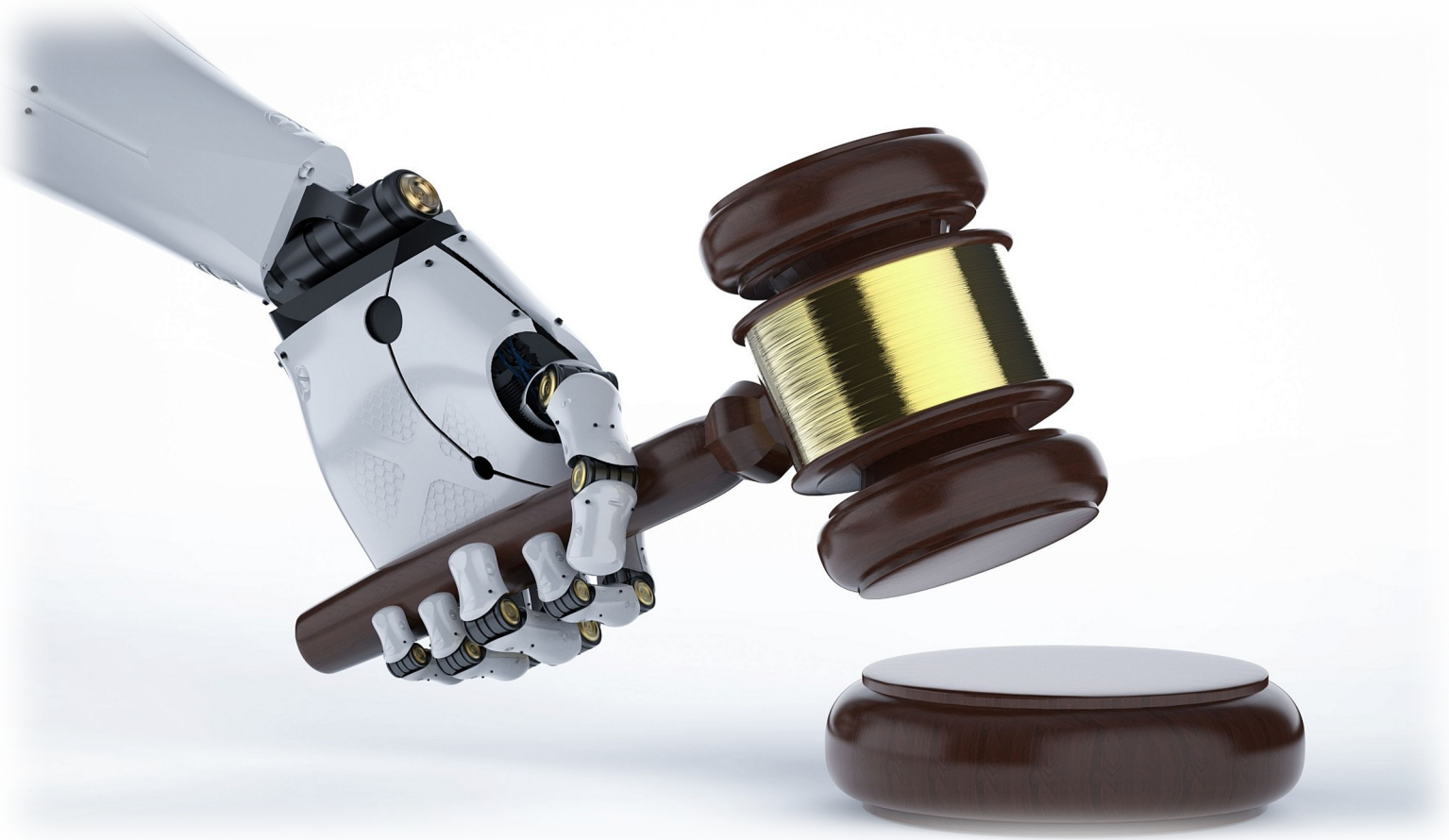
Capstone Amsterdam University College

The Implementation of Robotic Judgement: Requirements for the Regulation of Judicial Decision-Supporting Algorithms

Final Thesis Submission Date: May 27th, 2020



Justin Smael

Universiteit van Amsterdam

Amsterdam University College

VU Vrije Universiteit Amsterdam

Amsterdam University College

The Implementation of Robotic Judgement: Requirements for the Regulation of Judicial Decision-Supporting Algorithms

Capstone submitted in partial fulfilment of the

requirements for the Degree of Liberal Arts & Sciences

at

Amsterdam University College (AUC)

May 27th, 2020

Word Count: 16543

Author: Justin Smael, justinsmael1611@hotmail.com

Major: Social Sciences (Law)

Minor: Sciences (Information Science)

Tutor: Dhr. dr. C. (Cor) Zonneveld

Supervisor: Prof. dr. A. (Albert) Bomer, VU, a.h.bomer@vu.nl

Reader: Dhr. B.S. (Breanndán) Ó Nualláin BSc, AUC, o@uva.nl

Abstract

The possibility of implementing robotic judgement to solve legal disputes has been an object of study in the field of Law and Natural Language Processing (NLP) for decades, which has allowed the development of Machine Learning (ML) models that often revolved around the prediction of case outcomes. Rather than mere outcome prediction, algorithms trained on judicial data to support contemporary judges arguably have the potential to improve the efficiency and quality of legal decision-making, therewith providing more access to a higher standard of justice. The scholarly legal field often argues that algorithms remain unable to support judicial decision-making for reasons of input bias, opaqueness and the lack of a reasoned explanation. With the implementation of the General Data Protection Regulation's Article 22, stipulating the right to an explanation and a general prohibition of automated decision-making, this discussion has arguably complicated. This thesis shall address the contemporary status of the legal framework for algorithmic transparency, and its requirements for explainability. This research shall therewith evaluate the possibilities of applying a working definition of explainability to recently developed technical capabilities of ML and NLP, whilst classifying the analysed models based on their complexity. Through synthesizing academic literature, evaluating model performance measurements and based on expert advice, this thesis's main purpose shall be to demonstrate whether a future with robotic judgement is possible with the introduction of explainable judicial decision-supporting algorithms.

Keywords:

Judicial Decision-Supporting Algorithms (JDSA), Natural Language Processing (NLP), Machine Learning (ML), Algorithmic Transparency, Right to an Explanation

Acknowledgements

In conducting the research for this thesis, the contributions, opinions and feedback those who supervised and supported me have been of utmost value. I would herewith like to express my sincere gratitude. First and foremost, I want to express my gratitude to my supervisor, Professor Albert Bomer, who convincingly guided and encouraged me in the completion of this work. Besides my supervisor, I would like to thank the following people for giving valuable insights from the perspective their respective fields of expertise.

From the academic as well as technical perspective, I would like to specifically thank Jaap van den Herik (Leiden University), Floris Bex (Utrecht University), Masha Medvedeva (University of Groningen), Nikolaos Aletras (University of Sheffield), and the professionals Martin van Hemert, Jeroen Zweers, Jelle van Veenen, Willem Mobach and Roderick Lucas.

Besides, from a practical governmental perspective, I would like to thank Mildo van Staden from the Dutch Ministry of the Interior and Kingdom Relations. My gratitude also extends to the organisation of the NetLaw Media LegalTech talks and the Harvard Law School Center on the Legal Profession, for granting me the privilege to access the views of Richard Susskind and US Supreme Court Justice Ralph Gants on the subject matter of this thesis.

Furthermore, I whole-heartedly appreciated the advice and experience on the functioning of the judiciary as provided by Peter Cools, Gerard Tangenberg and Willem Korthals Altes. In relation to the practical adoption of the suggestions for my research, also considering the educational sphere for future judges, Gerard Tangenberg provided valuable insight of particular value. With the persistent and ever-available help of these experts, this thesis has been shaped into this final version.

Finally, I wish to acknowledge the support of my family in keeping me going, since this work would not have been possible without their support and input.

Table of Contents

List of Abbreviations

AI = Artificial Intelligence

ADM = Automated Decision-Making

BERT = Bidirectional Encoder Representations from Transformers

CEPEJ = European Commission for the Efficiency of Justice

CMLA = Computational Models of Legal Argumentation

COMPAS = Correctional Offender Management Profiling for Alternative Sanctions

DPIA = Data Protection Impact Assessments

ECHR = European Convention on Human Rights

ECtHR = European Court of Human Rights

EDPB = European Data Protection Board

DPIA = Data Protection Impact Assessments

GDPR = General Data Protection Regulation

JDSA = Judicial Decision-Supporting Algorithms

LIME = Local Interpretable Model-agnostic Explanations

MCE = Model-Centric Explanation

ML = Machine Learning

NLP = Natural Language Processing

SCE = Subject-Centric Explanation

SHAP = Shapley Additive Explanations

SVM = Support Vector Machine

WP29 = Working Party 29

xAI = Explainable Artificial Intelligence

Introduction

Robotic judgement becomes a reality when, through automatically predicting the outcome of a court case based on textual input describing the case's merits, judicial decision-making processes are handled automatically by algorithms. A "robotic judge", based on those algorithms, is then capable of delivering judicial decisions to the same, or higher, standard as contemporary analogue judges. The robotic judge would however do so through "computational power, vast amounts of precedential data and remarkable algorithms", rather than through nuanced human reasoning.[1] The implementation of robotic judgement-support for contemporary judges could, besides increasing speed and efficiency in the legal profession, allow more access to justice and arguably a higher quality of judicial decision-making.

As described in recent literature, the scenario of automatically predicting ex ante decisions based on quantitative models has advanced with technological improvements, specifically in the field of NLP.[2] Within the context of this advancement, various challenges can be identified, many of which are related to the interpretability of these algorithms and the reasoning behind the eventual output. Addressing these challenges, literature recognises the existence of a "right to explanation" within the GDPR regime, therewith setting thresholds for the interpretability of algorithmic modelling.[3] As generally recognised regarding technological advancement, the contemporary legal framework inadequately addresses recent developments. Furthermore, the paradox arises that whereas algorithmic performance increases, the capabilities of interpreting the models decreases, therewith

[1] Richard Susskind, Online Courts and the Future of Justice, (1st edn, OUP 2019) 280. Susskind argues for the implementation of online court proceedings to remove worldwide million-case court backlogs (partially as a consequence of the Covid-19 pandemic) and to allow more access to justice, by removing inefficiencies and costliness of legal advice and an overall improvement of efficiency and quality of adjudicational processes.

[2] Ilias Chalkidis, Ion Androutsopoulos & Nikolaos Aletras, 'Neural Legal Judgment Prediction in English' [2019] arXiv.

[3] Bryce Goodman & Seth Flaxman, 'European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"' [2017] AI magazine 50.

reducing the overall value of using those algorithms in a supporting capacity. To allow the implementation of robotic judgement, there are various challenges that need to be addressed.

This thesis shall delineate the extent to which decision-supporting algorithms are capable of delivering reasoned judicial decisions through hypothesizing that on the basis of technical developments, which transcend rule-based export systems and consider the application of state-of-the-art ML and NLP models, reasoned and explainable judgements can be delivered. Firstly, this thesis will address the contemporary status of the legal framework for algorithmic transparency, and its requirements for explainability. Secondly, the possibilities of applying the aforementioned working definition of explainability to the recently developed technical capabilities of ML and NLP shall be evaluated, whilst classifying the need to do so through analysing model complexity. Finally, applying the proposed legal framework to recent technical developments will enable the evaluation as to whether judicial decision-making benefits from the implementation of JDSA, therewith acknowledging or rejecting the core potential of robotic judgement in the near future (i.e. five years).

Chapter 1: Subject matter & Research Strategy

1.1 Research Context

Ever since the time of Harris's & Van Den Herik's articles on judicial decision-making and computers, the academic field of AI and Law has recognized the potential of digital transformation, originally referred to as "the computer revolution" in the judicial sphere, otherwise known as the "robot judge" in the field of jurimetrics or LegalTech.[4] In recent years, processing power, algorithmic capabilities and research in applied NLP and ML have developed to the extent that arguably allows for the implementation of algorithmic judgement-support based on textual data analytics.[5] As Susskind describes in *Online Courts and the Future of Justice*, the field aimed at developing ML systems to predict court behaviour has been researched extensively and advanced considerably in the past few years.[6] In earlier ML efforts, the work of Aletras and others, and Katz, Bommarito and Blackman is widely commended, achieving notable performance with relatively interpretable models.[7] In light of this advancement, more recent studies have demonstrated the effectiveness of ML in predicting judicial reasoning and case outcomes.[8] Whereas the amount of data and the legal scope of the datasets of these recent studies differ, ranging from the European Court on Human Rights (ECHR) to the Philippine Supreme Court and Chinese datasets, the overall trend shows improved performance of predictive models, whilst however increasing model

[4] Allen Harris, 'Judicial Decision Making and Computers' [1967] Villanova Law Review 272; Jaap van den Herik, 'Kunnen Computers Rechtspreken?'' [1991] Gouda Quint.

[5] Jesse Beatson, 'AI-Supported Adjudicators: Should Artificial Intelligence Have a Role in Tribunal Adjudication?' [2018] Canadian Journal of Administrative Law & Practice 307.

[6] Susskind (n 1) 277.

[7] Nikolaos Aletras and others, 'Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective' [2016] PeerJ Computer Science 2; Daniel Martin Katz and others, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' [2014] 12(4) PLoS ONE 1.

[8] Max R. S. Marques and others, 'Machine learning for explaining and ranking the most influential matters of law' [2019] In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law 239; Masha Medvedeva, Michel Vols & Martijn Wieling, 'Using machine learning to predict decisions of the European Court of Human Rights' [2019] Artificial Intelligence and Law 1; Shangbang Long and others, 'Automatic Judgement Prediction via Legal Reading Comprehension' [2019] Springer 558-572; Wenmian Yang and others, 'Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network' [2019] arXiv.

complexity.[9] A recent exemplification of delineating state-of-the-art ML and NLP capabilities is the evaluation of a variety of neural networks tasked with English legal judgement prediction on cases from the ECHR, which demonstratively outperform previous rule-based models.[10] Concerned with this advancement, Surden describes the importance of considering the limitations of the realistic and demystified use of ML, specifically arguing for realistically applying task-specific AI, bearing in mind the limitations.[11] As the scholarly field acknowledges, including Beatson and Yu & Ali, one of the fundamental limitations exists in the opaqueness of ADM, which results in the well-known "algorithmic black box".[12] Addressing this limitation, research in methods for interpreting and explaining models yields new opportunities for the implementation of judiciary-supporting algorithms, which raises both regulatory concerns and allows for new technical capabilities.[13] As Guidotti and others for example explain, understandable algorithms that weigh the input features of predictive models allow for global interpretability, and therewith arguably reduce limitations that would prevent the deployment of JDSA.[14] With incorporating interpretation methods, decision-supporting algorithms would be capable of delivering relatively explainable decisions, which identifies a current research gap in the scholarly field.[15]

Recognizing this research gap, it is important to consider that ensuring a higher quality of judicial decision-making can only be attained when transparency and

[9] Chalkidis and others (n 2) 1; Virtucio M B L and others, 'Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning' [2018] IEEE 42nd Annual Computer Software and Applications Conference 76; Yang (n 8) 1.

[10] Chalkidis and others (n 2) 7; Emad Elwany, Dave Moore & Guarav Oberoi, 'BERT Goes to Law School: Quantifying the Competitive Advantage of Access to Large Legal Corpora in Contract Understanding' [2019] arXiv.

[11] Harry Surden 'Artificial Intelligence and Law: An Overview' [2019] Georgia State University Law Review 35.

[12] Beatson (n 5) 34; Ronald Yu & Gabriele Spina Ali, 'What's Inside the Black Box? AI Challenges for Lawyers and Researchers' [2019] Cambridge Univeristy Press 2 5

[13] Riccardo Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' [2018] ACM computing surveys 1; Christoph Molnar, 'Interpretable machine learning. A Guide for Making Black Box Models Explainable' [2020] (10) https://christophm.github.io/interpretable-ml-book/.

[14] Guidotti and others (n 13) 6; Molnar (n 13) 104.

[15] Philipp Hacker and others, 'Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges' [2020] Artificial Intelligence and Law.

interpretability are accounted for in the decision-making process, and when the possibility of legal expert feedback is incorporated.[16] Contemporary academic research in the legal framework for algorithmic explainability predominantly concerns the GDPR, specifically Articles 13(2)(f), 14(2)(g) and 15(1)(h) and 22(1)(4) thereof.[17] Articles 13(2)(f), 14(2)(g) and 15(1)(h) require data controllers, which would be the judiciary in the case of JDSA, to provide information to data subjects, the parties in a court proceeding in the case of JDSA, about:

"[T]he existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, *meaningful information* about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[18]

Based on the wording of these provisions, scholars are divided on the type of explanation that is required. Edwards and Veale provide an in-depth analysis on the shortcomings of a right to an explanation in relation to ML, specifically through "algorithmic war stories.", and argue that it is unsatisfyingly defined at what point of the decision-making procedure data subjects can trigger the right to an explanation.[19] Edwards and Veale also state that ML algorithms are unlikely to provide the type of explanations required by the GDPR, mostly since ML explanations are restricted by the type of explanation that is sought, the complexity of the domain and the type of data subject. [20] To identify the type of explanation that would comply with regulatory standards, literature finds the "subject-centric approach", also referred to as local interpretability, most promising.[21] Using counterfactuals,

[16] Beatson (n 5) 21.

[17] Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' [2019] Columbia Law Review 1829; Lilian Edwards & Michael Veale, 'Enslaving the algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?.' [2018] IEEE Security & Privacy 46.

[18] EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament OJ 2016 L 119/1. (emphasis added).

[19] Edwards & Veale (n 17) 3.

[20] Lilian Edwards & Michael Veale, 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for' [2017] Duke Law & Technology Review 18. 64.

[21] Deeks (n 17) 13; Edwards & Veale (n 20) 47.

which entails the adjustment of input factors, data subjects are capable of understanding if and how a different outcome could be generated by the algorithm.[22] Other means of explainability employ "surrogate models", easier models that resemble the behaviour of the original ML model, or decompose the algorithm in entities of source code, which often has limitations for reasons of incompatibility, opaqueness and intellectual property violation.[23]

As a further method allowing interpretability, Molnar describes the LIME algorithm, which performs well in creating selective explanations based on counterfactuals.[24] Such analysis is however insufficient for complete causal attributions, and thus might not fulfil legal requirements, which are however currently too unclear to determine.[25] Another specific method of determining feature importance, SHAP, was introduced by Lundberg and Lee and described for practical application by Molnar.[26] SHAP combines game theory and local explanation methods to produce consistent and local feature importance values. SHAP is similar to LIME in the sense that it utilizes a locally fitted model to determine feature importance.[27] The difference between SHAP and LIME however is that SHAP does not produce a simpler model. As a result, it is arguably not possible to provide reasoned explanations and more technical expertise is required. With the introduction of the "*EU Guidelines on Ethics in Artificial Intelligence*" and the "*White Paper on Artificial Intelligence*", standards are in the process of being set for the regulation of interpretability of ML models.[28] Such a framework for the applicability of models such as LIME and SHAP lends itself to scholarly as

[22] Deeks (n 17) 8.
[23] Deeks (n 17) 9.
[24] Molnar (n 13) 113.
[25] Beatson (n 5) 30.
[26] Molnar (n 13) 120; Scott M. Lundberg & Su-In Lee, 'A Unified Approach to Interpreting Model Predictions' [2017] In Advances in neural information processing systems 4765.
[27] Molnar (n 13) 129.
[28] European Commission for the Efficiency of Justice (CEPEJ), 'European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment' (2018); European Commission, 'White Paper on Artificial Intelligence: A European approach to excellence and trust' (2020).

well as practical interpretation, which shall be analysed further in the context of the legal framework developed in this thesis.

1.1.1 Research Questions

This research shall explore the following main question:

A) Are recently developed models, based on ML and NLP, capable of implementing robotic judgement based on reason-generating and explainable JDSA?

In exploring this main question, the following sub-questions will be addressed:

I. What would be adequate requirements for model explainability in the context of JDSA, taking into account the legal framework of the right to an explanation?

II. Which ML and NLP models can be considered high-performing yet explainable with potential judicial capabilities, and what are the main challenges for these models?

III. Which specific requirements should be considered for the implementation of JDSA?

1.2 Methodology

1.2.1 Methods

For this thesis, the main methodology is based on an amalgamation of a literature review with expert interviews, synthesizing primary and secondary academic literature, also including white papers, governmental legislative sources and corporate investigations into the current state, and proposed future directions, on the topic of JDSA. The overall goal of this research will be achieved through conducting a literature review, developing a legal framework and evaluating implemented models. The aim of these methods is to test the hypothesis of demonstrating the technical capabilities of developing models capable of reasoned, explainable, and therewith legally enforceable judgements. The literature review shall bridge the gap between the fields of Law and Computer Science, whilst going in-depth where this proves necessary. This work shall attempt to avoid unnecessarily complicated technical explanations, through for example using analogies and simplifications. Where necessary, footnotes will refer to more in-depth technical literature. The selection of relevant literature shall take into account the timeliness, and recent applications, of studies, whilst considering which sources are deemed to contribute most to the rapidly developing discussion on the implementation of robotic judgement.

Besides, this research will provide meticulous documentation on interviews with experts, that will represent their main thoughts and consider the future application of their suggestions. To evaluate technical models, experts in the practical field of applied ML and NLP will function in an assisting capacity, whilst acknowledging that certain details might not be shared for reasons of intellectual property.[29] This research will thus conduct interviews

[29] Throughout this work: ML is defined as: "the process by which machines acquire direct task-specific knowledge. ML generally focuses on analysing historical data for patterns and relationships", whereas NLP is defined as "The ability of a machine to parse and deduce meaning from natural language as well as the ability to express knowledge and intentions from a digital system into natural language" Surden (n 11) 6; Jack Krupansky, 'Untangling the Definitions of Artificial Intelligence, Machine Intelligence, and Machine Learning' [2017].

with several experts from different roles in the triple helix consisting of academic, governmental and corporate perspectives, thus ranging from legal scholars to corporate programmers or lawyers, and from scientific scholars to current judges and policymakers. In its essence, the used methods shall thus be descriptive (literature review on technical capabilities in the legal field), analytic (determining the requirements for a definition of model explainability, and a legal framework regulating algorithmic transparency, both in the specific context of JDSA) and evaluative (whether the aforementioned requirements are adequate and comprehensive).

1.2.2 Structure

Chapter II of this thesis shall aim to provide an adequate definition for model explainability in the context of JDSA. The current European framework of explainability will hereto be analysed in light of recently published academic articles on the *right to an explanation* and algorithmic interpretability. Chapter III shall introduce the technical capabilities of NLP and ML in assisting the implementation of robotic judges through synthesizing the most recently used techniques. This chapter shall comprise a literature review on technical methods and will evaluate model performance in combination with interviews with technical experts, and will classify relevant models as being simple, moderate, or difficult to interpret and explain. Hereto, the differentiation between simple (such as regression models, decision trees or support vector machines), moderate (such as random forests and simple neural networks) or difficult (attention models or transfer learning) models and their capabilities of supporting judicial decision-making shall be described.

Subsequently, chapter IV shall explore suggestions for the improvement of the current legal framework regulating algorithmic transparency in predictive modelling. Critical analysis shall be attained through interviewing experts in the field and through comparing and

contrasting sources from the triple helix of academic, governmental and corporate views on the implementation of rules regulating the use of transparent and explainable algorithms. As such, this thesis shall evaluate the use of the definition of explainability (chapter II) in relation to the implementation of techniques described in more technical literature (chapter III), which shall result in critically synthesizing the extent to which robotic judges, based on JDSA, are capable of explainable judicial reasoning (chapter IV).

1.2.3 Limitations

This thesis shall have to constrain my interests in this field with the recognition of various limitations. Firstly, in experiencing the swiftly advancing academic literature on the "hot topic" of robotic judgement, this research shall focus on the legal concerns for the applicability, and usability of explainable recently suggested ML models, based on currently developed ML techniques and technical capabilities. Secondly, although recognising the fast-developing and self-learning capabilities of AI, this research does not argue for full-fledged judicial application of state-of-the-art models for reasons of the desire of preventing black box decision-making and the inability of delivering reasoned decisions. Since AI regularly gets mentioned in relevant literature, it is important to consider the actually deployed models, rather than the buzzwords that contextualize the development of judicial models. Thirdly, as a more fundamental limitation to the scope of this work, it shall not go in-depth on all the possibilities of process automation in the legal field, whether through blockchain or ML technology, since this research focuses on the judicial aspects and considerations for the implementation of supportive algorithms in judging cases. Since there often appear to be various misunderstandings regarding the technical aspects from a legal perspective, the explanations in this work will make use of interpretable descriptions and analogies, rather than in-depth mathematical explanations. This work will evaluate the recent scholarly aspects

of explainability, interpretability and transparency of model performance, and put no further emphasis on more philosophical considerations hindering the implementation of robotic judgement. Although ethical considerations or highly technical examinations might be mentioned, this research will not extensively address them in context of their extensive scholarly field. In conducting interviews, the scope of this research shall mainly comprise and thus be limited by the opinions of the Dutch judiciary combined with a variety of mostly European experts.

Chapter 2: Legal framework for explainability in JDSA

To determine the feasibility of implementing JDSA, it is important to consider the generally applicable contemporary legal framework related to the application of ADM. Within this general framework, as set out in for example the *European Commission's White Paper* on the use of AI, the explainability of ML models is particularly important in the context of an administrative capacity (e.g. the judiciary).[30] In the judiciary, the requirement of explainability is both required by law and allows contestability in judicial proceedings.[31] Since scholarly and regulatory sources often apply the terms transparency, interpretability and explainability in close relation to each other or even interchangeably, it is foremost important to consider the adequate definitions of transparency, interpretability and explainability in relation to judicial algorithmic modelling. Adequate definitions need to incorporate technical flexibility to allow comprehensive application for different types of ML models whilst ensuring legal certainty and harbouring the quality standards of judicial decision-making. Within the European legal framework, these terms are often used within the meaning of the "*right to an explanation*", as codified in Article 22 of the GDPR. Various other authoritative documents issued at European, domestic and academic levels shed light on the specific interpretation attributable to the explainability of algorithms functioning in a judicial capacity, but in doing so, there appears to be significant room for improvement, which shall be explored throughout this chapter.

[30] European Commission (n 28) 6. "[T]here is a need to build bridges between disciplines that currently work separately, such as machine learning and deep learning (characterised by limited interpretability, the need for a large volume of data to train the models and learn through correlations) and symbolic approaches (where rules are created through human intervention). Combining symbolic reasoning with deep neural networks may help us improve explainability of AI outcomes."

[31] Cary Coglianese & David Lehr, 'Regulating by robot: Administrative decision making in the machine-learning era.' [2016] The Georgetown Law Journal 105 1147. 1207. '[I]n administrative applications of machine learning, agencies will need to disclose algorithmic specifications, including the objective function being optimized, the method used for that optimization, and the algorithm's input variables.'

2.1. Academically explaining the GDPR "right to an explanation"

2.1.1 Relevant provisions

The GDPR "right to an explanation" is concerned with the legal implications of ADM, and particularly the explainability thereof. Article 22 specifically states that

"[T]he data subject shall have the right not to be subject to a *decision based solely on automated processing*, including profiling, which produces *legal effects* concerning him or her or similarly significantly affects him or her."[32]

Combined with specific provisions codified in articles 13(2)(f), 14(2)(g) and 15(1)(h), "*safeguards against automated decisions*" for data subjects are provided. In the debate regarding the involvement of algorithmic judicial models, it stands indisputable that if a robotic judge were to be autonomously implemented, "*a data subject*" would be subjected to a "*decision based solely on automated processing*" with clear "*legal effects*". Claiming otherwise would constitute a contradictio in terminis because automated judicial decisions have legal effects by definition.

Bearing in mind the wording of Article 22(1), the introduction of a robotic judge as various AI optimists might have in mind is simply prohibited by law for any "*data subject*" to whom the GDPR is applicable, unless the exceptions in Article 22(2) and Article 22(3) are applicable. Under Article 22(2)(c) and Article 22(3), "*explicit consent*" in combination with European Union or Member State Law that harbours "*safeguards for the data subject's rights and freedoms and legitimate interests*", allows automated processing. Any such safeguarding law should however at least allocate the data subject "*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*".

Furthermore, Articles 13(2)(f), 14(2)(g), and 15(1)(h) set forth that a data controller must provide data subjects with "*meaningful information about the logic involved*" of an automated

[32] Article 22(1) EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament OJ 2016 L 119/1. (emphasis added).

decision.[33] Since Article 13(2)(f), 14(2)(g), and 15(1)(h) are concerned with "*fair and transparent processing*" and refer to Article 22, various intricacies can be identified. In discussing Article 22(3). Recital 71 of the GDPR famously states that the "*safeguards*" should, inter alia, include "*the right [....] to obtain an explanation of the decision*".[34] Since the true "right to an explanation" is thus only contained in a non-binding recital, an academic discussion arose as to whether the right to an explanation follows from the GDPR.[35] Overall, the provisions construct a regulatory framework that applies to "robotic" judgement, since the role and responsibilities of the data processor are allocated to the judiciary. Thus, seemingly intelligible yet academically controversial provisions are to be followed, and the vaguely defined *right to an explanation* is to be respected.

Besides the GDPR, Article 6 of the ECHR sheds valuable light on the regulatory framework required for JDSA, since it establishes the *right to a fair trial*.[36] Since the specific provisions are arguably equivocal, it is important to consider the fundamental rights enshrined within the meaning of the *right to a fair trial* under various generally applicable standards of justice, particularly in light of (customary) international law related to the explainability and transparency of technological support in the judiciary.

[33] Edwards & Veale (n 17) 4.
[34] GDPR (n 18) Recital 71.
[35] Ibid.
[36] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended, hereinafter cited as 'ECHR') Article 6: Right to a fair trial. Fundamental rights will be more extensively addressed in Section 4.2.

2.1.2 Definitions

In a broad sense, explainability in legal modelling can be associated with transparency and interpretability, which can be defined in accordance with the following recently developed definitions.[37]

1. Transparency describes how an algorithmic model is to be understood, and why it returns a certain output. Transparency can be regarded at the level of the model as an entity (simulatability), the individual components (decomposability) and the learning algorithm (algorithmic transparency). For algorithmic transparency, input features such as bias are often considered. Within the meaning of the EU regulatory framework, the primary role of transparency is identified as a tool to enable algorithmic accountability.[38] If it is not known what an organisation is doing, it cannot be held accountable and cannot be regulated. Transparency standards often require different levels of detail for the general public, regulatory staff, computer scientists and researchers. The degree of transparency of an algorithmic system often depends on a combination of governance processes and technical properties of that system.

2. Interpretability describes the degree to which a cause or effect (e.g. a decision in judicial decision-making) in an algorithmic model can be explained. Interpretability can, post hoc, be achieved through visualizing what and how a model has learned (visualization), analysing parameters for a single decision (local explanations) or demonstrating the most similar output examples, for the reason of showing the logic of the process involved.[39] Although the aforementioned definition is often applied, ML

[37] Hacker and others (n 15) 17.
[38] CEPEJ (n 28) 54.
[39] ECHR Article 6: Right to a fair trial.

literature does not have a consensus on a definition of interpretability and state-of-the-arts methods often evaluate ad hoc interpretability through proxy characteristics to prevent opaque models.

3. Explainability is often equated with interpretability in legal literature. Explainability is however arguably more concerned with the ability to explain the internal mechanics of a ML or deep-learning algorithm in intelligible human terms.[40]

2.1.3 Scholarly debate

Legal scholars are fundamentally divided on the extent to which a right to an explanation exists, and if it exists, how to allow its effective enforcement. The debate started when Goodman and Flaxman argued for the existence, and implications, of the right to an explanation and set a first step in delineating challenges in designing algorithms and evaluation frameworks that enable explanation.[41] Wachter and others thereafter proclaimed that a right to an explanation does not exist, since the GDPR only effectively stipulates a "*right to be informed*", which constitutes a right to an ex ante explanation of the functioning of a decision-making algorithm.[42] Both the work of Goodman and Flaxman and Wachter and others shaped the public debate around the right to explanation, but were later regarded as not meaningfully addressing the relevant provisions of the right to an explanation, especially not the implications of "*meaningful information about the logic involved*" in ADM.[43] On the placement of the wording in the GDPR, scholars on the one hand argue that "*meaningful*

[40] Tim Miller, 'Explanation in artificial intelligence: Insights from the social sciences' [2019] (267) Elsevier Artificial Intelligence 1; Hacker and others (n 15) 17.

[41] Goodman & Flaxman (n 3) 55.

[42] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law 76.

[43] Goodman & Flaxman (n 3) 55; Andrew D. Selbst & Julia Powles, 'Meaningful information and the right to explanation' [2017] 7(4) International Data Privacy Law, 233.

information" is used interchangeably between Article 13, 14 and 15. On the other hand, it is argued that because of the wording of the context of the Articles, 13(2)(f) and 14(2)(g) require an overview of a system prior to processing, which constitutes an overall obligation for the data processor, whereas 15(1)(h) requires deeper disclosure, particularly for decisions that affect the data subject in a specific context. Arguably, this conflict has not been adequately resolved thus far.

Regarding the matter of "*meaningful information*", scholars remain widely divided.[44] Technical literature acknowledges the importance of the structure of architecture of the processing model and questions the weighing of input features as sufficiently "*meaningful*".[45] On the other hand, legal scholars advance that information is meaningful if it helps the data subject to exercise contestation rights to decision-making. If weights and input factors influence these contestation rights, then such information would need to be provided.[46] Miller argues how, in the context of applied ML, explanations can be made "*meaningful*" from the perspective of philosophical, psychological, and cognitive science and concludes that for human explanations, it is important to note that explanations are to be deemed:

1. contrastive and comparative,
2. selective, based on cognitive bias,
3. explanations are social, involving social values, and
4. probabilities are not as important as causal links in explaining reasoning.[47]

Based on this knowledge, counterfactual reasoning as proposed by Hume meets these four criteria and would thus be a good step in the direction of regulating and creating

[44] Hacker and others (n 15) 4.
[45] Brendt Mittelstadt, Chris Russell & Sandra Wachter, 'Explaining explanations in AI' [2019] Proceedings of the conference on Fairness, Accountability and Transparency 279.
[46] Supra 38; Hacker and others (n 15) 4.
[47] Miller (n 40) 6.

explainability in legal models.[48] In relation to automated support for the judiciary, "*meaningful information about the logic involved*", these values of explainability shall need to be considered, which will be further addressed in Chapter 4.

2.2. Applying the Right to an Explanation to JDSA

In order for ML to successfully function in the judicial context, *decisions with reasons* need to be constructed.[49] As Susskind describes, the scholarly AI and law community has sought to answer the question whether, through analysing fact patterns, identifying applicable laws and generating legal arguments, robotic judges could compute legally reasoned decisions.[50] Although thirty years of research have advanced during which numerous studies have been conducted, algorithmic systems remain unable to outperform judges at their own game.[51] Within the debate on algorithmic explainability, there however seem to be considerations that might advance the knowledge of the requirements for judicial determinations. In this regard, it is important to consider the aforementioned debate on the GDPR, where scholarly literature seemed to have reached a consensus on the expected substance of the right to an explanation. This consensus however resulted in further discussion on the extent to which the right to an explanation was to be interpreted and applied.[52] In this discussion, the interpretative "*Guidelines on Automated individual decision-making*

[48] Hacker and others (n 15) 16 "Since 1748, when David Hume presented his ideas on the importance of causality and counterfactual reasoning for explanations, many new theories have been proposed, but they are all more or less extensions of the idea of counterfactuals. While a causal chain explains a certain state or decision technically, this is typically not accepted as an explanation by the non-expert users of a system".

[49] Susskind (n 1) 281.

[50] David B. Wilkins interview with Richard Susskind, President of the Society for Computers and Law, Oxford University (Harvard Law School Center of the Legal Profession virtual book talk 'Online Courts and the Future of Justice', 23 April 2020).

[51] Ibid.

[52] Margot E. Kaminski, 'The right to explanation, explained' [2019] Berkeley Technology Law Journal, 189; Trevor Bench-Capon, 'The Need for Good Old Fashioned AI and Law' [2020].

and Profiling" by the Article 29 WP29 have significant influence.[53] The WP29 establishes that the right to an explanation does not have to be evoked by the data subject and that it should be interpreted as a general prohibition on ADM, if and only if ADM has legal effect. Thus, the *modus operandi* that would allow ADM has to deploy JDSA in an advisory capacity (e.g. providing predictions or identifying similarities in fact patterns), supporting the judiciary. According to the CEPEJ and the European Commission's White Paper, transparency is however an absolute requirement if algorithms provide an advisory function. Within the context of the CEPEJ Charter and GDPR Recital 71, it is however important to note that for reasons of protecting trade secrets, technical details and documentation would in certain cases fall outside the scope of the right to an explanation.[54] Other than this exception, standards for transparency and interpretability be adhered to in the implementation of judicial models.

2.2.1 Considerations in Regulating Interpretability

In the context of regulating interpretability of ML algorithms, the work of Casey, Farhangi and Vogl, Edwards and Veale and Brkan yield valuable insights.[55] First, it is important to consider that the GDPR framework does not differentiate in scope and purpose of ADM, which might lead to misleading generalizations.[56] In order to effectively distinguish such model-specific characteristics, it is important to consider characteristics of the cases for which the model would be applied, which Chapter 4 shall further elaborate on for JDSA. As

[53] Kaminski (n 52) 6; European Commission, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' [2018]; Working Party 29, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' [2018].

The WP29 is the predecessor of the current European Data Protection Board (EDPB), constituted by the presidents of all EU Member State data protection authorities and thus the leading authority in enforcing the GDPR.

[54] Kaminski (n 52) 15. This "loophole" is out of scope for this specific research.

[55] Bryan Casey, Ashkon Farhangi & Roland Vogl, 'Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise' [2019] Berkeley Tech 143; Edwards & Veale (n 20); Maja Brkan, '"Do algorithms rule the world?" Algorithmic decision-making and data protection in the framework of the GDPR and beyond.' [2019] International journal of law and information technology 91.

[56] Edwards & Veale (n 20) 10.

Edwards and Veale discuss with example of "*algorithmic war stories*", it is unclear when data subjects actually trigger the right to an explanation, and if the right is triggered, whether ML algorithms are capable of providing the explanations in accordance with the standards of interpretability.[57] To support this argument, Edwards and Veale demonstrate that ML explanations are restricted by the type of explanation that is sought and the complexity of the legal domain. Although overall sceptical of judicial modelling, Edwards and Veale acknowledge the promising ways of generating subject-centric explanation (SCE).[58] SCE is a method of "locally" interpreting a model, which is only concerned with specific sets of input data.[59] SCE can be contrasted with model-centric explanation (MCE), which is concerned with interpreting the global reasoning that a model uses to make certain decisions.[60] To achieve global interpretation, the behaviour of the used features, the trained components and many other parameters need to be explained. The lack of such explanation often leads to what is generally known as the "algorithmic black box", which Chapter 3 shall address further.

Secondly, as Brkan and the WP29 advance in line with Recital 58 of the GDPR as a response to the algorithmic black box, the complexity of algorithms is not a legitimate defence to fail transparency requirements.[61] If complex algorithms would thus be implemented in a judicial capacity, although in a supporting capacity, they would need to meet certain the requirements of transparency and explainability. As Casey, Farhangi and Vogl describe, these requirements are, in practice, interpreted by European Data Protection Authorities, who are currently shaping precedent concerned with the interpretability of ADM. An example of the establishment of such precedent is described by Brkan, who

[57] Edwards & Veale (n 20) 25.
[58] Edwards & Veale (n 20) 44.
[59] Edwards & Veale (n 20) 39.
[60] Deeks (n 17) 7.
[61] Brkan (n 55) 19; GDPR (n 18) Recital 58.

delineates how the Court of Justice of the European Union could interpret the right to an explanation in line with the interpretation method used in the adjudication of *Google v. Spain*.[62] In the same way of constructing and enforcing *the right to be forgotten* in *Google v Spain*, an interpretation of the legal principles enshrined in Articles 13(2)(f), 14(2)(g) and 15(1)(h) and 22 of the GDPR could thus construct the right to an explanation as applicable to the regulation of ADM and JDSA. For the subjects of the right to an explanation, this confers a data subject the right to know the reasoning behind an automated decision with legal effect.

Thirdly, as Hacker and others argue, explainability faces a highly uncertain future under the GDPR.[63] Beyond the scope of data protection law, Hacker and others therefore inquire the role of explainability in the legal domains of contract and tort law.[64] Besides, the GDPR does not specify what kind of human involvement seems to exempt the data controller from the constraints of the right to an explanation in ADM, given that a decision must be "based solely on automated processing".[65] The WP29 further adds that data controllers are not able to "*bypass*" GDPR Article 22(3) by "*fabricating human involvement*" meaning the involvement of analogue human judgement without any significant influence in the decision-making process, since the final decision would then remain to be "*based solely*" on automated processing. Morison and Harkens further delineate that meaningful human involvement exists if the human-in-the-loop acts in the capacity of overturning the final decision, which would be the case in JDSA.[66] Furthermore, it can be argued that in analysing whether a decision

[62] Google Spain, Google Spain SL and Google Incorporated v Agencia Española de Protección de Datos ('AEPD') and Costeja González, Judgment, Case C-131/12, ECLI:EU:C:2014:317, ILEC 060 (CJEU 2014), 13th May 2014, Court of Justice of the European Union; European Court of Justice (Grand Chamber); Edwards & Veale (n 20) 25.

[63] Hacker and others (n 15) 4.

[64] Hacker and others (n 15) 2.

[65] Wachter and others (n 42); Selbst & Powles (n 43).

[66] John Morison & Adam Harkens 'Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making' [2019] 39(4) Legal Studies 618.

should follow the algorithm, the involved analogue judge should be able to form a decision of the basis of all relevant data.[67]

2.3. Preliminary Conclusion Research Sub-Question I: An Adequate Definition of Model Explainability in JDSA

In judgements the explanation is what matters. Although the winning party is oftentimes most interested in the *dictum*, describing which demands will be met and what the overall outcome is, the losing party is more interested in the motivation why the court decided to rule this way.[68] In light of JDSA, it is important to consider the regulatory regime of the GDPR "right to an explanation", which is concerned with model interpretability and explainability. It is hereto important to consider the various scholarly perspectives, which set forth the limitations related to the interpretation of an inadequately defined right to an explanation. To reach an adequate definition for the legal framework of model explainability, two main aspects need to be considered. First, to construct a judicial legal explanation is not to simply follow the rule of law, but to reason why that particular rule is the rule that is to be followed given the circumstances, whilst weighing the influencing variables and arguments. Secondly, it is essential to set minimal thresholds regulating the degrees to which algorithmic decisions are constructed. Such thresholds should include, at least, a clarification of the input data that is used to reach a decision, and information on how certain features are weighed in reaching outcomes.[69] As Coglianese and Lehr further delineate, regulating explainability should not only consider the output and weights used in an algorithm but should also provide insight into the internal algorithmic functioning of a decision-making

[67] Interview with Willem Korthals Altes, Former Judge, Amsterdam District Court (Telephone call, 18 May 2020).

[68] Interview with Gerard Tangenberg, Senior Raadsheer, Gerechtshof The Hague (FaceTime, Online, 25 March 2020); Interview with Willem Korthals Altes, Former Judge, Amsterdam District Court (Telephone call, 18 May 2020).

[69] Goodman & Flaxman (n 3) 6.

process, or at least an accurate human-understandable approximation.[70] In the judiciary, this could entail a description of the algorithm's purpose, design and basic functioning. To apply ADM, the feasibility of applying these three aspects should be further researched, since their amalgamation enables an adequate means of regulating model explainability. Conclusively, this means that for an outcome to be classified as explainable, local interpretability as constructed by an understanding of the input data and the assigned weights in the decision-making process allows for meeting the minimum requirements of achieving explainable models. With the acceptance of these requirements, explainable AI (xAI) can allow the identification of case-by-case judicial bias, and machine-learned decisional support can be provided to the judiciary.[71]

[70] Coglianese & Lehr, 'Transparency and algorithmic governance' [2019] Administrative Law Review 1 6; Coglianese & Lehr, (n 31) 1207.

[71] Deeks (n 17) 4 "Common law xAI offers real promise as we head deeper into the age of algorithms. Courts will only be able to work xAI issues at the edges, looking across legal categories to draw on xAI developments in different doctrinal areas, but that work—and the response to that work by the creators and users of machine learning algorithms—may get us where we need to be."

Chapter 3: Explainable Judicial Capabilities of ML and NLP Models

Technical feasibility of digital data-driven support for the judiciary should, as Susskind argues, focus on the question whether computational systems are capable of delivering "*the social and economic outcomes*" that can be expected of judges, through the unhuman ways of ML, computational processing power, decision-predicting algorithms and legal argumentation models.[72] Research throughout half a century in rule-based, expert and now ML-based systems has yielded continuous progress and various perspectives on the answer to this question. In evaluating these answers, it is important to consider the actual benefits of digital support for the judiciary, which Jongbloed and others summarise as being three-folded: allowing efficiency, impartiality and the avoidance of the risk of human error.[73] Besides, a benefit of digital support is the ability of xAI to identify algorithmic error and input bias, especially from the model's input data based on previous analogue human judgements.[74] Since the more philosophical desirability of these benefits is outside the scope of this research, this chapter shall focus on the capabilities of achieving the benefits of digital support in the judiciary with the recognition and implementation of explainable capabilities of recent state-of-the-art ML models.

As recognised by both the academic field and the regulatory perspective of the CEPEJ, it is important to consider that the NLP and ML capabilities discussed in this chapter are concerned with modelling based on so-called "weak" AI, as opposed to the thus far science-fiction version of "strong" AI, also known as Artificial General Intelligence.[75] This distinction prevents the error of assuming that robotic judges would replace the work of

[72] Susskind (n 1) 281.

[73] Ton A. W. Jongbloed and others, 'The Rise of the Robotic Judge in Modern Court Proceedings' [2015] International Conference on Information Technology 59 8.

[74] Deeks (n 17) 8.

[75] CEPEJ (n 28) 30.

analogue judges in its entirety with overall ML intelligence and machine conscience that reasons and adjudicates in the same way an analogue judge does.[76]

3.1. Challenges in Developing Explainable ML

In the field of applied ML for judicial analytics, the challenges in creating explainability is two-fold. First, the question is whether ML can be made explainable for its users and subjects.[77] The second question is whether ML is able to develop computational models that perform legal reasoning, thereby explaining the modelled outcomes based on computational reasoning and ML logic.[78] Concerning the latter, extensive research has boiled down to two questions as answered by Ashley:

1. *"[H]ow text analytic tools and techniques can extract the semantic information necessary for argument retrieval"*, and
2. "*[H]ow that information can be applied to achieve cognitive computing*"

the latter would include constructing reasoned judgements with the desired aforementioned *"social and economic outcomes"* of the judicial function.[79] In answering these questions, Ashley delineates the differentiation between statutory (rule-based) and case-based legal reasoning. For rule-based reasoning, complications exist mostly in the sphere of semantic as well as syntactic ambiguity, whereas for case-based reasoning the elucidation of relationships between legal concepts and cases constitutes a computational complication.[80] With the development of models of evidentiary legal argument, ML models for classifying sentences as propositions, premises and conclusion and transforming legal information retrieval into argument retrieval, the question as to whether ML text analytic can retrieve

[76] Susskind (n 1) 280.
[77] Susskind (n 1) 281.
[78] Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, (1st edn, CUP 2017) 31; Morion & Harkens (n 60) 15.
[79] Ashley (n 78) 32, 39; Susskind (n 1) 281.
[80] Ashley (n 78) 39, 100.

reasoning is close to being answered in the affirmative.[81] For the matter of cognitive computing and sound legal reasoning, more complications play a role, including the matter of explainability for the users of and subjects to automated judicial systems. Of particular relevance in relation to the question of explainability for users is a study on perceptions on ADM.[82] Whilst for perceived usefulness the results were rather optimistic, the study showed that a large majority of the study's respondents (66%) perceived ADM by AI as risky. Based on these perceived risks, the academic challenges are mostly attributable to two main categories: the opacity of the algorithmic black box, or more generally the internal challenges to the modelling approach, and the challenges of input bias and other external challenges.[83]

3.1.1 Algorithmic Black Box and Internal Challenges of Opacity

The most generally established concern about algorithmic interpretability is that algorithms operate as black boxes, mostly because it is complex for a lay person to detect how an algorithm finds correlations.[84] Depending on the level of complexity and structure of the modelling approach, algorithmic decisions are easier or harder to comprehend, but understanding all the parameters is, for most models, simply infeasible, which should not however be the main aim of opening the black box.[85] The often-used example of a black box algorithm is a neural network, where the number of layers in a deep-learning approach might be good indication of the ability to trace how well the output can be traced back from the original input, but says very little about the overall process of getting from input to output.[86]

[81] Ashley (n 78) 5, 160, 287 & 327.

[82] Theo Araujo and others, 'In AI we trust? Perceptions about automated decision-making by artificial intelligence' [2020] AI & Society 1.

[83] Beatson (n 5) 27.

[84] Deeks (n 17) 9.

[85] Hacker and others (n 15) 16; "In the same way as it is impossible to understand all parameters of a complex model, we also don't put your brain's neurons under a microscope to find out how you make decisions" Interview with Willem Mobach, Senior Manager, Deloitte (Microsoft Teams, 22 April 2020) and confirmed in an interview with Rachel Rietveld, CEO, ArbeidsmarktResearch UvA (Telephone Call, 4 May 2020)

[86] Yu & Ali (n 12) 4.

For most ML algorithms, tracing the construction of the loss function and assigned weights for determining the correlations between different features is especially hard.[87] As Yu & Ali argue, algorithms have difficulty in distinguishing between causation and correlation, which might result in certain conclusions that might be right but are essentially based on the wrong inferences.[88] In establishing many false correlations and attributing them to a pattern of causation, the algorithm might undetectably construct an unaffordable risk when functioning in a partially judicial capacity, since the judge using it could consequently completely distrust the algorithmic outcomes. Furthermore, in terms of the internal challenge of opacity, unconscious bias in the minds of those designing algorithms strongly affects the internal challenge of opacity, which need to be prevented at all times in the context of JDSA.

3.1.2 Input Bias and other External Challenges

In hitherto implemented models with a judicial decision-supporting functionality, with the underlying ML in the COMPAS recidivism model as a prime exemplification, bias in input data have demonstrated severe implications, especially when such bias gets further supported with a positive feedback loop strengthening bias creating a discriminatory algorithm.[89] Especially in the context of achieving equality, fairness and transparency in judiciary and codified *right to a fair trial,* input bias in judicial support algorithms needs to be carefully evaluated.[90] Variance in the input of legal cases needs to be distributed to an extent to not merely allow fairness, but also a good representation of society at a certain point in

[87] Molnar (13) 22.

[88] Yu & Ali (n 12) 4.

[89] Alice Xiang & Inioluwa Deborah Raji, 'On the Legal Compatibility of Fairness Definitions' [2019] arXiv, COMPAS' risk assessment was used to inform a judge's decision on granting bail or even sentencing and proved to be highly racist because of data bias.

[90] Maria Dymitruk, 'The Right to a Fair Trial in Automated Civil Proceedings' [2019] Masaryk UJL & Tech 27 7.

time.[91] Timeliness of the used data through both political and societal influencing factors are all external challenges that can both be embedded, as well as discovered with the search for patterns in input data.[92] At this moment, the publicly available data for judicial modelling remains limited, because only a fraction of all judgements get published, thus not given an overall representative overview of societal developments addressed by the judiciary.[93] Moreover, the publicly available data mostly represents the special cases, and thus the outliers, which further complicates drawing good conclusions based on the available data.[94] Because ML algorithms inherently "learn" from past examples, especially in a supervised modelling approach, it is advisable to develop systems that mix semi-supervised or unsupervised classification techniques with the overall supervised argumentation modelling approach to find possible input data biases. For deployment in a common law system, the judiciary is influenced more by the societal implications of a judgement than in a statutory civil law system, which demonstrates the greater potential benefits as well as risks for the implementation of algorithmic support in the judiciary.[95]

Another fundamental external factor in model development exists in the need for expert annotation, which is especially relevant for classification algorithms. As generally recognised in the field of data science and big data, the training of large corpora (i.e. where n equals all) results in improved model performance. In the legal field, a lot of the available data

[91] Daniel L. Chen, 'Machine Learning and the Rule of Law' [2019] Computational Analysis of Law, Santa Fe Institute Press 1.

[92] Morison & Harkens (n 66) 18.

[93] Interview with Peter Cools, Raadsheer, Hoge Raad der Nederlanden (Telephone Call, 27 March 2020); Interview with Gerard Tangenberg, Senior Raadsheer, Gerechtshof The Hague (FaceTime, Online, 25 March 2020).

Cases that are published are generally considered the "outliers" that are of particular relevance to the development of the domestic rule of law, whereas the cases that are not published are often less influential. Since an algorithm learns best on larger datasets, it would be important to use as much of the available data as possible in implementing JDSA.

[94] Interview with Gerard Tangenberg, Senior Raadsheer, Gerechtshof The Hague (FaceTime, Online, 25 March 2020); Interview with Rachel Rietveld, CEO, ArbeidsmarktResearch UvA (Telephone Call, 4 May 2020); Interview with Willem Korthals Altes, Former Judge, Amsterdam District Court (Telephone call, 18 May 2020).

[95] Dymitruk (n 90) 7.

however remains too unannotated and too unstructured to use for modelling practices, which requires the input of legal experts before becoming valuable for the development of models in more specific and more complex legal domains.

3.2. State-of-the-Art ML and NLP Models

In recent years, research in the field of ML and NLP to transform and improve the legal field has advanced, especially in the application of neural networks and text-driven classification models. With the application of state-of-the-art models, such as Bidirectional Encoder Representations from Transformers (BERT) in an adapted form named hierarchical BERT (hier-BERT), research by Chalkidis and others has attained notable performance with a model transcending linear models based on more traditional bag-of-word approaches.[96] Elwany, Moore & Oberoi conclude that pre-trained and a thereafter fine-tuned BERT model adds significant improvement for accuracy and training speed in legal classification tasks, even without the need for complicated neural model architecture and expert-annotated legal text corpora.[97]

Despite high performance, the attention scores of such high-performing models only provide marginal indications of feature importance, which do not suffice as a legally sound justification of the model's decision-making process. In order to construct justifications that do meet the legal requirements of explainability, the research of applied ML and NLP models

[96] Aletras and others (n 7); Chalkidis and others (n 2) 7; Medvedeva and others (n 8). BERT is an open source applied ML in NLP model developed by researchers at Google in 2018. BERT has outperformed several models in NLP and provided top results in Question Answering, Natural Language Inference and more. BERT makes use of a so-called Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text, rather than the traditional bag-of-words or Word2Vec approach. BERT has a lot of applications for the legal market, as also recognised by the 2020 CodeX FutureLaw conference on "Is Law's Moat Evaporating?: Implications of Recent NLP Breakthroughs" retrieved: https://conferences.law.stanford.edu/futurelaw2020/sessions/bert-and-the-future-of-legal-analytics/

[97] Elwany, Moore & Oberoi (n 10) 4.

for judicial tasks is divided into three main categories, respectfully the work on case outcome prediction, models for legal argumentation and methods that allow users to interpret models.

3.2.1 Case Outcome Prediction Models

As Susskind quoted almost four decades ago from Lawlor, who stated in 1972:

> [T]he day should come… when you will be able to feed a set of facts to a machine that has cases, rules of law and reasoning rules stored in it, and in which the machine can then lay out for you, step by step, the *reasoning process* by which you may be able to arrive a certain conclusion. You can study it and then decide whether the machine is right or wrong. In some cases, the machine may not tell you exactly what the conclusion may be, but may say there is a probability that such-and-such is correct, and this probability is 90%.[98]

Recent research, arguably initiated by the work of Katz and others and Aletras and others, and practical application of initiatives like Lex Machina from LexisNexis, TAX-I from Deloitte and start-up LexIQ, show that this day has come.[99] Whereas academic interest in this matter is not a novel development, since there are remarkable systemic predictions methods from decades ago, recent developments in ML and NLP demonstrate the technical feasibility of implementing robotic judgement-support that succeed in strengthening the judiciary with reasoned legal explanations.

---

[98] Reed C. Lawlor, 'Excerpts from Fact Content of Cases and Precedent - A Modern Theory of Precedent' [1971] Jurimetrics Journal 245; Susskind (n 1) 281 (emphasis added).

[99] Katz (n 7) Aletras (n 7); Susskind (n 1) 282; Interview with Roderick Lucas, Manager, Deloitte (23 April 2020); Interview with Martin van Hemert, CEO LexIQ (Telephone Call, 31 March 2020).

As Ashley described, a study by Mackaay & Robillard from 1974 showed one of the first uses of a *k* nearest neighbour (kNN) algorithm for the prediction of outcomes of tax cases on capital gains tax.[100] As input, the algorithm considered a list of 46 binary text-based values, such as whether the capital gain was made by a "company", and if so (1) or not (0). The output is depicted in figure 1 and provides an outcome prediction in a two-dimensional placement of a new case in relation to previous cases, on which the algorithm was learned. Following the principle of *stare decisis*, a judge can argue for certain decision in line with earlier rulings.[101]

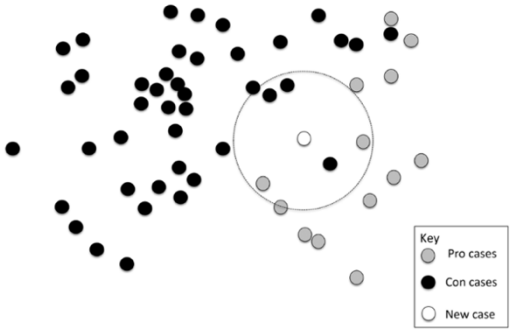


Figure 1: Representation of the two-dimensional output of the Mackaay & Robillard kNN model, developed in 1974

Although this case-based reasoning model can be deemed explainable, primarily because reasoning on 46 features remains relatively comprehensible for human understanding, it does not generate any form of legal reasoning.[102] The outcome of Mackaay's & Robillard's model is, one of the first attestations demonstrating the value of JDSA, because it helps in the identification of similar cases. To construct a sound legal explanation for a specified case in a legal domain, the prediction should however provide domain arguments that are tested in light of the judicial reasoning in the decision-making process of previous cases.[103]

Constructing judicial domain arguments is generally attained in one of two ways: legal argument-mining algorithms or rule-based systems. Since argument-mining will be addressed in Sub-Section 3.2.2, the latter shall first be evaluated. As Surden exemplifies in the tax domain, the rule-based system approach requires the specific input of features expressed in

[100] Ashley (n 78) 302.
[101] Trevor Bench-Capon, 'Arguing with cases' [1997] JURIX 85 1.
[102] Ashley (n 78) 114.
[103] Ashley (n 78) 115.

numerical values (e.g. if income >= 91.000, then tax rate == 28%).[104] The downside of the rule-based top-down approach to computation is that all relevant laws and decision-processes must be explicitly programmed, whereas in the bottom-up ML approach an algorithm can construct its own correlations. As Chen argues, the prediction of judicial decisions by bottom-up ML can especially yield value in detecting judicial indifference, since the facts of a case and the relevant circumstances can be processed and visualized with the overall aim of de-biasing the decision-making process.[105] In the work of Katz and others, this is attempted through applying randomized decision trees to the last case decided before the target case, resulting in accuracy scores of 70.9% over a 60-year period, notable without overfitting on the input data.[106] The most influential features were related to behavioural trends by supreme court justices, which is arguably in line with the value of predictive analytics for judicial decision-making, since it allows analysing judge's responses to societal trends. In the work by Aletras and others, it is particularly noteworthy that the predictions were strongly based on patters in facts and procedure of cases, rather than on relevant laws, which demonstrates the true potential of the use of ML, rather than rule-based systems.[107] For consideration of comprehensive judicial implementation of ML, it is however important to note that, since the work of Aletras and others merely concerned European Court of Human Rights cases with undisputed facts, such patterns in facts will be less identifiable when the facts remain disputed, requiring human interpretation, and when systems are applied to multiple legal domains.[108]

[104] Surden (n 11) 14.
[105] Chen (n 91) 4.
[106] Ashley (n 78) 114.
[107] Morison & Harkens (n 66) 21.
[108] Morison & Harkens (n 66) 19.

3.2.2 Models for Judicial Argumentation

Throughout his extensive analysis on CMLA, Ashley describes various ways of constructing the judicial syllogism that is attributed to deduction-based decision-making in disputes with a substantially evident statutory framework.[109] Overall, interpretation is an important characteristic of judicial reasoning that has thus far not been accounted for in computational models.[110] Partly though the Mining and Reasoning with Legal Texts (MIREL) project, there however is a variety of available algorithms to apply argument-mining on previous rulings, which arguably permits the creation of judicial argumentation, in line with previous judicial interpretation. The most applicable models for judicial reasoning should include the following capabilities:

1. Evaluating the same legal argument in the context of different cases based on contextual distinction.[111]
2. Reasoning about the relevance of certain cases for the context of a case in the same domain.[112]
3. Drawing analogies to precedential cases.[113]

Although case-based reasoning yields promising results in the sphere of identifying judicial argumentation, the aspects of balancing on the basis of values and purposes has remained beyond the scope of implemented models.[114] For more specific argument-mining and argument-identification models, the extraction of semantic concepts and relations could over time demonstrate promising results for application the judiciary.

---

[109] Ashley (n 78).

[110] Henry Prakken, 'Komt de robotrechter er aan?' [2018] 2018(4) Nederlands juristenblad 269

[111] Ashley (n 78) 92; on the description of the case-based CATO algorithm, dealing with trade secret misappropriation in terms of legal factors. CATO could downplay or emphasize distinctions, for example when a side's argument cites a particular distinguishing Factor in the current facts, the program could downplay it by pointing out another Factor in the cited case that mattered for the same reason.

[112] Ashley (n 78) 114 describes the CABARET algorithm analysing the number of most on-point cases relevant to the matter to be adjudicated.

[113] Ashley (n 78) 119 describes of the GREBE case-based legal reasoning algorithm.

[114] Bench-Capon (n 52) 3.

3.3. Model Differentiation and Interpretability

In addition to predictive analytics and legal argument extraction to support adjudication, the possibility of generating model explanations through various methods, which open the black box of more complex algorithms, yields promising results. One of the extensively researched methods for doing so is determining feature values in algorithmic reasoning through the LIME, SHAP or DeepLift algorithms.[115] As Molnar however describes, the short-term memory of humans is incapable of processing a large quantity of parameters.[116] In ML applied for NLP matters, this short-term memory issue has been dealt with the use of the so-called Long Short-Term Memory algorithm, which addressed the incapability of capturing word meaning for processing longer word sequences. With the introduction of BERT, contextual value could be captured, which would be especially valuable for judicial processing.

In terms of interpretability, the aforementioned methods, including LIME, SHAP or DeepLift, apply local interpretability, since they are mainly based on a counterfactual approach for assigning weights to single input features. Such methods can assist a judge in determining the impact which a certain feature has on the outcome of the overall decision-making process. Asides from this notion of post hoc interpretability through for example feature-weighing methods, interpretability can be attained with simply using simpler models, more generally known as intrinsic interpretability.[117] As Rudin describes, implementing intrinsically interpretable models avoids the potential bad practice of using less-than-satisfactory explanation methods that simply meet regulatory requirements.[118] In the judiciary, a mandate could for example be "if there is an interpretable model with the same

[115] Molnar (13) 78.

[116] Molnar (13) 18; Andrew Slavin Ross, Michael C. Hughes & Finale Doshi-Velez, 'Right for the right reasons: Training differentiable models by constraining their explanations' [2017] arXiv.

[117] Cynthia Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.' [2019] 1(5) Nature Machine Intelligence 233.

[118] Rudin (n 117) 9.

level of performance as a black box model, then deploy the interpretable one".[119] As earlier research by Goodman and Flaxman acknowledges, which is in line with technical literature from Rudin and Molnar, the balance in the performance vis-à-vis explainability trade-off should at all times be considered.[120]

The differentiation in choosing the right model to perform a judiciary-supporting task should however also take into account the factors of the user's and subject's technical fluency, and the ability for a lay person to understand the model. As demonstrated in the upper-left corner of figure 2, and as generally agreed in the ML community, the highest functional performance can be achieved with more complex models, such as neural networks. For such models, quantifying the influence of input variables through interpretation methods arguably meets the requirements of explainability, but this does not mean the system becomes human-understandable, since there might simply be too many features, and features weights, that impact the model's overall outcome.[121] In the trade-off depicted in figure 2, ensemble methods (e.g. random forests) have explainable output, but the randomised aggregation of other methods makes global interpretability hard to attain. For neural networks, separate interpretation models have been researched widely, and alternatives to LIME, SHAP and DeepLift have been an active area of research.[122]

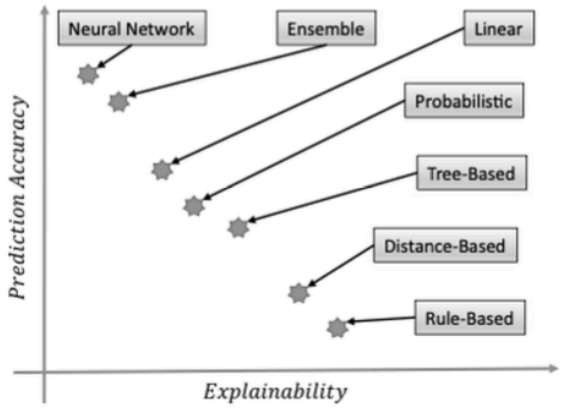


Figure 2: Accuracy and explainability trade-off as described by Hacker et al. (2020, p. 17)

[119] Rudin (n 117) 9.
[120] Rudin (n 117) 2; Molnar (13) 43.
[121] Deeks (n 17) 8.
[122] Goodman & Flaxman (n 3) 6.

3.4. Preliminary Conclusion Research Sub-Question II: Explainable Judicial Capabilities of ML and NLP Models

As delineated, scholarly research demonstrated the capabilities of applying ML and NLP models in the judge-supporting function of predictive analytics for case outcomes, modelling legal argumentation, and more general legal information retrieval. Classification models trained on legal text can be used for similarity matching between cases, and persuasive arguments can be detected in relevant previous cases, which can both support the analogue judge in adjudicating novel cases, and arguably also in constructing legal reasoning when combined with the legal expert knowledge that a judge possesses. Based on pattern detection and developing models that identify the most decisive elements in judicial reasoning, challenges for the judiciary such as bias, both algorithmic as well as societal, can be detected and the overall justice system can therewith be improved. The main challenges for such models can be differentiated on being internal or external, and each challenge requires different approaches for creating interpretability. Although the scholarly field and experts agree that full automation of the judicial function will not be reached in the foreseeable future, there already exist ML and NLP models that have the capability of improving adjudicational processes, especially for the bulk of simpler cases, which can be comprehended by analogue judges, who will for now be responsible for formulating the final explainable reasoning of judgements.

Chapter 4: Requirements for the Implementation of JDSA

To realize the benefits of applied ML and NLP, the judiciary ought to be reformed through a regulatory environment fostering innovation that is fact-based and data-driven, replacing the traditional notions of a more prescriptive regulatory framework.[123] Whilst ensuring that the traditional standards of justice are maintained, including the principles of fairness, equal treatment, integrity and transparency, this regulatory environment should allow algorithmic support in adjudication whilst sustaining the precedential rule of law.[124] In line with the principle of transparent and open justice and in allowing data-driven implementation of JDSA, all judgements from all courts, especially the bulk of simpler cases, have to be made available for legal analytics and the development of ML algorithms. Rather than merely publishing the outliers and complex cases, and quite arbitrarily determining this notion of "*publishable*" by the importance that a judge attributes to a certain judgement, judicial authorities thus need to require the publication of all judgements.[125]

With the support of ML algorithms, past discriminatory biases can be identified, and lessons can be learned from the data of past decisions.[126] The structure of legal argumentation can be analysed, and persuasive arguments can be generated based on the identified links between similar cases, all based on NLP models that have demonstrated the possibility of continuous improvement. Access to justice issues will be addressed, and the legal system can be allowed to operate more efficient, less costly (e.g. reduced clerk costs), and more open to those seeking justice with the use of digital tools, arguably whilst ensuring the same or higher standards of justice based on legal expertise. The challenges of digital illiteracy, the

[123] Interview with Supreme Court Justice Deno Himonas at the Codex FutureLaw 2020 Conference, hosted digitally at https://conferences.law.stanford.edu/futurelaw2020/sessions/regulatory-sandboxes-for-access-to-justice/.

[124] Dymitruk (n 90) 12.

[125] Interview with Willem Korthals Altes, Former Judge, Amsterdam District Court (Telephone call, 18 May 2020)

[126] Tania Sourdin, 'Judge v. Robot: Artificial Intelligence and Judicial Decision-Making' [2018] 41 UNSW Law Journal 1114.

accessibility of digital tools and the means of accessing complex legal knowledge will however need to be addressed appropriately. Furthermore, to realistically ensure that robotic support in the judiciary is implemented for all its desirable intents and purposes, standard requirements must be identified, and this chapter shall address two of the main factors to take into consideration. Section 4.1 shall hereto address the differentiation of simplicity and complexity in applying JDSA in the context of certain legal matters, whereas Section 4.2 will delineate the use of algorithms in an advisory capacity, for which algorithmic reliability should be a fundamental requirement.

4.1. Differentiating Simplicity and Complexity in JDSA

An often considered and widely addressed matter in judicial decision-making is the question of how to differentiate simple from complex legal matters.[127] Firstly, from the algorithmic perspective, Gardner's heuristic algorithm for distinguishing hard and easy legal questions provides a conundrum, since for a method of differentiation to be effective, it must be "easy" in itself. [128] Secondly, as Ashley explains, the factors of legal semantic ambiguity and vagueness further complicate making this distinction, and make it hard to model the distinction on the mere basis of a case its reasoning premises.[129] Thirdly, there is a more fundamental question of emotional values that arguably calls for specific requirements for differentiations of "hard" and "easy" in different legal domains. To exemplify this, consider the case of determining the simplicity of a tax evasion case in a corporate context, as compared to a matter of child custody in a family law setting. Although the facts might arguably be straight-forward for both legal questions, the influencing factors relatively easily determined and both cases are concerned with "hurt", the financial hurt in a tax case is

[127] Susskind (n 1) 147 "*A case that requires the finest judicial minds would for example be Donoghue v Stevenson.*
[128] Ashley (n 78) 19.
[129] Ashley (n 78) 39.

incomparable with the emotional values and "hurt" in the child custody case, making the latter seldomly an easy matter for a judge to rule on.[130]

Because of these three aspects in differentiating judicial cases, legal domain-specific requirements need to be established. Such requirements could stipulate that the liability of the use of ML techniques might differ depending on the applicable legal domain, which might regulate the extent to which a judge might rely on statistical insights to reach a decision concerned with a matter of tax, or family law. *The European Small Claims Procedures* provides an example of such regulation, particularly in establishing the determination of the necessity of oral proceedings in cases concerned with a monetary value below EUR 2000, which can be solved entirely through written procedure.[131] The ability to use computational insights might be regulated in the same way, as well as for the matter of determining the need for physical presence and allowing the possibility of technical means for alternative adjudication.[132]

As Susskind's consideration of whether court is a service or a place argues for online judging, cases could be decided "on the papers alone", thus without physical gatherings, for a wide range of lower-value civil claims, which would be particularly helpful for lower court judges.[133] In light of the Covid-19 crisis, the overall effects of the shift development of such solutions will arguably soon become a reality. For ML and NLP based legal reasoning and judicial support, an important consideration remains that well-functioning simpler models do not have the ability of constructing many-sided argumentation required for "hard" legal matters, but the bulk of simpler cases yields the promise of implementing robotic support in

[130] David B. Wilkins interview with Chief Justice Ralph Gants, Supreme Court Judge (Harvard Law School Center of the Legal Profession virtual book talk 'Online Courts and the Future of Justice', 23 April 2020) Quoted from Supreme Court Justice Gants.

[131] The European Regulation No. 2015/2421 on Small Claims Procedures. Article 5:

The European Small Claims Procedure shall be a written procedure. The court or tribunal shall hold an oral hearing if it considers this to be necessary or if a party so requests. The court or tribunal may refuse such a request if it considers that with regard to the circumstances of the case, an oral hearing is obviously not necessary for the fair conduct of the proceedings. The reasons for refusal shall be given in writing. The refusal may not be contested separately.

[132] The European Regulation No. 2015/2421 on Small Claims Procedures. Article 8: *The court or tribunal may hold an oral hearing through video conference or other communication technology if the technical means are available.*

[133] Susskind (n 1) 144.

the nearer future than many anticipated, since ADM with legal effect has arguably already been implemented to a certain degree.[134]

4.2 Advisory Algorithms as Secondary Judge; Trustworthiness of Algorithmic Support

As described throughout chapter 3 and as recognised by various scholars, no algorithm has thus far demonstrated the capacity to adjudicate new legal cases with a legally sound justification of its decision-making process.[135] Since the overall accountability of the judicial process will formally remain in the judge's hands for the foreseeable future, decision-supporting algorithms need to be trusted in their recommendations, and will have to demonstrate consistent reliability. To allow adequate functioning, advisory algorithms should arguably be evaluated to the same extent as any full-fledged ADM system, for which the CEPEJ provides valuable insight.[136] Uses that are encouraged within the framework of the CEPEJ include:

1. Case-law retrieval enhancement, particularly when combined with data visualization demonstrating the relevance of previous such as described in Sub-Section 3.2.1 with the Mackaay & Robillard model.
2. Access to legal documents and legal expert knowledge through applied NLP for automatically generating document and argumentation templates, such as delineated in Sub-Section 3.2.2 with the example of argument-mining.

[134] Marlies van Eck, 'Computerbesluiten, kunstmatige intelligentie en de bestuursrechter' [2019] OpenRecht. In line with the property valuation (Dutch 'WOZ') case ruled by the Dutch High Council in 2018, taxation software based on self-learned ML has already been used with legal effect. Since the used algorithms were considered an unexplainable 'black box', the court ruled that its usage was in violation of general Dutch administrative law ('Algemene wet bestuursrecht) in particular article 7(4) thereof, which is concerned with the contestability of decisions.

[135] Evert Verhulp & Rachel Rietveld, 'Hoe expertsystemen de rechtspraak kunnen helpen' [2019] 2019(2) Rechtstreeks 39. 5; Stefan Philipsen & Erlis Themeli, 'Een introductie op de robotrechter' [2019] 2019(2) Rechtstreeks 46.

[136] CEPEJ (n 28) 65.

3. Data-driven strategic tools for the efficient administration of justice and projected case outcomes, especially when developed in combination with legal professionals such as suggested in Sub-Section 3.2.1, which addressed predictive analytics.

Furthermore, alternative dispute settlement procedures, such as arbitration and mediation, and online dispute resolution could benefit from the use-cases defined above as long as they follow Article 6 and 13 of the ECHR. [137]

More careful considerations need to be made through additional scientific studies related to judge profiling and the anticipation of judgements, since these potential use-cases are highly dependent on external factors (such as input bias and a rapidly changing regulatory landscape) and have the potential of "hurting" the judiciary.[138] Just like the differentiation can be made in civil disputes for either single-judge rulings or full-bench chambers consisting of multiple, usually 3 judges, an advisory algorithm could also be applied as a secondary judge. Since full-bench chambers often adjudicate more complex disputes, algorithmic support with the aim of strategic efficiency and modelling argumentations is arguably better suited for application in single-judge matters.[139] For matters concerning appeals or cases for which full-bench chambers are required, there is arguably a lot more room for applying case-law retrieval for heavily documented longer-term cases, whereas strategic efficiency and argumentation modelling might be out of scope for the first generation of JDSA. As algorithms "learn" and improve, especially through the expert judgement feedback from judges capable of providing a third-party perspective, trust in the algorithm's performance

[137] ECHR Article 6: Right to a fair trial; Article 13: Right to an effective remedy.

[138] Gerard Tangenberg, Senior Raadsheer, Gerechtshof The Hague, confirmed this CEPEJ statement with the exemplification of a potential loss of trust in the judiciary because of the following revealing that tax cases brought to court in the Dutch Northern Provinces would generally have a higher chance of winning as compared to other courts. Retrieved: https://cdn.prod.elseone.nl/uploads/2020/03/Persbericht-data-analyse-rechtspraak-in-belastingzaken.pdf

Interview with Gerard Tangenberg, Senior Raadsheer, Gerechtshof The Hague (FaceTime, 25 March 2020)

[139] Interview with Jaap van den Herik, Professor Law and IT, Faculty of Science and Faculty of Law, Leiden University (Telephone Call, 30 March 2020)

and its overall applicability will strengthen, and eventually make JDSA a valuable secondary advisor for the judiciary.

4.2.1 Developing the Regulatory Framework for JDSA

Fundamentally, analogues judges are allowed to enforce the rule of law on the basis of the characteristics of being independent and impartial. For the deployment of ADM in the judiciary, the fundamental rights as enshrined in the ECHR, particularly Articles 6, 8, 14 and 17 thereof, should be adhered to at all times.[140] The more judge-related articles 40(1) and in particular 45 on the ECtHR provide more guidance on the requirements of robotic judgement through ADM.[141] Following these rights, the regulatory framework should aspire to deliver justice in an efficient, inexpensive yet expert-based manner.[142] As Susskind also sets out in line with decades of academic research, legal theory, judicial writing and policy thinking, there are seven main principles characterizing justice, applicable to both JDSA and analogue human judgement: [143]

- Substantive justice, concerned with fair decisions, in accordance with societal development and the concept of being *just*.
- Procedural justice, establishing fair judicial processes through the principles *audi alteram partem*[144], *nemo iudex in causa sua*[145], and addressing the facts and applicable laws rather than the parties.

---

[140] ECHR Article 6, 8, 14 and 17.

[141] ECHR Article 40: Public hearings and access to documents:
1. Hearings shall be in public unless the Court in exceptional circumstances decides otherwise.
ECHR Article 45: Reasons for judgments and decisions:
1. Reasons shall be given for judgments as well as for decisions declaring applications admissible or inadmissible.
2. If a judgment does not represent, in whole or in part, the unanimous opinion of the judges, any judge shall be entitled to deliver a separate opinion.

[142] Interview with Floris Bex, Scholar AI & Law, Universiteit Utrecht (Skype, 18 May 2020)

[143] Susskind (n 1) 73.

[144] "*All litigants should be given the opportunity to state and defend their cases in court*" Susskind (n 1) 77

[145] "*No one should be judge in a case in which they have an interested"* Susskind (n 1) 78. Considering the notion of bias based on past decisions, this principle might be especially applicable to robots"

- Open justice, requiring transparency in judicial procedures and required a judgement to be explainable, intelligible and appealable.
- Distributive justice, meaning the judiciary should be accessible and intelligible to all.
- Proportionate justice, requiring decision-making to be appropriately balanced, also in ensuring the costs of handling a case should be balanced with the nature and value of that case[146].
- Enforceable justice, allowing for the authority and enforceability of a judicial decision to be backed by the state[147]
- Sustainable justice: courts should be stable, secure, adequately funded, and aligned technologically with the societies they serve.[148]

Bearing in mind these standards of delivering justice, the acceptance criteria for ADM in the judiciary should establish a decision-making system that is controllable, transparent and testable. Especially in the early stages of deployment, algorithmic systems should be subjected to proportionality tests, which balances the benefits of their support with the liability in case of algorithmic mistakes, and the time and costs associated with improvement.[149] In terms of further proportionality, the mistake frequency and error-impact ratio, which will differ for the legal domain related to the cases to which JDSA is applied, should be evaluated by humans in

[146] Susskind (n 1) 82.

[147] Susskind (n 1) 83 *"Courts have the remarkable capacity to deprive people legitimately of their money, property, liberty"*.

[148] Susskind (n 1) 84.

[149] Directive (EU) 2018/958 of the European Parliament and of the Council of 28 June 2018 on a proportionality test before adoption of new regulation of professions PE/19/2018/REV/1; European Parliament, 'Draft Opinion on Artificial Intelligence: Questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice' [2020] (2020/2013(INI)). As Andreas Schwab (rapporteur for the opinion) writes: *"[I]t follows from Directive (EU) 2018/958 that humans must always bear ultimate responsibility for decision-making that involves risks to the achievement of public interest objectives*" whilst urging "*Member States to assess the risks related to AI-driven technologies before automating activities connected with the exercise of State authority, such as the proper administration of justice*" and "*consider the need to provide for safeguards, foreseen in Directive (EU) 2018/958, such as supervision by a qualified professional and rules on professional ethics*"

the iterative development process.[150] Based on this evaluation, JDSA can be deployed to the legal domains for which this risk factor is lowest, for example differentiating between tax and criminal law. To achieve this goal, models need to be auditable through validation with strict documentation, which should be made publicly available, therewith allowing open justice.

For JDSA, algorithmic accountability and liability can best be determined with methods generally applied within the regulatory framework of data protection law, combined with the suggestions provided in a recent motion for a European Parliament Resolution on a Civil Liability Regime for AI.[151] To identify the potential "harms" JDSA could inflict on those affected by and subjected to its functioning, fundamental rights impact assessments inspired by Data Protection Impact Assessment (DPIA) might prove a concrete example and valuable tool.[152] Since DPIA's have been effectively implemented in providing benchmarks for assessing ADM models at early design and development stages and are based on GDPR Article 25 on "*data protection by design*", fundamental rights enshrined in the ECHR and the Charter of Fundamental Rights of the European Union, a DPIA-driven regulatory approach would be capable of regulating the appropriate functioning of models and managing the potential risks effectively. [153] With the application of impact assessments on the potential violation of

[150] Addendum Dutch Second Chamber Letter on Risks as a result of using Governmental Data Analytics, 4; Interview with Mildo van Staden, Innovation Advisor, Ministry of the Interior and Kingdom Relations (Telephone Call, 20 April 2020); Johannes Bijlsma, Floris Bex & Gerben Meynen 'Artificiele Intelligentie en Risicotaxatie: Drie Kernvragen voor Strafrechtjuristen' [2019] Nederlands Juristenblad (44) 3-4

[151] European Committee on Legal Affairs, 'Motion for a European Parliament Resolution with recommendations to the Commission on a Civil Liablity Regime for Artificial Intelligence (2020/2014(INL)); European Parliament, 'Draft Report with recommendations to the commission on a civil liability regime for artificial intelligence (2020/2014(INL)) 14, 17 & 25.

[152] GDPR (n 18) Article 35(3) states:

[A] data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of: (a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; in accordance with this Article, fundamental rights that are within the meaning and objective of the ECHR that might be at risk can be evaluated before resulting in "harm".

A DPIA is a generally applicable privacy-related impact assessment with the prime objective of identifying and analysing how data privacy might be affected by a data controller's actions or activities.

[153] Consolidated versions of the Treaty on the Functioning of the European Union (TFEU) [2016] OJ C202/1. Specifically: *Articles 20 (equality), 21 (non-discrimination) and 47 (Right to an effective remedy and a fair trial)*; Heleen L. Janssen, 'An approach for a fundamental rights impact assessment to automated decision-making' [2020] International Data Privacy Law 76 9, 31, 15: "*Controllers envisaging ADM in their data processing have to comply*

fundamental rights combined with a regime of strict liability, data controllers, the judiciary in case of JDSA, are required to implement appropriate technical safeguards to ensure compliance with fundamental and more specific GDPR rights, such as *the right to an explanation*. In order to meet the minimal requirements of the *right to an explanation* as set out in Chapter 2 through continuously analysing the patterns identified by the used algorithms (e.g. with the techniques described in Section 3.3), relevant rights can be respected, and liability can be attributed appropriately.

4.2.2 Implementation Process and Framework Applicability

Considering the implementation process of JDSA, it should be evident that although legal innovation improves efficiency and decreases costs for judicial processes, governmental investment in the judiciary should enable the use of sufficient resources. To develop the costly and relatively advanced technology, the legal industry should learn from the FinTech industry, and could benefit from the implementation of a regulatory sandbox.[154] This innovation-fostering strategy allows private and public companies to conduct experimentation under the supervision of the appropriate regulator (e.g. the CEPEJ), who should regulate effectively and proportionally without the long process of adopting new laws, whilst allowing overall development without the risk of fines and requirements for high compensation due to liability.[155]

To provide an example of applying the regulatory framework of explainability, described in Chapter 2, to the algorithms described in Chapter 3, an example of an

with a natural person's fundamental rights as well. The wording of Article 35(7)(c) GDPR, stipulating that an assessment of the risks to the rights and freedoms of data subject need to at least be contained in such assessment, endorses this approach."; Interview with Mildo van Staden, Innovation Advisor, Ministry of the Interior and Kingdom Relations (Telephone Call, 20 April 2020)

[154] Jorge Gabriel Jimenez & Margaret Hagan, 'A Regulatory Sandbox for the Industry of Law', [2019] Thomson Reuters Legal Executive Institute.

[155] Ibid. 'The sandbox enables a safe environment for business to test services or products without the risk of being sued for the unauthorized practice of law.'

evaluation of a JDSA could be attained through analysing the requirements of human explainability in the following manner.[156] An appreciated and established model for legal analytics is the relatively explainable SVM algorithm, which achieves high performance in classifying similar cases.[157] In line with the aforementioned legal framework, the main requirements that are to be considered for evaluating an SVM would for example be:

- Reasoning capacity: is the applied model capable of generating argumentation structures, or of showing argumentation in relevant past judgements?
- Previous deployment: have academically developed models demonstrating promising performance for the usage of the model in a legal context?
- Overfitting risk: what is the risk of overfitting on past data?[158]
- Possibility for expert feedback: can the model be adjusted to incorporate expert user feedback?
- Training data volume: what is the amount of data needed to train the model?

On overview of the results of evaluating these requirements would result in an evaluation such as in Table 1:

[156] In light of Directive (EU) 2018/958 of the European Parliament and of the Council of 28 June 2018 on a proportionality test before adoption of new regulation of professions PE/19/2018/REV/1, and the proportionality test that courts often conduct for balancing fundamental rights, an exemplification of applying the minimum explainability requirements could also be set out in light of so-called low-value Mulder cases (e.g. parking tickets), contrasted with the high-stake recent Dutch System Risk Indication (SyRI) decision.

Whereas the SyRI decision clearly states that the applied system was insufficiently transparent and verifiable, thereby failing to comply with Article 8(2) of the ECHR, the court decided that the SyRI legislation does not strike a fair balance, as required under the ECHR, warranting a sufficiently justified violation of private life. As such, the SyRI legislation was ruled unlawful, because it violates higher law and, as a result, has been declared as having no binding effect.

Nederlands Juristen Comité voor de Mensenrechten v The State of the Netherlands [2020] C/09/550982/HA ZA 18-388.

In contrast with the failed proportionality test regarding SyRI, the algorithmic support as introduced in Narayan's thesis, a functional decision-support system would likely be proportionate to allow automation and efficiency in the paralegal case-processing workflow, because it merely concerns the bulk of low-value cases and does not require judicial decision-making with substantial legal effect.

Narayan Nitin 'A decision support system for the Court of East Brabant' (Professional Doctorate in Engineering, Jheronimus Academy of Data Science 2019) 39.

[157] Virtucio (n 9) 5; Aletras and others (n 7) 12.

[158] Floris Bex & Henry Prakken 'De Juridische Voorspelindustrie: onzinnige hype of nuttige ontwikkeling?' [2020] Ars aequi, 69, 255-259. 256 Overfitting could sustain past bias and cause the model not to be applicable for use cases outside the scope of the training set.

Algorithm	***Judicial Argument Generating Capacity***	***Previously applied for judging capacity***	***Risk of overfitting***	***Training Data Volume***	***Legal Expert Feedback Possibility***
Rule-based (most CMLA's)	No	Yes	-	High	Yes
Decision Trees	No	Yes	High	Low	Yes
Ensemble models (e.g. Random Forest)	No	Yes	High	Low	Yes/No
kNN (non-linear)	Yes	No	Medium	High	Yes
SVM (linear)	Yes	Yes	Low	Low	Yes
BERT (neural network)	Yes	Yes	Low	Extremely High	No
RNN – LSTM (neural network)	Yes	No	Low	Extremely High	No

Table 1: Overview of evaluation requirements for the deployment of judicial models, inspired by the work of van Amelsfoort (n 155) 19 and Hacker and others (n 15) 19

Based on the framework as applied in Table 1, an SVM legal classification model has a high potential of judgement support based on the combination of explainable visualisation for multiple dimensions allowing reasoning, previous scholarly and practical examples, low risk of overfitting, high performance on small datasets for specific domains, and the possibility of improvement with adjusting weights based on expert feedback.[159] Overall, an SVM model would thus be a suitable JDSA, especially because it can be explained how input features relate to the classification of relevant cases, which was the overall purpose of the proposed of explainability framework from Chapter 2. [160]

[159] Aletras and others (n 7); Corbin van Amelsvoort, 'Predicting judicial decisions in Dutch tax law by Natural Language Processing and Machine Learning' (Master's thesis, Open Universiteit 2019) Especially for differentiating types of text, an SVM can prove its benefits.

[160] Section 2.3. An SVM would potentially pass the minimum requirements of explainability, given that based on the determining the relevance of rulings in cases, argumentation structures could be determined, and the final legally binding decisions remains with the presiding analogue judge.

4.3. Preliminary Conclusion Research Sub-Question III: Requirements for Implementing JDSA

In line with the principle of transparent and open justice and in allowing the data-driven implementation of JDSA, judicial and regulatory authorities should require all judgements from all courts to be made available for legal analytics and the development of ML algorithms. Having considered the aspects of explainable (Chapter 2) capabilities of ML and NLP models (Chapter 3), the main requirements for implementing JDSA can be summarised as to account for four main factors. First, it should be noted that JDSA should first be applied to the high-volume of relatively easy legal matters submitted for judgement. Second, JDSA should be deployed for those use cases that can be better performed with well-trained models and computational power. Such use cases include performing the relatively easy tasks of case law retrieval, which will arguably eventually outperform humans in terms of costs and efficiency and will constitute a first case for building trust in JDSA. Third, the seven principles of justice should constitute the main standard that is to be achieved at all times, therewith also establishing generally applicable regulations capable of enduring the quickly shifting technical landscape. Fourth and finally, the framework proposed by academic literature, as enriched by the suggestions in Sub-Section 4.2.2, constitutes a step in the direction of practical implementation.

Conclusion

The objective of this thesis has been to answer the question:

"Are recently developed models, based on ML and NLP, capable of implementing robotic judgement based on reason-generating and explainable JDSA?"

Conclusively, in light of the regulatory requirements of explainability, the answer to this question is that current self-learning models are unable to implement reasoned and explainable robotic judgement through JDSA when concerned with complex legal cases. Both simple and more complex NLP and ML models thus far remain unable to meet the minimum requirements of explainability and reasoning, with research into the functioning of the SVM algorithm as a notable exception. The established minimum requirements of demonstrating the relations between a model's input and output in an intelligible manner that elucidates and reasons why weights are determined have often not been met, rendering analogue judges unable to currently apply algorithmic support in an intelligible manner. The scholarly field thus agrees that robotic judgement based on decisions with reasons will not be implemented in the foreseeable future (i.e. five years).

However, in arguing for the advantages and requirements of implementing JDSA, therewith allowing the first generation of robotic judgement, the rapid advances in the field of ML and NLP have demonstrated the potential of improving adjudicational processes in the future. With the implementation of the proposed innovative regulatory framework for the technical development of JDSA, considerable benefits, including increased access to justice, the removal of adjudicational backlogs, identifying past unfairness and partiality and overall efficiency, can be attained. A main challenge to developing such algorithms however exists in the availability of training data, since a mere fraction of case outcomes and rulings is currently published. To implement JDSA, models should be trained on large corpora, which generally results in higher model performance, and will allow the development of a

representative model based on training data that accounts for all previously ruled cases, therewith addressing the challenges described in Section 3.1. Based on the reasoning patterns and rulings of a representative input of previous cases, the bulk of simpler cases could particularly benefit from JDSA whilst ensuring that the minimum requirements of model explainability are adhered to.

The regulatory requirements regarding explainability should require models to be explainable by design, taking into account the reasoning behind the relations between input and output for specific features and case-specific facts. It is particularly noteworthy that relatively simple classification models have demonstrated great potential for modelling patterns of precedential relevance, and that NLP argument-mining models can support reasoning in the judiciary, especially for the bulk of easier legal matters and smaller claims procedures. For the implementation of JDSA, fundamental conceptions of justice should be combined with model-specific requirements, which can substantially increase the quality of judgement standards. If implemented in line with the GDPR's right to an explanation and the suggested statutory framework of liability and interpretability, the future of adjudicating cases could benefit from the insights, considerations and suggestions from this thesis. As a next step towards the future of the legal industry, judges, policymakers and regulatory authorities shall thus hopefully acknowledge the potential of robotic judgement through JDSA.

Broader Relevance and Future Research

As decades of research in the combination of AI and Law has demonstrated, there are many potential applications and use cases of improving justice systems with technological tools. Previous research has largely focused on either the legal, or the technical implications of technological development, whereas this thesis has attempted to bridge the disciplines, and provide valuable insights from both the legal and the technical perspectives. For the specific development of decision-support systems, the minimum requirements of the proposed regulatory framework of JDSA can be applied to many newly developed models. There however remains a lot of technical and mathematical depth in the functioning of the models used throughout this thesis that should be considered in future research. A promising research gap for future technical research, as advised by Professor van den Herik, would be the application of zero-shot learning to the judicial domain. As the field of applied ML and NLP will however continue to advance, the broader relevance of this research exists in demonstrating how to construct, in particular with the suggestion for a regulatory sandbox, the legal framework that allows ascertaining the desired improvements in the judiciary based on principles, such as accountability and explainability and the seven principles of justice described in Sub-Section 4.2.1, rather than on specific statutory regulation.

Since interpreting the regulatory landscape will remain an endeavour tackled through human reasoning, future research in ML and NLP techniques should attempt to develop reasoning models and compare them with the actual performance of rule-based good old-fashioned AI and Law. Future research could thus for example include: can an input of a set of specific facts and a desired outcome lead to a computational output that shows the probability of achieving the desired outcome, reasoned and explained on a ML basis, with NLP-mined legally sound argumentation? Besides, xAI will remain a field in which many

model-specific questions remain unanswered, particularly in the legal domain under the GDPR.

Overall, future research should however consider that, before the Covid-19 pandemic, many of the innovations of today seemed years away from actual implementation, with the use of online court proceedings as the main example. Given the circumstances of the pandemic resulting in online judgement and the necessity of rapid adjudicational processes, further legal innovation through the application of ML and NLP should be prioritized by policymakers to reach an affirmative answer to this thesis's research question.

On a final note, the recent remote functioning of the judiciary has however reiterated the value of a physical analogue court proceedings, particularly in complex legal matter dealing with cases involving a substantial degree of "hurt", since many of such cases have not been appropriately addressed in the last few months. Future research in the value of physical appearances in court rooms, in combination with the societal acceptance of remote proceedings, should take into consideration the available technological means, whilst ensuring the highest attainable quality of adjudication is maintained. A final question that arises should thus be how to adjust court proceedings to remain functioning as intended in the best possible way, whilst ensuring innovation and developing JDSA?

Bibliography

Primary Sources

Statutory Law

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) [1950]

Directive (EU) 2018/958 of the European Parliament and of the Council of 28 June 2018 on a proportionality test before adoption of new regulation of professions PE/19/2018/REV/1

Regulation (EU) 2015/2421 of the European Parliament and of the Council of 16 December 2015 amending Regulation (EC) No 861/2007 establishing a European Small Claims Procedure and Regulation (EC) No 1896/2006 creating a European order for payment procedure' [2016]

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, OJ 2016 L 119, 1) [2018]

Cases

Donoghue v Stevenson [1932] UKHL 100

Google Spain SL v. Agencia Española de Protección de Datos (AEPD) (Google v. Spain) [2014] CJEU C131/12

Nederlands Juristen Comité voor de Mensenrechten v The State of the Netherlands (SyRI-wetgeving Procedure) [2020] C/09/550982/HA ZA 18-388

Legislative Instruments

Addendum Dutch Second Chamber Letter on Risks as a result of using Governmental Data Analytics [2019] https://www.rijksoverheid.nl/documenten/rapporten/2019/10/08/tk-bijlage-over-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid Accessed 30 March 2020

Consolidated versions of the Treaty on the Functioning of the European Union (TFEU) [2016] OJ C202/1.

European Commission, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' [2018]

European Commission, 'White Paper on Artificial Intelligence: A European approach to excellence and trust' [2020]

European Commission for the Efficiency of Justice (CEPEJ), 'European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment' [2018]

European Committee on Legal Affairs, 'Motion for a European Parliament Resolution with recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence' [2020] (2020/2014(INL))

European Parliament, 'Draft Report with recommendations to the commission on a civil liability regime for artificial intelligence [2020] (2020/2014(INL))

European Parliament, 'Draft Opinion on Artificial Intelligence: Questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice' [2020] (2020/2013(INI))

Working Party 29, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' [2018]

Secondary Sources

Books

Ashley K D, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, (1st edn, Cambridge University Press 2017)

Susskind R, Online Courts and the Future of Justice, (1st edn, Oxford University Press 2019)

Academic Articles

Aletras N and others, 'Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective' [2016] PeerJ Computer Science 2

Araujo T and others, 'In AI we trust? Perceptions about automated decision-making by artificial intelligence' [2020] AI & Society 1

Beatson J, 'AI-Supported Adjudicators: Should Artificial Intelligence Have a Role in Tribunal Adjudication?' [2018] Canadian Journal of Administrative Law & Practice 307

Bench-Capon T, 'Arguing with cases' [1997] JURIX 85

Bench-Capon T, 'The Need for Good Old-Fashioned AI and Law' [2020]

Bex F & Prakken H 'De Juridische Voorspelindustrie: onzinnige hype of nuttige ontwikkeling?' [2020] Ars aequi, 69, 255-259

Bijlsma J, Bex F & Meynen G 'Artificiele Intelligentie en Risicotaxatie: Drie Kernvragen voor Strafrechtjuristen' [2019] Nederlands Juristenblad (44)

Brkan M, '"Do algorithms rule the world?" Algorithmic decision-making and data protection in the framework of the GDPR and beyond.' [2019] International Journal of Law and Information Technology 91

Casey B, Farhangi A & Vogl R, 'Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise' [2019] Berkeley Tech 143

Chalkidis I, Androutsopoulos I & Aletras N, 'Neural Legal Judgment Prediction in English' [2019] arXiv

Chen D L, 'Machine Learning and the Rule of Law' [2019] Computational Analysis of Law, Santa Fe Institute Press 1

Coglianese C & Lehr D, 'Regulating by robot: Administrative decision making in the machine-learning era.' [2016] The Georgetown Law Journal 105 1147

Coglianese C & Lehr D, 'Transparency and algorithmic governance' [2019] Administrative Law Review 1

Deeks A, 'The Judicial Demand for Explainable Artificial Intelligence' [2019] Columbia Law Review 1829

Dymitruk M, 'The Right to a Fair Trial in Automated Civil Proceedings' [2019] Masaryk UJL & Tech 27

Edwards L & Veale M, 'Enslaving the algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' [2018] IEEE Security & Privacy 46

Edwards L & Veale M, 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for' [2017] Duke Law & Technology Review 18

Elwany E, Moore D & Oberoi G, 'BERT Goes to Law School: Quantifying the Competitive Advantage of Access to Large Legal Corpora in Contract Understanding' [2019] arXiv

Goodman B & Flaxman S, 'European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"' [2017] AI magazine 50

Guidotti R and others, 'A Survey of Methods for Explaining Black Box Models' [2018] ACM computing surveys 1

Hacker P and others, 'Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges' [2020] Artificial Intelligence and Law

Harris A, 'Judicial Decision Making and Computers' [1967] Villanova Law Review 272

Van den Herik J, 'Kunnen Computers Rechtspreken?" [1991] Gouda Quint.

Janssen H L, 'An approach for a fundamental rights impact assessment to automated decision-making' [2020] International Data Privacy Law 76

Jimenez J G & Hagan M, 'A Regulatory Sandbox for the Industry of Law' [2019] Thomson Reuters Legal Executive Institute

Jongbloed T A W and others, 'The Rise of the Robotic Judge in Modern Court Proceedings' [2015] International Conference on Information Technology 59

Kaminski M E, 'The right to explanation, explained' [2019] Berkeley Technology Law Journal, 189

Katz D M and others, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' [2014] 12(4) PLoS ONE 1

Krupansky J, 'Untangling the Definitions of Artificial Intelligence, Machine Intelligence, and Machine Learning' [2017] https://medium.com/@jackkrupansky/untangling-the-definitions-of-artificial-intelligence-machine-intelligence-and-machine-learning-7244882f04c7

Lawlor R C, 'Excerpts from Fact Content of Cases and Precedent - A Modern Theory of Precedent' [1971] Jurimetrics Journal 245

Long S and others, 'Automatic Judgement Prediction via Legal Reading Comprehension' [2019] Springer 558-572

Lundberg S M & Lee S I, 'A Unified Approach to Interpreting Model Predictions' [2017] In Advances in neural information processing systems 4765

Marques M R S and others, 'Machine learning for explaining and ranking the most influential matters of law' [2019] In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law 239

Medvedeva M, Vols M & Wieling M, 'Using machine learning to predict decisions of the European Court of Human Rights' [2019] Artificial Intelligence and Law 1

Miller T, 'Explanation in artificial intelligence: Insights from the social sciences' [2019] (267) Elsevier Artificial Intelligence 1

Mittelstadt B, Russell R & Wachter S, 'Explaining explanations in AI' [2019] Proceedings of the conference on Fairness, Accountability and Transparency 279

Molnar C, 'Interpretable machine learning. A Guide for Making Black Box Models Explainable' [2020] (10) https://christophm.github.io/interpretable-ml-book/

Morison J & Harkens A, 'Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making' [2019] 39(4) Legal Studies 618

Philipsen S & Themeli E, 'Een introductie op de robotrechter' [2019] Rechtstreeks 2019(2) 46

Prakken H, 'Komt de robotrechter er aan?' [2018] 2018(4) Nederlands juristenblad 269

Ross A S, Hughes M C & Doshi-Velez F, 'Right for the right reasons: Training differentiable models by constraining their explanations' [2017] arXiv

Rudin C, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.' [2019] 1(5) Nature Machine Intelligence 233

Selbst A D & Powles J, 'Meaningful information and the right to explanation' [2017] 7(4) International Data Privacy Law, 233

Sourdin T, 'Judge v. Robot: Artificial Intelligence and Judicial Decision-Making' [2018] 41 UNSW Law Journal 1114

Surden H, 'Artificial Intelligence and Law: An Overview' [2019] Georgia State University Law Review 35

Van Eck M, 'Computerbesluiten, kunstmatige intelligentie en de bestuursrechter' [2019] OpenRecht

Veale M, Binns R & Van Kleek M, 'Some HCI Priorities for GDPR-Compliant Machine Learning' [2018] arXiv

Verhulp E & Rietveld R, 'Hoe expertsystemen de rechtspraak kunnen helpen' [2019] Rechtstreeks 2019 nr 2 39

Virtucio M B L and others, 'Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning' [2018] IEEE 42nd Annual Computer Software and Applications Conference 76

Wachter S, Mittelstadt B & Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law 76

Xiang A & Raji ID, 'On the Legal Compatibility of Fairness Definitions' [2019] arXiv

Yang W and others, 'Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network' [2019] arXiv

Yu R & Ali G S, 'What's Inside the Black Box? AI Challenges for Lawyers and Researchers' [2019] Cambridge University Press 2

Theses

Narayan N. 'A decision support system for the Court of East Brabant' (Professional Doctorate in Engineering, Jheronimus Academy of Data Science 2019)

Van Amelsvoort C, 'Predicting judicial decisions in Dutch tax law by Natural Language Processing and Machine Learning' (Master's thesis, Open Universiteit 2019)