Legal AI Systems in the EU's proposed Artificial Intelligence Act

Sebastian Felix Schwemer
Centre for Information and
Innovation Law (CIIR)
University of Copenhagen
Denmark
sebastian.felix.schwemer@jur.ku.dk

Letizia Tomada
Centre for Information and
Innovation Law (CIIR)
University of Copenhagen
Denmark
letizia.tomada@jur.ku.dk

Tommaso Pasini
Department of Computer Science
University of Copenhagen
Denmark
tommaso.pasini@di.ku.dk

ABSTRACT

In this paper we examine how human-machine interaction in the legal sector is suggested to be regulated in the EU's recently proposed Artificial Intelligence Act. First, we provide a brief background and overview of the proposal. Then we turn towards the assessment of high-risk AI systems for the legal tasks as well as the obligations for such AI systems in terms of human-machine interaction. We argue that whereas the proposed definition of AI system is broad, the concrete high-risk area of 'administration of justice and democratic processes', despite coming with considerable legal uncertainty, is narrow and unlikely to extent into many uses of legal AI and IA systems. Nonetheless, these regulatory developments may be of great relevance for current and future legal AI and IA systems.

KEYWORDS

European Union, Artificial Intelligence Act, AI regulation, high-risk AI systems, legal sector, intelligent assistance, human oversight, data governance

1 Introduction

On 21 April 2021, the European Commission presented its long-awaited proposal for a Regulation on Artificial Intelligence (AI), also referred to as Artificial Intelligence Act (AIA).[1] The proposal is the culmination of the EU's work on regulating AI, which started several years ago. In February 2020, the Commission had published its White Paper on AI [1], which set policy requirements on how to achieve the twofold aim to both promoting the use of AI and to address its associated potential risks. The proposal continues in the vein of the White Paper. With the aim to develop an 'ecosystem of trust', it sets out a legal framework for trustworthy AI with 'human centric' rules for AI, taking into account i.a. the recommendations of the European Parliament in its Resolution on a Framework for Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies [2].

The increasing use of AI and IA systems in the legal sector is transforming the legal practice by automating different parts of legal tasks. Legal actors will need to foster their professional skills, learning both how to use the new tools and also to supervise, question and interpret AI system outcomes. Yet, due to the diversity of the legal practice [3], for example in terms of areas of practice and organisational and business structures, it is not possible to generalise either on the impact of AI in this context or on what would be the appropriate level of AI-human interaction, which can occur at different stages.[2]

In some instances, algorithms are used in legal practice for purely administrative and organisational tasks, such as in the context of administration of justice. For example, in Poland an AI system for random allocation of cases has been implemented in 364 ordinary courts. Once per day it assigns cases to the judges of the specific court [5][6]. In other instances, AI systems are used in legal practice to perform tasks of contract review in the context of due diligence analysis or to carry out legal research. Concrete examples include eBrevia, which uses Natural Language Processing (NLP) to extract textual data from contracts and other documents, and LawGeex, which combines Machine Learning (ML) with text analytics and statistical benchmarks to check if contracts are within predefined parameters [7]. Another example, ROSS Intelligence, provides legal practitioners with natural language search capabilities [8]. AI systems are also relied on to automate document drafting [9][10][3]. Whereas several of these systems may be primarily relied on in private practice, some of their functionalities can also be relevant for uses in the public sector.

An even more advanced use of AI systems in legal practice is the adoption of data analytics or 'predictive' analytics. These methods can be used for example to regulate the provision of welfare [11] or to inform decisions and sentencing in criminal justice systems [12]. Methods based on statistical probabilities have been already used in these fields and current developments in ML techniques suggest

In: *Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021),* held in conjunction with ICAIL 2021, June 21, 2021, Sao Paulo, Brazil.

[1] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 21.4.2021, COM/2021/206 final.

[2] Arguing for a broad interpretation of human intervention, encompassing human action at early stages of design, training and testing see [4].

that they will be used in the future to assist in predicting legal outcomes in different types of cases [3]. Current examples of legal outcome prediction services include LexMachina (now LexisAdvance) [13] or Ravel Law [14]. The first was created to analyse decisions in the patent law sector. It extracts information on the patent, the parties and on legal findings and outcomes, with the aim to find patterns to provide insights on how future cases may be solved. The second combines NLP and ML to communicate insights on persuasive language to be used depending on the judge and to formulate predictions [14][15]. The AI system can inform legal actors on patterns, correlations and predictions upon observation of huge amounts of data [16]. These AI systems, however, make predictions based on ML rather than on legal reasoning and sometimes apply the learning to facts that are only assumed and not found in proceedings [3][17]. Also statistical analysis of decisions may be limited, for example because settled cases are excluded from databases or when few judgements are available due to the small size of the jurisdiction [3]. Both possible machine limitations and delicate implications that the AI outcome can have, raise the question what safeguards such AI systems shall be accompanied with.

In this context, our paper examines to what extent legal IA and AI systems are proposed to be subject to the new EU regime and maps challenges that the implementation of the proposed rules may pose for the uses of AI systems in legal practice.

2 Brief overview of the Artificial Intelligence Act

The proposed Regulation puts forward a legal framework with harmonised rules on AI. It introduces *inter alia* ‘rules regulating the placing on the market and putting into service of certain AI systems’ (recital 4). Notably, it does not introduce new rights for individuals affected by such AI system. Instead, it focusses on the regulation of the provider as well as the user[3] of such AI system.

Scope: The Broad Definition of “AI” System

The proposal defines an AI system in Art. 3(1) as ‘software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with’. This generic definition, inspired by the OECDs definition of an AI system, is complemented by Annex I, which contains a detailed list of approaches and techniques for the development of AI. These techniques and approaches, too, appear at first glance to be broad. They include not only (a) various ML approaches, but also (b) ‘Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems’ as well as (c) ‘Statistical approaches, Bayesian estimation, search and optimization methods’. This list can be amended by the Commission in order to be kept up to date (recital 6) in light of new market and technological developments, but only based on characteristics that are similar to the techniques and approaches (Art. 4).

By looking at the mentioned techniques and approaches, it is difficult to think of programs that do not fit into the broad definition. This latter, indeed, also includes software based on ‘handcrafted’ rules, which require no learning, such as Logic approaches, that can be entirely based on handcrafted rules expressed in some formal language. Another example are search methods which can be entirely based on heuristics that optimise the search in a large space of hypotheses. All these systems, while falling under the definition, do not learn from data (and therefore will not be much affected by the parts of the proposed regulation which focuses on ensuring perfectly curated datasets). It is also unclear, what exactly statistical approaches would entail. In any case also legal expert systems, including those with a manual knowledge acquisition process [18] or tools for constructing expert systems [19], might be considered AI systems that fall within the scope of the proposed Regulation. Furthermore, the definition does not negatively delimit to IA systems; in effect, it would cover both legal AI and IA systems as long as such system uses any of the mentioned techniques and approaches.

The question is whether the proposal includes indications that call for a restrictive reading of the broad definition. The very reliance on the notion ‘AI’ could imply that it needs to be interpreted more narrowly. Recital 3, for example, mentions that AI is a ‘fast evolving family of technologies’. The Impact Assessment accompanying the proposal does not provide much help with whether to construct the definition of AI system in a more restrictive manner either. The Commission puts its proposed definition in the context of the recent definition by the Organisation for Economic Co-operation and Development (OECD) [20], according to which an AI system ‘is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.’ But then it merely points towards a technological development and notes that AI systems traditionally ‘have focused on “rule-based algorithms”’ and that ‘AI systems currently in use often include both rule-based and learning-based algorithms’ [21]. Similarly, recital 6 notes that AI systems ‘can be designed to operate with varying levels of autonomy’.

Suffice it here to conclude that the mentioned techniques are broad and encompass a variety of –more or less advanced– systems. At the same time, the scope of application of the Regulation can only be understood in an overall view: Art. 6 in connection with Annex III (high-risk AI systems) provides for a far more restricted scope as will be explored further below.

Consequences: The Risk-based Approach

The proposal follows a ‘[c]learly defined risk-based approach’ (recital 14). Risk is defined by the International Standards Organisation (ISO) [22] as the effect of uncertainty on objectives.

[3] A user in the context of the AI Act proposal is defined in Art. 3(4) as ‘any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity’.

The proposed Regulation focuses on risks to the health or safety or the protection of fundamental rights of natural persons concerned (see e.g. recitals 1, 13, 27, 32, Arts. 7(1)(b), 65). It differentiates between four types of risk: AI systems with unacceptable risks that are prohibited; AI systems with high-risk that are permitted but subject to specific obligations; AI systems with limited risk, which are subject to certain transparency obligations; and finally, AI systems with minimal risk[4], which are not addressed by the Regulation (illustrated in Figure 1 below).

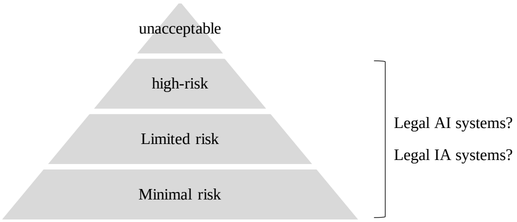

Figure 1: Risk model of AI Regulation

For legal AI and IA, only high-risk and below appear relevant. In the following, we do therefore not further address the specific and narrow AI practices that are deemed to carry unacceptable risks and proposed prohibited under Art. 5. Also, transparency requirements for AI systems with limited risks are outside the scope of our paper. Here, for *all*[5] AI systems in the legal sector, however, Art. 52(1) may be of interest: it introduces an obligation to inform natural persons of the fact that they are interacting with an AI system, unless obvious from the circumstances and context of use. In other words, an advanced legal chatbot may have to carry a label disclosing that the interaction is not taking place with a human.

Furthermore, we will not look at large parts of the proposal that deal with the procedural setup around e.g. *ex ante* self-assessments, conformity assessments or notified bodies, regulatory sandboxes as well as governance and enforcement. Suffice it here to note that the proposal should be seen in the context of product regulation and that a large part of concepts stem from the so-called 'New Legislative Framework'.[6]

Instead, in the following we focus on the category of high-risk AI systems (and the legal sector), before briefly commenting on non-high risk AI systems, where the voluntary application of obligations is encouraged.

3 High-Risk AI Systems and the Legal Sector

High-risk AI systems that are placed on the market or put into service are subject to certain specific obligations for *inter alia* providers, users, and importers. Firstly, systems that are used as a safety component of a product or a product covered by existing legislation in Annex II (e.g. machinery, medical devices or toys) or that are required to undergo a third party conformity assessment, are considered high-risk.

Secondly, and more relevant to our analysis, the placing on the market or putting into service of AI systems that are covered in Annex III, are considered high-risk (Art. 6(2)). The list in Annex III contains 8 pre-selected 'areas'[7], where the use of AI systems is deemed high-risk. The accompanying Impact Assessment further explains the Commission's methodology for this initial risk-assessment. It draws on a variety of sources including high-risk use cases from EP reports, a report by ISO, as well as from the pilot of the draft ethic guidelines by the High-Level Expert Group (HLEG) and the public consultation of the White Paper.[8] Each of the 8 areas contains at least one concrete 'use case'. Only these concrete 'use cases' in Annex III can be amended by the Commission (Art. 7(1)) with a view to include additional AI systems that fulfil two conditions: first, they are intended to be used in any of the areas listed in the Annex III; second, they pose a risk of harm to health and safety or of negative impact on fundamental rights which is equivalent or worse, in terms of severity and frequency, than the one posed by the systems already indicated in the Annex III. Because of the cumulativeness of the two conditions, any use case of an AI-system not falling within the scope of one of the pre-selected 8 areas, cannot become high-risk without a legislative intervention.

Annex III contains several high-risk 'areas', which can be of interest in relation to AI and IA in the legal sector (e.g. law enforcement; migration, asylum and border control management; or access to and enjoyment of essential private and public services and benefits). In the following we focus on the area of 'administration of justice and democratic processes'.

3.1 Administration of Justice and Democratic Processes

The 'area' of 'administration of justice and democratic processes' in Annex III point 8(a) contains only one 'use case' of a high-risk AI system:

> AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

Considering the broad definition of AI system (above), this 'use case' may seem to encompass a broad range of AI and IA systems in the legal sector at first glance. It is useful to break down this definition into a *positive* (what is covered by the definition) and a *negative* scope (what is excluded from it).

Assistance to a Judicial Authority

Firstly, the AI system must be intended for the assistance of a judicial authority. A first sub-question is thus, what 'judicial authority' encompasses since the proposal refrains from further defining the concept. In a narrow reading this could be restricted to

[4] According to the Commission, 'the vast majority of AI systems fall into this category', see European Commission, 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence', Brussels, 21 April 2021, IP/21/1682.

[5] I.e., irrespective of whether considered high-risk or not, cf. recital 70.

[6] See, e.g, Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC, *OJ L 218, 13.8.2008, p. 82–128.*

[7] While the proposal to a large degree continues in the vein of the Commission's White Paper [1], the 'sector' approach was dropped to the advantage of 'areas'.

[8] Annex 5 to [21], p. 41.

the 'authority capable of providing the effective judicial protection' guaranteed by Art. 47 of the Charter of Fundamental Rights of the European Union. [9] Such a reading *stricto sensu* would concomitantly imply that, e.g., the work of a Public Prosecutor's Office would not fall under this specific high-risk category. It is also unclear, whether e.g. complaint boards, widely established in Scandinavian countries (e.g. Personvernsnemda in Norway or Forbrugerklagenævnet in Denmark) would be covered.[10] Similarly, it seems questionable whether judicial authority would encompass alternative or online dispute resolution or arbitration.

In any case, AI systems intended for the private legal sector, including law firms (e.g. due diligence and contract review such as eBrevia or Lawgeex) or academic legal research, appear to clearly not fall within this category, with the consequence of being considered minimal risk.[11] This reading is also supported by the very title of the high-risk area ('Administration of justice and democratic processes'). In other words, only certain AI systems for the public legal sector appear to be considered high-risk.

It is less clear, however, how an AI system that is used both by practitioners and judicial authorities would be treated. Consider the following example: a system for legal information retrieval and case-law search, such as RossIntelligence used in private practice, is also be relied on by a judge. Would this change the risk category of said system? Both Annex III point 8(a) as well as the high-risk framework (cf., e.g., Art. 7(2)(a)) appear to emphasise the *intended purpose*–as opposed to *expected use*. Such purpose is defined in Art. 3(12) as meaning 'the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation'. In our reading this implies that such purpose is unilaterally defined by the provider of an AI system. Consequently, if an AI system is directly marketed towards judicial authorities, it would fulfil the first part of the requirement of the use case in Annex III point 8(a). Conversely, if an AI system is marketed exclusively towards private practice (but –unintendedly– used by a judge) it would likely not fulfil the requirement and thus not be considered high-risk. Thus, the lawmaker opted, in line with the risk approach, for a differentiated treatment of the *same* system dependent on its concrete use.

A second sub-question relates to the understanding of 'assistance'. How much/little of a human's work needs to be outsourced to the AI system in order to be considered 'assistance'? A look at the concrete role of the AI system in the following part may shed some light on this.

What AI Assistance is Covered?

The second *positive* scope regards the function of that AI system, which is stipulated as the assistance 'in researching and interpreting facts and the law and in applying the law to a concrete set of facts'. In our view, this point relates to the degree of automation of that assistance. How much human augmentation must the AI (or IA) system provide to be considered high-risk? And conversely, how much 'human agency' must be preserved for such system to be not considered high-risk?

A literal interpretation suggests that the listed functions are to be understood as cumulative. Thus, only an AI system performing *both* research and interpretation of *both* facts and the law *and also* applying the law to the facts would fulfil this criterion. It could be argued that AI systems for data analytics and predictions, such as e.g. LexisAdvance or RavelLaw, might fall within the scope of application. In a literal interpretation, however, it is important to stress that such system would need to not only assist in interpreting but also in *researching* facts. What exactly this entails remains vague. Legal information retrieval and case law search systems, such as e.g. RossIntelligence, in any case, would likely not be covered. Even when AI systems for case-law search and information retrieval are directly used by judicial authorities, they are neither as such assisting the authority in factfinding nor in the direct application of the law to the facts, despite that the design of search algorithms may present the risk of biases concerning what would be deemed as a relevant case and information that they display to the user [23]. A literal interpretation implies furthermore that intertwined tasks of a judge can be compartmentalized into decision-making and non-decision-making parts, which may not necessarily be the case [24].

Recital 40 helps understand *inter alia* the *negative* scope of Annex III point 8: it clarifies that the high-risk qualification should not encompass 'AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases' and brings as examples anonymization or pseudonymization of judicial decisions, documents or data, communication between personnel, or administrative tasks and the allocation of resources. The scope of the use case of the 'administration of justice and democratic processes' area, in any case, appears to be extremely narrow.

The question is therefore whether the *telos* of this sub-area is to only cover such integrated jack-of-all-trades legal AI systems, which may currently not exist. Bear in mind that the use case has been identified *inter alia* because of the 'increased possibilities' for use by judicial authorities in the EU.[12] Furthermore, we may ask: Does it make a difference–from the proposed Regulation's risk perspective–whether a judicial authority uses *one* AI systems with all these capabilities or *several* separate AI systems that collectively fulfil the criteria?[13]

A teleological interpretation might give leeway to a broader reading. Recital 40 provides further clarification of the lawmaker's intention regarding the area of administration of justice and democratic processes. It specifies that such AI systems should be

[9] See in this context, e.g., Case C-509/18, *Minister for Justice and Equality v PF*, Opinion of Advocate General Campos Sánchez-Bordona delivered on 30 April 2019, ECLI:EU:C:2019:338, point 18.

[10] Note, however, that 'AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services' are considered high-risk AI systems in Annex III point 5(a).

[11] Unless covered by another area in Annex III.

[12] Annex to [21], p. 46.

[13] See in this context also recital 6, noting that AI systems 'be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded).'

considered as high-risk 'considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial' and with the purpose to addressing 'the risks of potential biases, errors and opacity.' We suggest that AI systems performing only case-law search and information retrieval could be covered by the specific high-risk area in Annex III, to the extent that their search results may have an influence on 'democracy, rule of law, individual freedoms', on 'the right to an effective remedy and fair trail' and may pose a risk of 'potential biases and errors'. Similarly, despite recital 40 expressly excluding AI systems used for 'purely ancillary administrative activities' from being qualified as high-risk, the classification of a task as ancillary or not, may be not always straightforward. For example, the above-mentioned AI system for allocating cases to judges used in Poland [5][6], may be referred to as ancillary as it can be deemed to fall within the 'administrative tasks or allocation of resources' scenario. At the same time, however, a completely automated system of case allocation may still present risks of biases, errors, and opacity, which from a systematic perspective, might justify its classification as high-risk.

3.2 Future Cases of High-Risk Legal AI Systems

As explained above, the European Commission can add new use cases of high-risk AI systems to the 'Administration of justice and democratic processes' area. Notably, this is restricted to the addition of new use cases *within* the existing main 'areas' (in our example: administration of justice and democratic processes). Consequently, legal AI systems outside this area cannot become high-risk without legislative intervention. The very existence of mechanisms for adjusting high-risk areas could also be taken as an indication that the room for teleological interpretations of concrete high-risk AI use cases is restricted.

As demonstrated above, a lot remains unclear. The question is whether there is more clarity around potential future use cases of high-risk legal AI systems that could be added. Importantly, any future addition to the area of 'administration of justice and democratic processes' by the Commission, must according to Art. 7(1)(b) be 'in respect of its severity and probability of occurrence, *equivalent to or greater* than the risk of harm or of adverse impact posed by the high-risk AI systems'[14] mentioned in Annex III point 8(a). The Impact Assessment accompanying the proposal identifies two potential 'harms' in the area: firstly, '[i]ntense interference with a broad range of fundamental rights', e.g. relating to effective remedy and fair trial, non- discrimination, right to defence, presumption of innocence, right to liberty and security, human dignity as well as all rights granted by Union law that require effective judicial protection; and, secondly, a 'systemic risk to rule of law and freedom'.[15] The pre-identification of the analysed sub-area in Annex III point 8(a) is based on several indicative criteria, namely: (1) increased possibilities for use by judicial authorities in the EU; (2) potentially very severe impact and harm for all rights dependent on effective judicial protection; (3) high potential to scale and adversely impact a plurality of persons or groups; (4) high degree of dependency (due to inability to opt out) and high degree of vulnerability vis-à-vis judicial authorities; and (5) indication of harm (high probability of historical biases in past data used as training data, opacity).[16] All these aspects would like have to be considered when adding further use cases of high-risk AI systems.

In our view, also the accompanying recital could be drawn upon to help not only understand the scope of the specific area but also potential future high-risk use cases. The exact relation between recitals and Annex III, however, is unclear. Surprisingly, the corresponding recitals appear to relate not only the specific area but also the concrete high-risk use case of that area (e.g., point 8(a)). From a systematic perspective, this is peculiar: Annex III can be amended by the Commission, whereas corresponding recitals can only be changed by the legislator. We therefore suggest that the intentions expressed in recitals could be relevant not only for the area but also concrete use cases. In any case, the European Commission may be –under the aforementioned conditions– able to add additional use cases which are less restricted.

3.3 High-Risk Legal AI Systems, Quo Vadis

Figure 2 below illustrates the scope of high-risk AI systems and the area of 'administration of justice and democratic processes' in the context of legal AI/IA.

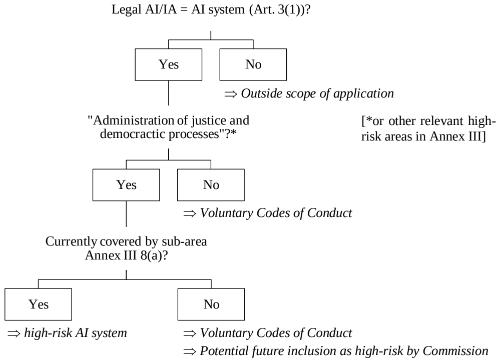
Legal AI/IA = AI system (Art. 3(1))?
Yes
No
⇒ Outside scope of application
"Administration of justice and democractic processes"?*
[*or other relevant high-risk areas in Annex III]
Yes
No
⇒ Voluntary Codes of Conduct
Currently covered by sub-area Annex III 8(a)?
Yes
No
⇒ high-risk AI system
⇒ Voluntary Codes of Conduct
⇒ Potential future inclusion as high-risk by Commission

Figure 2: Legal high-risk AI

The distinction between AI systems that may be deemed high-risk and the ones that may be deemed minimal-risk leaves, as described above, broad room for interpretation and thus legal uncertainty. Especially a clarification of which 'degree' of AI assistance would be required in order to fall within the scope of application, may be helpful in this regard. Selected consequences of this distinction, notably obligations for providers and users of high-risk AI systems will be analysed in the following section. The legal uncertainty of grey areas, however, may not necessarily undesired by the lawmaker. The proposed Artificial Intelligence Act encourages the voluntary application of the high-risk requirements by AI systems that are not considered high-risk (see below).

[14] Our emphasis.

[15] Annex to [21], p. 46.

[16] Ibid.

4 Requirements for High-Risk AI Systems

AI system deemed high-risk must comply with the requirements laid down in Chapter 2 of the proposed Regulation. These include a variety of obligations. Art. 9 provides for the establishment, implementation, documentation and maintenance (a 'continuous iterative process') of a risk management system in relation to the high-risk AI system. In this, foreseeable risks, for example, need to be identified and analysed (Art. 9(2)(a)) and other possibly arising risks evaluated (Art. 9(2)(c)). Further requirements relate, for example, to data governance, documentation, transparency and human oversight measures. The proposal also comes with a detailed oversight and enforcement regime, which is outside the scope of our analysis. Suffice it here to note that there exists a detailed setup and non-compliance can be fined with up to 6 % of a company's total worldwide annual turnover.[17] In the following, we highlight selected obligations, which can be of special interest with regards to legal AI systems.

4.1 Data and Data Governance

Art. 10 of the proposal sets quality criteria for training, validation and testing data sets to be used for the training of models of AI systems.[18] In particular, Art. 10(2) subjects the data sets to data governance and management practices.[19] Notably, such practices shall concern, e.g., examination in view of possible biases (Art. 10(2)(f)). Furthermore, Art. 10(3) requires data sets to be 'relevant, representative, free of errors and complete'. Legal AI often relates to the analysis of legal text and more advanced systems often rely on NLP. The requirement on data sets might pose challenges from a technical perspective. Different areas of NLP rely on transfer-learning techniques: that is, a neural network is first trained on large amounts of data to either predict the next word following a given sentence, or the word that is missing from the text, and then specialised on a particular task. Pre-training has been shown to work extremely well [27][26], and it is nowadays considered the standard approach to adopt. However, verifying the representativeness, completeness and correctness of the used datasets would be practically impossible since they usually count billions of tokens spanning across hundreds of languages.

Thus, one wonders how such models–which are nowadays used also in products–will be trained in the future. A similar problem is faced also when relying on large knowledge bases, e.g., Wikidata, Wikipedia, etc. In this case it is also not clear whether to consider such data as part of the training set and thus being subject to the Art.10(3), or if they can be overlooked and used straightaway. Both scenarios are not ideal as, in the first case, knowledge bases can still be a large source of bias thus resulting in unfair decision of automatic models. In the second case, the same doubts raised for large training sets apply. In fact, even though a knowledge base may be built manually, e.g., Wikipedia, it has no guarantee of being correct, and, even more, to be free of bias.

Recommender systems could also be largely affected by this requirement. Such models usually rely on signals generated by users (e.g., clicks, views, etc.) and their internal state is thus frequently updated based on them. While the initial training can be controlled, to some extent, by manually verifying the data, it is not conceivable to ensure the same high quality also after incorporating new data generated online by potentially millions of users.

A possible solution could be to also (if not only) regulate the behaviour of machine learning models by measuring their outputs bias and fairness with respect to protected groups depending to their application domain. Indeed, while data is surely a source of bias, models showed to also amplify bias or make spurious correlation that might not be evident by simply looking at the data [25]. Furthermore, this could leave more freedom to use large datasets, which are at the base of the current paradigm, while, at the same time, ensuring a fair and unbiased behaviour of models, also incentivising the development of new technique to algorithmically mitigate biases within data rather than fantasising on creating the perfect dataset.

Finally, Art. 10(5) introduces a legal basis for processing special categories of personal data for the purposes of debiasing. This clarification is of high relevance, since pursuant to the GDPR, modelers would have required an explicitly and freely given consent for the collection and processing of sensitive data.[20] Even if a justification may to some extent be obtained by interpreting debiasing as being a matter of 'public interest', thus falling under the exception of Article 9(2)(g) GDPR, which permits processing for reasons of substantial public interest [31], this may provide for a clearer legal basis.

4.2 Documentation, Transparency and Information

The proposed Artificial Intelligence Act also requires high-risk AI systems to be accompanied by a technical documentation, showing its compliance with the mentioned requirements (Art. 11), and developed with logging capabilities which enable the automatic recording of events and ensure the traceability of its functioning during its lifecycle (Art. 12). Moreover, the operation of the AI system must be in a transparent manner and accompanied by instructions for use (Art. 13). Of special interest in this context is that provided information not only needs to be concise, complete and clear but also 'relevant, accessible and comprehensible to users' (Art. 13(2)). Compliance with this requirement will call for an understanding and assessment of the expertise level of the user (Art. 3 and recital 49).

4.3 Human Oversight

The proposed Regulation also addresses AI-human interaction with a provision on 'appropriate' (recital 48) human oversight measures (Art. 14). The provision is far more detailed than the usual

[17] Similar to the administrative fines e.g. in the recently proposed Digital Services Act.

[18] Outside the world of trained algorithms, the proposed Regulation requires 'appropriate data governance and management practices' (Art. 10(6)).

[19] E.g. regarding design choices, data collection, data preparation including annotation or labelling, formulation of relevant assumptions, assessment of the availability, quantity and suitability of needed data sets or identification of data gaps.

[20] Cf. Art. 9 GDPR. See in this regard, e.g. [28], [29]. More specifically, on the challenges for the uses of sensitive data for debiasing purposes see, e.g., [30].

snippy human-in-the-loop lip service in other EU instruments (e.g. GDPR; Directive (EU) 2019/790; Recommendation (EU) 2018/334 etc.). Art. 14(1) requires high-risk AI systems to be *designed* and *developed* in a manner that 'they can be effectively overseen by natural persons' when the AI system is in use. Such manner 'includes' appropriate human-machine interface tools. The stipulated aim is to prevent or minimise the 'risks to health, safety or fundamental rights' (Art. 14(2)). Notably the benchmark are such risks that may emerge when the high-risk AI system is used 'in accordance with its intended purpose or under conditions of reasonably foreseeable misuse'. We have already commented on the concept of 'intended purpose' above. The boundaries of latter concept, 'foreseeable misuse', however, are not further defined in the proposal and remain vague.[21]

The measures which are meant to ensure human oversight must be either identified and built directly into the high-risk AI system, when technically possible, or identified by the provider and to be implemented by the user (Art. 14(3) (a) (b)). Further on, Art. 14(4) lists the goals that 'the individuals to whom human oversight is assigned' shall be able to achieve through those measures. Depending on circumstances, these include, e.g., a kill-switch or to be able to in a specific situation decide whether to override or reverse the output of that system. In this regard, it is interesting to explore what standard is set out for the 'human' that oversees the system use. It appears that their achievement requires an extent of technical expertise and knowledge. For example, the required ability to (a) 'fully understand the capacities and limitations of the high-risk AI system' and 'to monitor its operation' in order to detect and address possible dysfunctions, (c) the ability to 'correctly interpret the high-risk AI system's output' considering the characteristics of the system and the methods available and (e) the ability to 'intervene on the operation of the high-risk AI system' call for a certain degree of technical understanding of the system.[22] Possibly more easily approachable seems the required ability to (b) remain aware of 'automation bias', i.e. the 'possible tendency of automatically relying or over-relying on the output' produced by the high-risk AI system. This could be of special interest to certain legal AI systems –provided they are deemed high-risk– involved in the preparation of judgments, since Art. 14(b) refers specifically to systems that provide 'information or recommendations for decisions to be taken by natural persons'. Coming back to the standard for the human-in-the-loop, accompanying recital 48 adds that such measures guarantee that 'natural persons to whom human oversight has been assigned have the necessary competence, training and authority to carry out that role.' Interestingly, a previous leaked draft version of the Regulation contained specific provisions on 'organisational measures' in that respect.[23]

Importantly, Art. 14 only stipulates that high-risk AI systems must *feature* ('design and develop') appropriate human-machine interface tools. In other words, the obligation relates exclusively to the provider of such AI system; it does not stipulate an obligation for users to actually *perform* human oversight during operation.[24] According to Art. 29(1), however, users of high-risk AI systems are obliged to use such system in accordance with the accompanying instructions by the provider (which in turn may contain instructions on human oversight). Furthermore, users must monitor the operation based on the instructions of use (Art. 29(4)). Such clear and concise documentation (recital 46, see also above) must *inter alia* include a detailed description of needed human oversight measures. While not entirely clear and noting that there appears to be no clear obligation to *perform* human oversight in the Regulation, it seems that users may be obliged to implement the human oversight measures indicated by the provider and according to the specific instructions.

4.4 Obligations of Providers, Users and Other Parties

In addition to the requirements for AI systems addressed above, the proposal also establishes further specific obligations for providers, users and other parties. Providers, for example, need to implement a quality management system to ensure compliance (Art. 17), draw up technical documentation (Art. 18) and ensure that the AI system has been subject to a conformity assessment procedure (Art. 19), as well as to keep automatically generated logs (Art. 20). Furthermore, they are obliged to take immediate corrective actions when necessary and cooperate with competent authorities (Arts. 21 to 23). In addition to the obligations of providers, the proposed Regulation foresees obligations for product manufacturers (Art. 24), importers and distributors (Arts. 26 and 27). Finally, Art. 29 contains obligations of users of high-risk AI system, which requires –among other things– to use the system pursuant to the instructions of use, to monitor the system's operation on their basis and to ensure the relevance of input data when appropriate.

4.5 Self-Regulation

The scope of high-risk AI systems is restricted. As discussed above, many legal AI/IA systems would–despite the broad definition of AI system–likely not be considered high-risk. For non-high risk AI systems, the proposed Regulation instead foresees self-regulation. Both the European Commission as well as Member States are called upon to encourage and facilitate codes of conduct aimed at the voluntary application of the obligations set out for high-risk AI systems (Art. 69(1)). These codes of conduct can be implemented both on individual company-level as well as via broader industry collaborations. Thus, the above sketched requirements for high-risk AI systems might be relevant far beyond legal AI that falls within the limited scope of Annex III.

[21] The identification of such misuse would take place in the iterative risk management process by the provider of said AI system (cf. Art. 9(2)(a)).

[22] Compare also Art. 9(4)(c).

[23] The importance of organisational requirements had been previously stressed also in relation to human input in the context of the GDPR and of Article 29 Working Party interpretation's, where the requirement to ensure that the human has the 'authority and competence' to change the decision, has been identified as a 'social and organisational challenge'. See [26]. In this context, it is also interesting to highlight how the wording of point (d), envisaging the ability 'to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse [its] output' has changed in comparison to the previous version. The leaked draft had specified that the ability to decide not to use the high-risk AI system in any specific situation, could be exercised 'without any reason to fear negative consequences.'

[24] See, however, Art. 14(3)(b).

6 Conclusion

Margrethe Vestager proclaimed that the 'EU is spearheading the development of new global norms to make sure AI can be trusted' when presenting the proposal on 21 April 2021. Time will tell whether the proposal indeed will set the new global gold standard. In this context, it is noteworthy that the proposal does not follow a rights-based approach, which would, e.g., introduce new rights for individuals that are subject to decisions made by AI systems. Instead, it focuses on regulating providers and users of AI systems in a product regulation-akin manner.

In this contribution, we have looked at the relevance of the proposed Regulation in the field of legal AI systems. In the legal industry these recent regulatory developments are noteworthy. On the one hand, the definition of AI system is so broad that many existing legal AI/IA use cases would fall under the definition set forth by the proposed Regulation. On the other hand, only very few legal AI/IA systems would fall under the high-risk area of 'administration of justice and democratic processes'.[25] Legal AI/IA systems falling outside this area, notably AI systems in, e.g., private practice, will –provided they are not covered by one of the remaining 7 high-risk areas– likely not be considered high-risk, or at least not without further legislative intervention. Furthermore, also the specific use case of the analysed high-risk area is restricted in scope. At the same time, we highlighted areas of ambiguity and find that the proposal leaves significant grey areas. In these grey areas, however, self-regulation (in form of Codes of Conduct) might make the described requirements for AI systems relevant beyond the restricted high-risk areas and thereby for a larger variety of legal AI/IA systems.

It is important to note that the proposal will now be negotiated, changed and amended by the European Parliament and the Council in a process that can take up to several years.[26] Thus, it is very likely that we have not seen the final relevance of the EU's Artificial Intelligence Act for legal AI/IA systems yet.

ACKNOWLEDGMENTS

This research is part of the Legalese project at the University of Copenhagen, co-financed by the Innovation Fund Denmark (grant agreement: 0175-00011A). We thank Tobias Mahler for inspiring discussions and the anonymous peer reviewers for their helpful comments. All remaining errors are our own.

REFERENCES

[1] European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final.
[2] European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).
[3] Michael Legg and Felicity Bell. 2019. Artificial Intelligence and the Legal Profession: Becoming The AI-Enhanced Lawyer. *University of Tasmania Law Review*, 38, 2, 34 –59.
[4] Dimitra Kamarinou, Christopher Millard and Jatinder Singh. 2016. Machine Learning with Personal Data. Queen Mary School of Law Legal Studies Research Paper 247. Queen Mary University of London. 1–23.
[5] Alek Tarkowski. 2019. Report. Automating Society. Poland *AlgorithmWatch*, (29 January 2019) available at: https://algorithmwatch.org/en/automating-society-2019/poland/.
[6] Janneke Gerards and Raphaële Xenedis. 2020. Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law. 1-192. Available at https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1.
[7] https://www.lawgeex.com.
[8] https://rossintelligence.com/features.
[9] Kathryn D Betts and Kyle R Jaep. 2017. The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life into a Decades–Old Promise. *Duke Law & Technology Review*, 216–233.
[10] Benjamin Barton. 2014. The Lawyer's Monopoly: What Goes and What Stays. *Fordham Law Review*, 82, 6, 3067 – 3090.
[11] Dominique Hogan-Doran. 2017. Computer Says "No": Automation, Algorithms and Artificial Intelligence in Government Decision-Making. *The Judicial Review*, 13, 345-382.
[12] Nigel Stobbs, Dan Hunter and Mirko Bagaric. 2017.Can Sentencing be Enhanced by the Use of Artificial Intelligence?. *Criminal Law Journal*, 41, 5, 261–277.
[13] LexMachina, https://lexmachina.com/what-we-do/how-it-works/
[14] Ravel Law, http://ravellaw.com/products/#eluid7fcb4be2
[15] Frank Pasquale and Glyn Cashwell. 2018. Prediction, Persuasion, and the Jurisprudence of Behaviourism. *University of Toronto Law Journal*, 68, 1, 63–81.
[16] Dru Stevenson and Nicholas J Wagoner. 2015. Bargaining in the Shadow of Big Data. *Florida Law Review*, 67, 5, 1337–1399.
[17] Ajay Agrawal, Joshua Gans and Avi Goldfarb. 2018. *Prediction Machines — The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.
[18] Donald A. Waterman and Mark A. Peterson. 1981. *Models of Legal Decision Making: Research Design and Methods*. Rand Cooperation.
[19] Kevin D. Ashley. 2017. *Artificial Intelligence and Legal Analysis*. Cambridge University Press.
[20] OECD, Recommendation of the Council on Artificial Intelligence, 2019.
[21] European Commission, Commission Staff Working Paper, Impact Assessment, Accompanying the, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules rules on Artificial Intelligence, Brussels, 21.4.2021 SWD (2021) 84 final.
[22] ISO Guide 73, Risk management, ISO 31000.
[23] Susan Nevelow Mart. 2016. The Algorithm as a Human Artifact: Implications for Legal {Re}Search. 1–50. Available at: https://ssrn.com/abstract=2859720.
[24] Jason Millar and Ian Kerr. 2016. Delegation, relinquishment, and responsibility: The prospect of expert robots. In Ryan Calo, A. Michael Froomkin and Ian Kerr. *Robot Law*. Edward Elgar Publishing.
[25] Jieyu Zhao and Tianlu Wang and Mark Yatskar and Vicente Ordonez and Kai-Wei Chang. 2017. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. 2979–2989. DOI: 10.18653/v1/D17-1323.
[26] Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, Samuel R Bowman. 2019. SuperGLUE: A stickier benchmark for general-purpose language understanding systems. *Advances in Neural Information Processing Systems*.
[27] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy and Samuel R. Bowman. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 353–355.
[28] Niki Kilbertus, Adria Gasc et al. 2018. Blind Justice: Fairness with Encrypted Sensitive Attributes. *Proceedings of Machine Learning Research*, 2630 – 2639.
[29] Michael Veale and Lilian Edwards. 2018. Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. *Computer Law & Security Review*, 34, 398-404.
[30] Michael Veale and Reuben Binns. 2017. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 2, 1–17.
[31] Eirini Ntoutsi, Pavlos Fafalios et al. 2020. Bias in data-driven artificial intelligence systems – An Introductory Surver. *WIREs Data Mining and Knowledge Discovery*, 10, 1–14. DOI: 10.1002/widm.1356.

[25] Note, however, other high-risk areas in Annex III.

[26] The proposal suggests that most of the Regulation is applied only 24 months after its entry-into-force date.