, recognized by ECHR Article 6(1) and the principle of effectiveness (Dymitruk 2019a, 36).

[63] For the latter constraint, see Mireille Hildebrandt (2019, 87–96)'s argument that privacy concerns itself with the fundamental indeterminacy of human beings, which means that at least part of human existence does not lend itself to computational representation.

[64] Modern AI techniques have potentialized this connection, but it is not a novel concern: as Pałka (2020, 10) points out, the 1970s privacy literature argued for broadening privacy protections to cover not just the disclosure, but also the *use* of identifiable information. This movement can be associated with the decisional aspect of privacy, but it highlights the connection between decisions, automated or not, and the information that might be used to ground them, anticipating the *data-driven* hype of the 2010s.

[65] As an example, data mining in case documents might uncover information about people that are not parties to the proceedings, even if they are anonymised (Rocher et al. 2019).

[66] Either as simple automation — such as scheduling the upload of relevant documents — or in more complex scenarios to the extent permitted by law.

limits to computing or from capabilities that, despite being possible, are still beyond the current technological horizon.[67]

In computer science, there are many famous examples of non computable problems, which a computer cannot solve in principle (Immerman 2018, 1). Before there even were digital computers, Alan Turing proved that it is impossible to build a program that, for every input and every program, tells whether the evaluated program will finish running and provide an answer for the evaluated output (Kroll *et al.* 2017, 652). From those incomputable results, it follows that not all tasks can be solved by a computer,[68] which leads to the problem of identifying what techniques cannot be solved through algorithmic means.

Nevertheless, problems that are formally shown not to be computable are addressed every day. For example, even if it is impossible to express certain aspects of human personality through computable procedures,[69] bureaucracies and market actors alike increasingly rely on data procedures for categorising individuals (Fourcade and Healy 2017). This is possible due to the use of heuristic approaches, which rely on existing knowledge about a specific problem (Russell and Norvig 2010, 92) as rules of thumb for solving problems, especially those which lack a suitable, exact algorithm.

Those rules of thumb might fail to achieve their goals. Even if they are successful, they might produce undesirable side-effects — in the case of judicial automation, an example would be a decision algorithm that fails to properly consider the interests of third parties — or waste limited resources, such as time. Therefore, it is important to evaluate any heuristics adopted as part of a computer system. Since many of the justifications for adopting a system are described in quantitative terms — the volume of cases that a court must handle, the time spent on each case, and so on —, evaluation is usually performed in terms of metrics,[70] which can be extended to cover fairness aspects (e. g. Kulynych *et al.* 2020). This evaluation of metrics should be accompanied by quantitative evaluations of a system, intended to capture aspects of its usage that cannot in principle or in practice be adequately metrified.[71]

One obstacle to quantitative and qualitative evaluations of algorithmic performance is that complex automation systems are *opaque* to observers. Jenna Burrell (2016, 3–5) describes

[67] This discussion focuses on limits to algorithms, but it is crucial to keep in mind that real-world applications might also face other obstacles, such as the lack of the computational resources needed to run an algorithm that would otherwise be feasible, or information security problems, such as the existence of backdoors in the software or the hardware used for automating a given task.

[68] As Kroll *et al.* (2017, 652) highlight, this includes the task of testing software for incorrect outputs and other problems. No testing approach will cover all cases, and even a combination of tests cannot claim perfect coverage.

[69] Hildebrandt (2019) argues, based on the relationship between privacy and the "foundational indeterminacy of human identity" (p. 88), that human identity is incomputable, and that privacy should afford protection to these incomputable aspects of human life.

[70] Sommerville (2011, 668) distinguishes between *control* metrics, used to manage the software development process, and *predictor* (or *product*) metrics, that refer to the software itself.

[71] Even when exact algorithms are used, this sort of evaluation will be useful for understanding whether the system is solving the *right* problem.

three sources of opacity in algorithmic context: (i) intentional secrecy by the state or private actors that control algorithms, through mechanisms such as intellectual property laws; (ii) the specialised skills that are required for reading and writing code and mathematical models such as those used in AI-based automation; and (iii) the scale of the computational operations in contexts involving large data sets (*big data*), such as those that would be faced by most judicial automation systems. The dynamics between those factors can complicate the selection of what metrics and qualitative evaluations are relevant for understanding how a system functions, as well as the performance of the selected evaluations.

Since public trust in judicial institutions is closely related to procedural transparency (Dymitruk 2019a, 39), any solutions for judicial automation will need to address the issue of opacity.[72] The field of research known as Explainable AI (xAI) seeks to reduce the barriers for interpreting and explaining the outputs of artificial intelligence systems (Ehsan & Riedl 2020). However, the question of what is an explanation of an automated system remains open, and different answers to that question will require different technical solutions for ensuring that the explanation standards are met (Mittelstadt *et al.* 2019).

In the judicial domain, the need to promote public trust requires the possibility of identifying potential errors from automation systems and holding accountable those responsible for errors, just like human-performed judicial activities lead to accountability of human actors. Therefore, judicial automation algorithms should be transparent in the sense proposed by Joanna J. Bryson and Andras Theodorou (2019, 318): not in terms of complete comprehension of the inner workings of a system, but as a tool for ensuring human control over judicial proceedings.

When it comes to judicial automation, this form of transparency is essential not just to ensure that the systems function as intended, but also to the realisation of judicial principles. First, it can provide the information necessary to show that the relevant factors — the matters of law and fact presented by the parties or otherwise relevant to the proceedings at hand — were considered. If this information is provided in human-readable formats, automation can also help the parties to understand and accept the judicial decisions. Finally, transparency, and the human control that it enables, is necessary for understanding the reasons that drive an automated system, and understanding the relevant reasons is essential for effective disputation of a judicial decision (see Dymitruk 2019b).

While the proper outlines of the control mechanisms will vary for each judicial system, human control over judicial automation systems will require at least the possibility of understanding and changing the decision heuristics that have been adopted into a system. Human thought also relies on heuristics (Wheeler 2018), but procedure law has mechanisms for error correction and, if necessary, holding human decision-makers liable. Since those mechanisms rely on sanctions which, at least within the current technological horizon, cannot be applied to the systems

[72] However, as Hildebrandt (2019, 113, note 84) points out, not all kinds of opacity are undesirable: the protection of privacy contributes to preserve the private lives of people by increasing their opacity to data-driven approaches to classification and decision-making.

themselves (Bryson and Theodorou 2019, 320), then automation should be governed by rules targeted at the humans who control the design, adoption, and use of automated systems.

Finally, judicial decisions are based on modes of argumentation that are particular to the legal professions, such as the weighing of legal principles (Maranhão 2017). Since the structure of arguments plays such a central role in how a legal case is decided, machine learning systems, which rely on statistical properties of data, might have difficulties in accurately representing and predicting legal decisions.[73]

An alternative to the use of statistics-based approaches to automation would be to use knowledge-based systems, which rely on formal representations of legal knowledge and therefore could capture the specific dynamics of legal argumentation (Sartor 2018). But, despite their formal sophistication, providing those systems with the kind of data they need to operate involves laborious processes of data processing and formatting (Ratner *et al.* 2019), meaning that the theoretical sophistication of knowledge-based legal AI is not yet matched by its practical applications. Therefore, judicial automation systems will need either to find ways to address the shortcomings of machine learning or knowledge-based AI or to combine those approaches, for example by applying natural language processing techniques to extract the data that will be used by a knowledge-based system (Robaldo *et al.* 2019).

4. Safeguards in judicial automation

4.1. Data protection law safeguards

As explained in Subsection 3.1, in the context of the judicial automation there are at least three safeguarding measures required by the GDPR:

(a) the right to obtain human intervention on the part of the controller,
(b) the right to express data subject's point of view,
(c) the right to contest the decision.

These safeguards are compulsory only if the legal basis for the decision-making based solely on the automated processing is the data subject's explicit consent (Article 22(2)(c) of the GDPR) or the necessity for entering into, or performance of, a contract between the data subject and the data controller (Article 22(2)(a) of the GDPR). As already explained, the only possible legal grounds for the judicial automation are not these two indicated cases, but the European Union or a Member State regulation allowing such automated data processing; in that case, the GDPR does not lay down minimum requirements for suitable safeguards. Nevertheless, given the

[73] As an example, Verma *et al.* (2017) have developed a system that achieves about 75% of accuracy in predicting dissenting arguments between US appeal court judges. For that system, the main predictive variables were: (i) the seating arrangements of the judges in the trial session; (ii) the length of their votes; and (iii) the number of references to case law, all of which are of limited usefulness, at best, for understanding how judges build their arguments.

significance of the judiciary in modern societies and the importance of personal data processed by courts and other judicial authorities, these three basic rights should be considered as the absolute minimum.[74]

The right to obtain human intervention. The WP29 names human intervention as a "key element" in establishing appropriate safeguards for the data subject's rights, freedoms and legitimate interests (Article 29 Working Party 2018). The guidelines emphasise that any human review of the automatically-generated decision must be carried out by someone who has the appropriate authority and capability to change the decision, and that the reviewer should undertake a thorough assessment of all the relevant data, including any additional information provided by the data subject.

Reading between the lines, it seems that the European legislator perceives human intervention as an antidote to possible automated system errors. Nonetheless, it should be remembered that even if the human review is exercised, there is no guarantee that the decision will result in a different or better outcome (Bygrave 2019). Considering the tendency to overtrust the results of the automated decision-making and the possibility that the human intervenors might themselves introduce biases and errors to the decision-making (Almada 2019, Kamarinou *et al.* 2016), it can be concluded that the human intervention may or may not serve as a suitable safeguard in this respect.[75]

The second reason why human intervention may seem a suitable safeguard is the belief that the use of automated systems would otherwise undermine human autonomy (High-Level Expert Group on Artificial Intelligence 2019).[76] The Ethical Guidelines for Trustworthy Artificial Intelligence highlight that the distribution of functions between humans and automated systems should follow human-centric design principles and leave meaningful opportunity for the human choice (the principle of respect for human autonomy).[77]

In the context of the judicial use of automated systems, it is worth considering whether the human involvement requirement should be extended to the right to the proceedings based *solely* on the human decision-making (at least during appeal proceedings). The decision on the extension of "suitable safeguards" would clearly depend on the European or a Member State legislator, who, in accordance with Article 22(2)(b) of the GDPR, are entitled to permit the

[74] Malgieri argues that these three safeguards can all be inferred from other than Article 22(3) provision in the GDPR (Malgieri 2019).

[75] The idea of deploying the automated decision-making *in response* to the human decision (the adversarial sub-model of semi-automated judicial decision-making) seems to constitute a proper countermeasure against the indicated risks.

[76] As the Ethics Guidelines for Trustworthy Artificial Intelligence put it: "Humans interacting with AI systems must be able to keep full and effective selfdetermination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans" (High-Level Expert Group on Artificial Intelligence 2019).

[77] According to this principle, "AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight" (High-Level Expert Group on Artificial Intelligence 2019).

automated judicial decision-making, but are at the same time obliged to ensure the suitable safeguards, which may or may not be the same as indicated in Article 22(3) of the GDPR.

The right to express data subject's point of view. The second suitable safeguard referred to in Article 22(3) of the GDPR is the right to express the data subject's point of view. While Malgieri considers that the data subject is free to exercise this right *after* the automated decision has been made (Malgieri 2019)[78], Kamarinou *et al.* argue that the data controller must allow the data subjects to express their point of view *prior* to a decision being made (Kamarinou et al. 2016). Kamarinou e*t al.* also notice that the data controller is obliged to deploy the measures to prevent the situation when the final decision is taken before the data subject is consulted (Kamarinou et al. 2016).

The right to express the data subject's point of view is also linked to the ethical principle of respect for human autonomy, protecting him or her against the possible undue impact of the automated systems on the data subjects. Therefore, the Ethical Guidelines for Trustworthy Artificial Intelligence emphasise that the overall principle of user autonomy must be central to the automated system's functionality (High-Level Expert Group on Artificial Intelligence 2019). In the context of judicial automation, there is no need to reach for the newly adopted ethical guidelines as the concept of the right to be heard is well-grounded in the case-law of the ECtHR as an element of the right to a court (ECHR Article 6).[79] In the procedural context, the right to be heard means that the parties to the proceedings have the right to present the observations which they regard as relevant to their case. This right can only be seen as effective if the observations are actually "heard", that is to say duly considered by the trial court (Guide on Article 6 of the European Convention on Human Rights. Right to a fair trial (civil limb) 2019).

The right to contest the decision. The third suitable measure to safeguard the data subject's rights, freedoms and legitimate interests is the right to contest the decision. Fulfilling this right in the context of the automated judicial decision-making entails primarily lodging an appeal against the automated decision taken in the course of court proceedings. Article 22(3) of the GDPR does not indicate whether the second decision has to be taken by a human or can be made by another automated system. However, considering that the human intervention requirement is still in place, it should be concluded that the data subject is entitled to appeal to the human.[80]

Interestingly, Recital 71 of the GDPR explains that the suitable safeguards should include the right to "obtain an explanation of the decision (…) and to challenge the decision". Thus, unlike Article 22(3) of the GDPR, the preamble to the GDPR explicitly refers to the need for the *ex post*

[78] Quot. "Another automated decision-making 'suitable safeguard' mentioned at Article 22(3) GDPR is the data subject's right to express his/her point of view to the data controller after that an automated decision has been taken" (Malgieri 2019).
[79] See, *e.g. Donadze v. Georgia,* application 74644/01, paragraph 35.
[80] On the other hand, Kamarinou *et al.* proposed an interesting idea to grant the data subject the right to appeal to the automated system (instead of to the human controller), on the ground that "machine learning algorithms have the potential to achieve a high level of objectivity and neutrality, whereby learning techniques can be made to disregard factors such as age, race, ethnicity, religion, nationality, sexual orientation, etc., if instructed to do so, more effectively than humans" (Kamarinou e*t al. 2*016).

explanation of the automatically-generated decision. The question of whether the data controller is required to present such an explanation has ignited a debate in the literature (see Bygrave 2019 or Malgieri 2019 for further details). At the same time, the right to "meaningful information about the logic involved, as well as the significance and envisaged consequences" established by the GDPR cannot be forgotten.[81] In line with the position taken by WP29, the controller should provide the data subject, in an easily accessible form, with the information about the rationale behind, or the criteria relied on in reaching the decision, enabling the data subject to understand the reasons for the decision (Article 29 Working Party 2018).

The debate around whether or not Article 22 and other parts of the GDPR address the right to the *ex post* explanation of automated decisions is of particular importance for the automation of the judiciary. WP29 quite rightly emphasises that the data subjects will only be able to challenge a decision if they fully understand how it has been made and on what basis (Article 29 Working Party 2018). And as Bench-Capon reasonably points out that "(i)t has always been argued by AI and Law practitioners that the decision is of secondary importance, and *what matters is the explanation* (…)" (Bench-Capon 2018). The already mentioned XAI research might be a good way to ensure the right to explanation, but unfortunately most AI research focuses on the prediction tasks rather than on providing explanations for them.

The explanation delivered by the automated judicial systems must satisfy two conditions. Firstly, the explanation should first serve the human control over the automated system. Secondly, it should enable the data subject to challenge the decision. If the automated judicial system is not able to issue the explanation (irrespective of whether it is a fully- or semi- automated decision-making process), it is impossible to verify the reliability of the automatically-generated decision. When acting as the decision-aid, the automated system must give sufficient explanation to the human being suggested; when the automated system takes the final decision, its explanation is a prerequisite for the possibility to contest and then assess it.

The second function of the explanation in the automated judicial decision-making is ensuring the proper justification of the decision taken in the court proceedings (and from the point of view of the party to the court proceedings – the right to know the justification of the decision taken against him or her). This results above all from the right to court guaranteed in the international human rights framework (e.g. ECHR Article 6, CFR Article 47). As "reasoned judgement" is one of the pillars of the right to a fair trial, the ECtHR indicates that the guarantees enshrined in ECHR Article 6 include the obligation for courts to give sufficient reasons for their decisions.[82] The ECtHR case-law also presents the need to justify judicial decisions. A reasoned judicial decision: (1) shows the parties that their case has truly been heard, (2) affords the possibility to appeal against the decision, (3) provides a public scrutiny of the justice system[83]. These points resemble the discussed safeguards laid down in Article 22(3) of the GDPR. This clearly

[81] Article 13(2)(f), Article 14(2)(g) and Article 15(1)(h) of the GDPR.
[82] *H v. Belgium*, application no. 8950/80; *Suominen v. Finland*, application no. 37801/97.
[83] *Tatishvili v. Russia*, application no. 1509/02, paragraph 58.

demonstrates that all three "suitable measures" are interconnected and underpin each other in the proper judicial automation.

All rights (discussed in Subsection 3.1) and the safeguards discussed above have to be taken into account both during the design stage of the automated system (when deciding to use the particular automated system in the judiciary) and during the system's overall operation (when the personal data are being automatically processed). The controller must take appropriate technical and organisational measures to ensure that all GDPR requirements are met, and be able to demonstrate the compliance with the Regulation. Considering the data protection *by design* and *by default* principles (Article 25 of the GDPR), the controller is obliged to adopt internal policies and implement measures which meet these principles.[84] The preamble to the GDPR also mentions the obligation to take into consideration the principles of data protection *by design* and *by default* in the context of public tenders[85], which can be considered particularly important for the courts and other judicial authorities willing to contract out the design stage of the automated system to a private company.

4.2. Safeguards from privacy law

Privacy law is applicable to judicial automation to the extent that the use of automation systems impacts a person's private or family life. Given the many possible uses of automation in judicial proceedings, an exhaustive description of the possible harms to privacy ensuing from automation would not be feasible, but Daniel J. Solove (2006)'s taxonomy provides an initial guide for navigating the possible threats to privacy posed by automation.

Solove (2006, 490) presents four main groups of privacy-impacting operations, based on the existing technological affordances. The first group consists of *information collection* operations, which, in a judicial context, refer not just to information about the parties, but about the legal professionals involved and the third parties connected to the proceedings. That information may be collected either from the textual data of the proceedings or collected through other mechanisms, such as the metadata involved in online form submission. Since automation may be both an instrument for information collection operations (e.g., a bot for scraping documents submitted by the parties) and a beneficiary of the outputs of such operations (e.g., by using the information as a basis for automated decision-making), it follows that the use of judicial automation must be based on privacy-compatible information collection procedures. Otherwise, even a system far removed from the collection might interfere with a person's informational self-determination by spreading or relying on improperly obtained data.[86]

[84] Recital 78 of the GDPR.
[85] *Ibid.*
[86] While the American "fruits of the poisoned tree" doctrine has not been explicitly incorporated into European case law, various rulings of EU Courts (such as C-583/13P, *Deutsche Bahn and others v. European Commission*, and Case C 110/10 P, *Solvay SA v. European Commission*) have annulled evidence obtained through means incompatible with the proper exercise of the rights to defence present in CFR Article 48(2) and ECHR Article 6(1).

The second group of privacy-impacting operations is that of *information processing*, referring to the procedures involved in using, storing, and manipulating data (Solove 2006, 504–505). From a judicial perspective, it is important to highlight two of the modes highlighted by Solove. One of them is *secondary use* (Solove 2006, 518–520): data handled for one purpose might be used for other ends, thereby frustrating the expectations that legitimated the original use.[87] There is also the matter of *exclusion* (Solove 2006, 521–523), where persons might be deprived of the possibility of exercising their privacy rights because they lack adequate information about the existence and nature of the information operations which affect them. In both cases, there is an intrusion into a person's private life, both because the unexpected uses or lack of information affect that person's right to informational self-determination and because the omitted or reused information might constraint their decisional aspect, for example by preventing that person from seeking recourse from an automated judicial decision.

Operations that result in *information dissemination*[88] may also interfere with a person's privacy, as they might allow access to information that the person would not want to share, either to specific targets or the public in general. In judicial automation, the impact of information dissemination is particularly relevant when one considers the automation of proceedings involving confidential information, such as those concerned with children's rights. While modern legal systems usually include robust measures against wrongful dissemination, automation can potentialise the reach of traditional dissemination approaches, while introducing a new possibility: the dissemination of information *inferred* from procedural data (Wachter & Mittelstadt 2019), sometimes even despite the best technical efforts for anonymising that data (Rocher *et al.* 2019).

Finally, Solove (2006, 548) describes *invasion* operations that intrude in a person's negative liberty or their decision-making processes regarding their private life, covering a broad understanding of what Subsection 3.2 of this paper describes as decisional privacy. Once again, the limits of judicial decision-making already constrain what can be achieved by judicial automation, but enforcing decisional privacy in this context requires that one understands how automation tools may empower humans decision-makers, and thus expand the negative as well as the positive reach of their decisions. The same problem of magnification of negative consequences can affect fully-automated decision-making processes, to the extent that the choices regarding the design and use of such systems may reproduce individual and systemic biases, lacunae of knowledge, and other sources of unfairness.

For some of the situations described above, data protection laws will provide the legally expected level of protection. Still, as discussed in Subsection 3.2, some cases that fall outside

[87] As an example, consider the different requirements (and prohibitions) each jurisdiction establishes for bringing into the proceedings evidence produced as a consequence of another lawsuit.

[88] The name suggests an intuitive unity between those operations, but, as Solove (2006, 523) clarifies, dissemination can happen through operations as diverse as disclosure, increased accessibility of existing information, and outright blackmail.

the personal and material scopes of data protection might still lead to undue interferences with a person's privacy, meaning that remedies must be sought within privacy law.[89]

Privacy law has two main effects on the design and use of automation systems. The direct effect relates to how the remedies provided by privacy law can be applied to address harms from undue automation or even prevent uses of automation that are incompatible with the law. Within the European human rights systems, ECHR Article 8, which concerns itself with the protection of private life, has been invoked to contest the use of automation by governments. As an example, a 2020 decision from the Dutch Court of The Hague has held that the law regulating the use of an automated fraud detection system by the Ministry of Social Affairs represented an interference with the private life of the evaluated persons that was incompatible with the possibilities for state action set by ECHR Article 8(2).[90]

At first glance, those remedies might seem to be weaker than those available within data protection, especially when one considers that the rights established by GDPR Article 22 — the right not to be subjected to automated decisions, the relatively narrow set of exceptions to that prohibition, and the safeguards against legally permitted decisions — are explicitly based on two elements: (i) the fully automated nature of the data processing; and (ii) the personal nature of the processed data.[91] A further complication is that the ECHR and the CFR both establish limits to the interference with privacy by public authorities,[92] leaving the protection of privacy against private actors to the national legislators. But, since judicial proceedings are intimately connected to State action, any automation of judicial tasks will still need to comply with the protection of privacy established by ECHR Article 8.[93]

According to ECHR Article 8(2), public authorities can only interfere with privacy rights if three conditions are fulfilled, beginning with the exigency that any interference must happen in accordance with the law. In the case of judicial automation, breaches of the law may occur either in individual proceedings — due to localised errors, inadequate data, or misuse, among other factors — or as a systemic failure in compliance, which would result from design or organizational factors. Judicial contestation of a system might thus end up declaring that a

[89] As Kokott and Sobotta (2013) point out, informational privacy and data protection are not identical concepts, but data protection law explicitly recognises the protection of privacy as one of its key goals (see GDPR Recital 4 and Convention 108+, Article 1).

[90] The *SyRI case* (*NJCM and others v. the Netherlands*, ECLI:NL:RBDHA:2020:865), judged on February 5, 2020.

[91] This provides another example of situations where privacy and personal data are distinct: privacy law might still be actionable in cases of partially-automated processing or for operations that, despite not involving personal data, still impact a person's private life.

[92] Similarly, CFR Article 51(1) states that the rights of the Charter, including the right to privacy established by Article 7, "are addressed to the institutions, bodies, offices and agencies of the Union with due regard for the principle of subsidiarity and to the Member States only when they are implementing Union law."

[93] A complicating factor that falls outside the scope of this article is that the performance of judicial automation tasks might be outsourced to *software-as-a-service* (SaaS) providers or to contractors.

specific decision is unlawful or that the system itself is unable to fulfil the legal requirements in some or even all of the cases in which it is deployed.[94]

ECHR Article 8(2) only recognises a few aims as legitimate grounds for privacy intrusions: national security, public safety or the economic well-being of the country, the prevention of disorder or crime, the protection of health or morals, or the protection of the rights and freedoms of others. While some of those goals, especially the last one, are promoted by a well-functioning judicial system, it does not follow that privacy-interfering judicial automation is automatically permitted, as a third test remains: any interference must be necessary in a democratic society.[95]

As Kokott and Sobotta (2013, 225) show, the ECtHR requires that any interference must be supported by relevant and sufficient reasons and must be proportionate to the pursued aims, while the CJEU requires the demonstration of a fair balance between the interests that the interference is meant to promote and the interests that are harmed by it. Since ECtHR case law establishes that States that deploy new technologies bear special responsibility for striking a suitable balance,[96] establishing that any given instance of judicial automation fulfils this requirement will only be possible through multiple levels of explanation, ranging from the high-level justification of why a system does what it does to the formal description of the steps involved in automated judicial reasoning tasks (Kaminski and Malgieri 2019). Therefore, judicial automation systems should be mindful of privacy, by avoiding preserving the private life of persons or showing that any effects in privacy meet the requirements for state interference.

Beyond the direct application of privacy rules to existing decisions or systems, rules and principles grounded on privacy concerns can be used to orient the decisions made by human actors regarding those systems: when to buy automation? What variables should the system take into account? What are the goals that it is meant to achieve? Those and other questions treat privacy law as a source of *constraints* that might affect whether and how automation is designed and used, especially in a judicial context, a topic that will be developed below.

---

[94] Here, it is important to keep in mind the issue of how algorithmic error is perceived, a theme discussed in Subsection 3.1.

[95] ECtHR case law has established some guidelines for evaluating whether a measure is necessary in a democratic society. According to the court's Guide to Article 8 (ECtHR 2019, 12), a necessary interference must be proportionate to a legitimate aim that is being pursued (see Dudgeon v. the United Kingdom, §§ 51-53; Z v. Finland, § 94; Paradiso and Campanelli v. Italy, § 179) and respond to a pressing social need (see Piechowicz v. Poland, § 212; Paradiso and Campanelli v. Italy, § 181). In evaluating necessity, states have some margin of appreciation, which will be broader on matters where there is no consensus among the Member States regarding the importance of an interest or the best means for its protection, but is still subject to ECtHR control (Paradiso and Campanelli v. Italy, §§ 182–184).

[96] See, e.g., *S. and Marper v. the United Kingdom*, § 112: "In the Court's view, the strong consensus existing among the Contracting States in this respect is of considerable importance and narrows the margin of appreciation left to the respondent State in the assessment of the permissible limits of the interference with private life in this sphere. The Court considers that any State claiming a pioneer role in the development of new technologies bears special responsibility for striking the right balance in this regard."

4.3. Technological safeguards for judicial automation

Software systems are built through engineering processes. The tasks performed within such processes can be grouped into four main activities (Sommerville 2011, 28): the *specification* of the tasks that the system must perform and of the constraints to its operation, the *design and implementation* of a software system according to the specification, the *validation* of whether a system meets its specified requirements, and the *evolution* of an already-existing system in response to changes. Moreover, software systems are designed as components of complex systems, which include, besides the software itself, the hardware in which the software is run, and the organisational and social contexts in which the software is applied (Sommerville 2011, 264). All of those tasks and elements may be leveraged to control how automation systems are designed and used, thus providing additional mechanisms for the protection of rights in the context of judicial automation.

Automation systems can be complex, especially if they make use of AI techniques. That complexity can be an obstacle to privacy-respecting uses of automation, since the ensuing opacity[97] makes it difficult to anticipate negative side-effects. However, the effects of complexity can be mitigated through a modular approach to software design (Bryson and Theodorou 2019, 310), which breaks down projects into submodules that can be handled more or less independently in terms of their functionalities, priorities, and goals, and by keeping accurate logs of choices and changes that happen within the various software construction activities. Those practices increase the developers' capabilities for looking into a system, allowing for improved diagnosis and fixes to potential issues.[98]

Design practices can be used to empower not just system developers, but also the people that will be affected by automation practices once they are in place. Two complementary paths for technology-based empowerment have been presented in the AI literature. First, a system can be designed to facilitate the access to information about its effects,[99] either in a condensed format that allows direct action[100] or in a format that might require more active effort[101] by information recipients.[102] The second approach consists in designing systems that incorporate

[97] The discussion of opacity presented in Subsection 3.3 focused mostly on external perspectives, such as those of people affected by automated decisions, but the same opacity factors are applicable to software developers, especially when they rely on solutions or modules developed by others.

[98] As well as being a useful tool for judicial control of automation, as the data obtained from logging can be used to identify where and why a system failed to reach legally required standards of rights protection.

[99] Paul Ohm (2018) has proposed the establishment of a forthrightness duty, which would mean that those responsible by software systems would not be merely required to address the opacity inherent to their systems, but also to actively inform people of potential risks and harms. Implementing a proposal along these lines will require technical means such as those discussed here, but has the potential to promote informational and decisional privacy by allowing persons to make informed decisions about software.

[100] As an example, providing a person with access to operation logs will still require the perusal of the logs themselves before the sources of harm are identified.

[101] Such as the notification obligations established by GDPR Articles 13 and 14.

[102] While the choice of an adequate approach will depend on the recipient, techniques as diverse as data visualisation tools and zero-knowledge proofs (Almada 2019, 9) can be used to provide interested

tools that allow people to seek redress from automation, for example by facilitating the human intervention established by GDPR Article 22(3) as a right for those subject to fully-automated decisions based on personal data processing.

Even within a specific domain such as judicial proceedings, the choice of suitable technical measures will vary substantially based on the domain of application — a system for automatically downloading open judicial data will require different controls than those applicable to a robot judge —, but existing software design approaches can be used to direct the construction of judicial automation systems. Privacy by Design (Hoepman 2018) approaches have been used to build systems that promote informational privacy, and recent work (Almada 2019; Bayamlioglu 2018) has discussed the technical measures needed to adequately implement the right to contest automated decisions established in GDPR Article 22(3). In principle, those approaches can be extended to cope with privacy violations beyond the scope of data protection law, and/or combined with *computable law* techniques (Casanovas *et al.* 2008) for explicitly handling situations in which judicial automation may interfere with the informational or decisional privacy of those involved, directly or indirectly, in the automated proceedings.

For those judicial systems that effectively make decisions that must take into account human behaviour, rather than merely performing bureaucratic tasks, there is an additional challenge: as discussed in Subsection 3.3, computational systems may rely on incomplete images of human beings, either because the proceedings are meant to evaluate incomputable aspects of human lives or because heuristics or lack of data prevent a complete appreciation of what is at stake. To address these issues from the first moments of software construction, a combination of participatory and value-sensitive design approaches (Davis 2009) can be used to ensure that multiple perspectives on judicial automation are taken into account, properly balancing the values at stake. However, not all privacy violations can be addressed *ex ante*, as the interference in a person's private life might happen only when a usually adequate system is applied to a particular case. Remedies for contesting decisions, such as GDPR Article 22(3), become particularly relevant in this scenario, as well as the need to develop the technological infrastructure for proper contestation.

Since the technological and organisational measures outlined above require substantial effort, a crucial question is how to ensure that they are actually applied to automation systems. Here, regulation plays a crucial role, both through direct commands — such as the adoption of practices for data protection by design and by default, mandated by GDPR Article 25 — and by authorising infra-legal normative acts that establish standards. Even when practices or outcomes are not directly mandated by law, they can be used as criteria for evaluating whether an automation system meets criteria such as those established by ECHR Article 8.[103] And, while

persons with the information they need to know whether and how automation has been used and what are the remedies available to any ensuing harms.

[103] The adoption of adequate design practices can play two non-excluding roles in demonstrating a system's compliance with legal requirements. It can *demonstrate* that the software design process was based on an adequate balancing of the interests at stake, and it can also *enforce* the balancing choices

many of the constraints to judicial automation will result from statute and case law, private sets of best practices, as well as market pressures, will also influence the judicial automation systems, especially when those rely on commercial off-the-shelf (COTS) solutions. Therefore, there are various paths for enforcing the adoption of design practices that mitigate risks and adopt substantial safeguards for the protection of privacy rights[104] in the context of judicial automation.

5. Concluding remarks

Automation has the potential to improve judicial performance, either through the full or partial delegation of labour-intensive tasks to the machines or as a decision-aiding tool for human judges. However, judicial automation is only acceptable if it promotes, rather than hinders, the judicial system's goals and complies with the existing legal requirements. In this paper, we have considered two interconnected sets of legal constraints that are directly applicable to judicial automation. First, there is data protection law, which is applicable to the handling of data about natural persons within judicial proceedings — with a few exceptions, such as fully anonymised data — and establishes specific safeguards for automation, in particular a right to contesting fully-automated decisions (GDPR Article 22(3)). Even in situations outside the scope of data protection law, judicial automation might unduly interfere with a person's personal or family life, thus attracting the safeguards from ECHR Article 8 (and CFR Article 7), since judicial automation happens in the context of the performance of state functions.

Neither GDPR Article 22 nor ECHR Article 8 (and CFR) constitute a comprehensive framework for judicial automation, as some types of judicial automation may not lead to the violation of privacy or data protection rights. But, even though those norms do not establish or imply a wholesale prohibition of automated decision-making such as the one adopted by France, they still prohibit some applications of judicial automation, either because any form of a given application is incompatible with privacy and data protection, or because the required safeguards cannot be achieved with the existing technological capabilities.[105] In both cases, the legal acceptability of a specific judicial automation system will depend on an analysis of the context where the system is used, the laws applicable to that context, and the effects that can be obtained by that system within the current technological horizon.

Technology thus provides substantial constraints that might make a particular automated system unacceptable even in scenarios where there is no general prohibition to automation. Nevertheless, technological approaches can be leveraged to embed legal constraints as part of the processing logic that drives judicial automation systems, and the investigation of how to adequately incorporate legal considerations within the software design and use cycle can be

made, either explicitly or implicitly, during the software specification process. An example can be seen in the soft law role played by EDPS and EDPB opinions and guidelines (Dimitrova 2019, 211).

[104] For measures meant to extend the protection by design of rights beyond the scope of privacy and data protection, see Kulynych *et al*. (2020).

[105] The e-Privacy Regulation currently undergoing the European Union legislative process promises to enrich the regulatory system that is applicable to judicial automation.

useful for expanding the set of judicial tasks that can be automated, even that set, at least for the foreseeable future, does not encompass all relevant aspects of judicial activity. Therefore, further study is needed to map and analyse the legal and technological requirements posed by an algorithmic judiciary.

References

Acemoglu, D., and Restrepo, P. (2018). *Artificial Intelligence, Automation and Work*. NBER Working Paper No. 24196.

Almada, M. (2019). Human intervention in automated decision-making: Toward the construction of contestable systems. *Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAIL 2019)*, Montreal, QC, Canada, 2–11. New York: ACM Press.

Article 29 Working Party. (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

Bayamlioglu, E. (2018). Contesting Automated Decisions: A View of Transparency Implications. *European Data Protection Law Review* 4, 433–446.

Bench-Capon, T. (2018). Legal Cases: Argumentation versus ML, *ArgSoc Workshop at Comma 2018*, Warsaw, Poland.

Brachman, R. J., and Levesque, H. J. (2004). *Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Bryson, J. J., and Theodorou, A. (2019). How Society Can Maintain Human-Centric Artificial Intelligence. *Human-Centered Digitalization and Services*, edited by M. Toivonen and E. Saari, Translational Systems Sciences 19, Springer Nature Singapore.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *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. Karen Yeung and Martin Lodge (eds.), *Algorithmic Regulation*. Oxford University Press 2019, Forthcoming; University of Oslo Faculty of Law Research Paper No. 2019-01.

Cáceres, E. (2008), EXPERTIUS: A Mexican Judicial Decision-Support System in the Field of Family Law, Proceedings of the Twenty-First Annual Conference on Legal Knowledge and Information Systems (JURIX 2008), Florence, Italy, 10-13 December 2008.

Casanovas, P. *et al*. (eds., 2008). *Computable Models of the Law*: Languages, Dialogues, Games, Ontologies. Lecture Notes in Artificial Intelligence, Springer.

Cobbe, J. (2019). Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies, 39*(4), 636–655.

Committee of experts on internet intermediaries (MSI-NET). (2018). Algorithms and Human Rights – Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications. Council of Europe.

D'Amato, A. (1977). Can/Should Computers Replace Judges? *Georgia Law Review* 11, 1277–1301.

Davis, J. (2009). Design methods for ethical persuasive computing. *Proceedings of the 4th International Conference on Persuasive Technology*, Claremont, USA. New York: ACM Press.

DeCew, J. (2018). Privacy. *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.

Dijkstra, J. *et al.* (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*. 17(3), 155–163.

Dijkstra, J. (1999). User agreement with incorrect expert system advice. *Behaviour & Information Technology* 18(6), 399–411.

Dijkstra, J. (2001). Legal Knowledge-based Systems: The Blind Leading the Sheep? *International Review of Law, Computers & Technology* 15(2), 119–128.

Dimitrova, Diana. (2019). Data Protection within Police and Judicial Cooperation. *Specialized Administrative Law of the European Union*: A Sectoral Review, edited by H. C. H. Hofmann, G. C. Rowe, A. H. Turk, 204–236.

Dymitruk, M. (2019a). The Right to a Fair Trial in Automated Civil Proceedings. *Masaryk University Journal of Law and Technology* 13(1), 27–44.

Dymitruk, M. (2019b). Need for "Explainable Artificial Intelligence" in Automated Judicial Proceedings. *Doctoral Consortium at the 17th International Conference on Artificial Intelligence and Law (ICAIL 2019)*, Montreal, QC, Canada.

Ehsan, U., and Riedl, M. O. (2020). Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach. *Proceedings of HCI International 2020: 22nd International Conference On Human-Computer Interaction*. Forthcoming.

European Commission for the Efficiency of Justice (CEPEJ). (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*.

Fourcade, M., and Healy, K. (2017). Seeing like a market. *Socio-Economic Review* 15(1), 9–29.

Guide on Article 6 of the European Convention on Human Rights. Right to a fair trial (civil limb). 2019. Council of Europe/European Court of Human Rights.

Hildebrandt, M. (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20(1), 83–121.

Hoepman, J.-H. (2018). *Making Privacy By Design Concrete*. Technical report. KPN CISO Office, The Hague.

Kaminski, M., and Malgieri, G. (2019). Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations. U of Colorado Law Legal Studies Research Paper No. 19–28.

Kroll, J. *et al.* (2017). Accountable Algorithms. *University of Pennsylvania Law Review* 165(3), 633–705.

Immerman, N. (2018). Computability and Complexity. *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. (2019). *Ethical Guidelines for Trustworthy AI.* European Commission. B-1049 Brussels.

Kamarinou, D. et al. (2016). Machine Learning with Personal Data. *Queen Mary School of Law Legal Studies Research Paper 247*. Queen Mary, University of London, United Kingdom.

Kehl, D., Guo P., and Kessler S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Kokott, J., and Sobotta, C. (2013). The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Protection Law* 3(4), 222–228.

Kulynych, B. *et al.* (2020). POTs: Protective Optimization Technologies. *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, 177–188. New York: ACM Press.

Langford, M., and Madsen, M. R. (2019). France Criminalises Research on Judges. *Verfassungsblog*, June 22, 2019.

Malgieri, G., and Comandé, G. (2017). Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law,* 7(3).

Malgieri, G. (2019). Automated Decision-Making in the EU Member States: The Right to Explanation and Other 'Suitable Safeguards' for Algorithmic Decisions in the EU National Legislations. *Computer Law & Security Review*.

Maranhão, J. (2017). Value assessment and revision in legal interpretation. *Proceedings of the 17th International Conference on Artificial Intelligence and Law* (ICAIL 2017), London, UK, 129–138. New York: ACM Press.

Mittelstadt, B. *et al.* Explaining Explanations in AI. *FAT* '19: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, Atlanta, USA, 279–288. New York: ACM Press.

Ohm, P. (2018). Forthright code. *Houston Law Review* 56(2), 471–504.

Pałka, P. (2020). Data Management Law for the 2020s: The Lost Origins and the New Needs. *Buffalo Law Review*, forthcoming.

Ratner, A. *et al.* (2018). Snorkel: Rapid Training Data Creation with Weak Supervision. *Proceedings of the VLDB Endowment* 11(3).

Robaldo, L. *et al.* (2019). Introduction for artificial intelligence and law: special issue "natural language processing for legal texts". *Artificial Intelligence and Law* 27, 113–115.

Rocher, L. *et al*. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10.

Roig, A. (2018). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). *European Journal of Law and Technology*, 8(3).

Russell, S. J., and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*, 3rd edition. Upper Saddle River: Prentice Hall.

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154(3), 477–560.

Sommerville, I. (2011). *Software Engineering*, 9th ed. Addison-Wesley.

van den Hoven, J. *et al.* (2019). Privacy and Information Technology. *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.

van der Sloot, B. (2017). Decisional privacy 2.0: the procedural requirements implicit in Article 8 ECHR and its potential impact on profiling. *International Data Privacy Law, 7(3),* 190–201.

Venkatasubramanian, S., and Alfano, M. (2020). The philosophical basis of algorithmic recourse. *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, 284–293. New York: ACM Press.

Verma, S. *et al.* (2017). The Genealogy of ideology: predicting agreements and Persuasive memes in the U.S. Courts of Appeals. *Proceedings of the 16th International Conference on Artificial Intelligence and Law (ICAIL 2017)*, London, UK. New York: ACM Press.

Wachter, S., and Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review* 2019(2).

Wacks, R. (2015). *Privacy: A Very Short Introduction*. Oxford: Oxford University Press.

Wheeler, G. (2018). Bounded Rationality. *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.