AI APPROACHES TO PREDICTIVE JUSTICE: A CRITICAL ASSESSMENT

Federico Galli[1]
Giovanni Sartor[2*]

Abstract
This paper addresses the domain of predictive justice, exploring the intersection of artificial intelligence (AI) and judicial decision-making. We will first introduce the concept of predictive justice, referring to the ongoing debate surrounding the potential automation of judicial decisions through AI systems. Then, we will examine the current landscape of AI approaches employed in predictive justice applications, providing a comprehensive overview of methodologies and technological advancements. Then, we delve into the phenomenology of predictive justice, highlighting the diverse spectrum of legal predictions achievable with contemporary AI systems. We also assess the extent to which these predictive AI systems are presently integrated into real-world judicial practices. Finally, the paper critically addresses recurrent fears and critiques associated with predictive justice. We sort these critiques into unreasonable objections, reasonable concerns with possible technical solutions, and reasonable concerns demanding further investigation. Navigating the complexities of these critiques, we offer some insights for future research and practical implementation. The nuanced approach taken in this study contributes to the ongoing discourse on predictive justice, emphasising the need for a balanced evaluation of its potential benefits and legal challenges.
Keywords
Predictive justice, Artificial Intelligence, Automated judicial decisions, Critiques to Predictive Justice, Black Box.

[1] Lecturer of Legal Informatics, Cirsfid-ALMA AI, Law Department of Bologna University – Italy.
[2] Full Professor of Legal Informatics, Cirsfid-ALMA AI, Law Department of Bologna University, and European University Institute – Italy.
[*] While the work is the product of a joint reflection, Section 2 should be attributed to Giovanni Sartor Section 3 to Federico Galli. Section 1, 4 and 5 are shared. Both Authors have been supported by the H2020 European Research Council (ERC) Project CompuLaw under the European Union's Horizon 2020 research and innovation programme (Grant Agreement no. 833647) and by the European Union's Justice Programme (2014-2020) Project ADELE: Analytics for DEcision of LEgal cases (Grant Agreement no. 101007420).

of code. - 4.3. Hard effect. - 4.4. Data. - 4.5. Black box. - 4.6. Intellectual property. - 4.7. Status quo effect. - 5. Conclusion: Towards cognitive-enhancing predictive justice.

1. PREDICTIVE JUSTICE: BRIDGING THE GAP BETWEEN HYPE AND REALITY

In recent years, "predictive justice" has become a prominent topic of discussion, not only within legal academia[3] but also in broader public discourse[4].

The increasing dialogue surrounding predictive justice is driven by rising expectations of its transformative potential within legal systems. Supporters of predictive justice claim its ability to enhance efficiency, boost legal certainty, and usher in a new era of judicial effectiveness. This concept holds the promise of streamlining legal processes, reducing case backlogs, and promoting greater uniformity and predictability in legal outcomes. These prospects have captured the attention not only of legal practitioners and lawyers, but also policymakers who aspire to enhance the certainty and efficiency of the judicial apparatus.

---

[3] Among internationally well-known scholars, we can refer, among others, to PASQUALE, CASHWELL 2018; HILDEBRANDT 2018; BEX, PRAKKEN 2021a; BEX, PRAKKEN 2021b; MEDVEDEVA, WIELING, VOLS 2023. In the Italian legal doctrine, among the many important contributions by distinguished Authors on this topic, we can refer to the contributions by FERRARI 2023; PAJNO 2022; SANTOSUOSSO, SARTOR 2022; BICHI 2020, CASTELLI, PIANA 2019.

[4] In Italy, for example, see the article by Gustavo Ghidini, Daniele Manca, Intelligenza artificiale: dubbi (e rischi) in Europa, Corriere della sera, 13 dicembre 2021, <https://www.corriere.it/opinioni/21_dicembre_13/intelligenza-artificialedubbi-rischi-europa>. In international newspapers, we can refer to Pranshu Verma, The never-ending quest to predict crime using AI, in Wahsington Post 15 July 2022. https://www.washingtonpost.com/technology/2022/07/15/predictive-policing-algorithms-fail/ and to Jack Newman, China develops AI 'prosecutor', in Mail Online 27 December 2021 < https://www.dailymail.co.uk/news/article-10346933/China-develops-AI-prosecutor-press-charges-97-accuracy.html>.

Opposed to this wave of optimism, apprehension and critiques have emerged. The integration of predictive algorithms into judicial processes has raised multiple concerns regarding the potential emergence of "robot justice"[5] and "robot judges"[6]. These fears paint a scenario in which judges and legal professionals delegate their work and duties to infallible machines, thus relegating humans to a subordinate position within the legal world governed by algorithms.

While the discourse oscillates between overly optimism and deep scepticism, governments worldwide are increasingly recognising the potential of predictive justice technologies. This growing interest is manifesting through increased governmental funding, initiatives, and projects aimed at exploring the multifaceted applications of predictive analytics, machine learning, and artificial intelligence within the realm of justice[7]. These initiatives span a wide spectrum, encompassing predictive policing, risk assessment in sentencing, case outcome forecasting, and the optimisation of court resource allocation.

Against this background, the objective of this paper is to navigate the evolving landscape of predictive justice. Our aim is to contribute to a nuanced understanding of this development, delineating its potential, limitations, and the legal concerns it raises.

Section 2 will look at the current approach in AI to develop predictive justice applications. After that, in Section 3, we will address the

[5] ZEKOS 2022.

[6] See MORISON, HARKENS 2019, among the many using these terms to anticipate future risks of AI in the judicial sector.

[7] For example, the use of AI in the judiciary was set as one of the proprieties in the 2019-2023 Action Plan on European e-Justice by the European Union. An annual call for project is published by the Justice Program to fund projects developing AI tools for the judiciary. In Italy, the recent PNRR funding program has included significant funding to digitalise justice, including the mandatory electronic management of all documents and fully telematic civil proceedings, as well as the use of artificial intelligence to automatise certain tasks. On these initiatives, see MARTORANA, SAVELLA 2021.

phenomenology of predictive justice by highlighting the different kinds of legal prediction that can be realised with current AI systems and the extent to which these systems are currently used in practice today. Finally, Section 4 will address some recurrent fears and critiques of predictive justice. We will provide arguments to discern from critiques that are unreasonable, those that are reasonable but can be easily addressed, and those that are reasonable and should be considered in further investigation.

2. AI APPROACHES TO PREDICTIVE JUSTICE

Prediction has always played a crucial and enduring role in the legal system. It involves foreseeing legal outcomes based on judicial and administrative precedents, legislative statutes, and legal principles. This predictive ability is a fundamental skill used by legal practitioners, judges, and scholars, as it helps anticipate legal decisions, assess the likelihood of favourable judgements, and evaluate the potential impact of legal arguments.

In the pursuit of automating reasoning and legal functions, constant efforts have been made toward predictive capabilities, which may assist legal professionals in assessing the prospects of new cases[8]. The limitations of traditional logic-based approaches, combined with limited computing power and data availability, have stood in the way of successful automation of legal forecasting. However, recent advancements in AI, particularly linked to machine learning and natural language processing techniques, have rekindled the possibility of a transformative shift in legal predictions. AI systems are today capable of processing vast amounts of legal data, identifying complex patterns, and generating predictions, often matching or even surpassing human accuracy.

[8] ASHLEY 2019.

Before moving on to predictive justice practices, in this section, we delve into the technologies that have been and are currently used in realising legal prediction machines.

2.1. Knowledge-based methods

Until a few decades ago, the general assumption was that creating an intelligent system required humans to provide a formal representation of relevant knowledge. This representation typically combines rules and concepts, along with algorithms that can draw inferences from this knowledge.

This same approach was prevalent in the early days of AI and law when scientists were trying to build legal expert systems[9]. These systems are provided with legal knowledge (such as legal rules, concepts and facts), represented in a computable language, and an inferential engine capable of applying the knowledge base to infer new knowledge. Different logical formalisms have been developed to represent legal knowledge (rule languages, classical logic, modal and descriptive logics, formal argumentation, etc.) and computable models for inferential processes (deductive, defeasible, inductive, probabilistic, case-based, etc.)

Legal expert systems represent the most prominent example of the s.c. symbolic approach in AI. These systems operate by engaging in reasoning exclusively through the manipulation of symbols, which are essentially abstract representations of objects, concepts, or information. Notably, these systems disregard the content or meaning behind these symbols, which is not captured by logical structures. This mode of reasoning is a purely formal process, often likened to what G.W. Leibniz would call "blind reasoning"[10].

[9] SUSSKIND 1987.
[10] LEIBNIZ 1849.

In this kind of reasoning, symbols are manipulated without any awareness of their underlying meaning.

Over the years, two main types of legal expert systems have been developed: systems based on rules and concepts, and those based on cases.

Rule-based systems have predominantly been used in administrative domains like social security and taxation[11]. These domains have complex networks of rules, each with specific meanings, typically governing well-defined and uncontested cases. For example, consider rules determining eligibility conditions for social security benefits. These rules might include criteria like age, family status, income, and assets, possibly detailed in other rules. In such contexts, rule-based systems excel because they can meticulously apply all the relevant rules to specific cases, provided these rules are accurately formulated using the language embedded within them[12].

It is important to note that, in many domains, rule-based systems can provide suitable solutions for the majority of cases. A noteworthy example of a successful system is Oracle Policy Automation (currently known as Oracle Intelligent Advisor)[13], which finds applications in various domains like immigration, social benefits, and taxation across countries like Australia, the United Kingdom, the United States, and France. In Italy, the ReMida software[14], developed by Gianfranco D'Aietti, has also achieved notable success.

We can imagine a knowledge-based system also functioning as a predictive tool. If the system is provided with all the attributes and characteristics of a new case encoded as rules, it would independently derive conclusions by

---

[11] CONTISSA 2015.

[12] For example, one of the most successful endeavours in building legal knowledge-based systems is provided by the TAXMAN I and II system by MCCARTY 1990.

[13] The Oracle Environment is available at the following link https://www.oracle.com/applications/oracle- policy-automation/index.html.

[14] The ReMida Family system is published by ReMida Editrice Giuridica (https://www.remidafamiglia.com/chi-siamo/autore), which distributes licences for its use.

applying those rules to the case. Essentially, this conclusion can be seen as a prediction of a future decision that a relevant decision-maker would make. For instance, a rule-based system could help someone estimate the likelihood of a successful claim for social security benefits. The accuracy of this prediction will depend on the ability of the legal knowledge engineer, who created the rules and concepts of the system, to anticipate how a competent decision-maker would interpret the law and evaluate the facts.

In judicial proceedings, where facts, concepts, or rules are typically contentious, rule-based systems face huge challenges. Judicial decisions often involve nuances, contextual interpretations, and considerations of equity and justice that cannot be encapsulated in rigid rules. Laws and legal precedents can evolve, and new case-specific information may emerge during proceedings, which cannot be taken into account by static rule-based systems. Rule-based systems also struggle when a certain degree of discretion is needed to consider factual scenarios in light of legal principles and political objectives. In essence, the limits of such systems are those of a "mechanical jurisprudence"[15], i.e., they can only address cases through the application of predetermined rules.

The second class of legal expert systems is based on judicial cases rather than rules. Instead of relying on predefined rules and logic, case-based systems draw their knowledge and decision-making capabilities from a repository of judicial cases. The system is designed to replicate legal reasoning by leveraging past case histories and their associated outcomes. An example of a case-based system is the HYPO system used in trade secret infringement cases[16]. This system's knowledge base collects several

---

15 We recall the famous paper by POUND (1908), where he argued that American common law or judge-made law had become sterile, unable to adapt to changing social and economic conditions, thus resulting in closed system of many archaic rules that judges and lawyers deducted from general "conceptions" and applied mechanically to the actual situations before them.

16 RISSLAND-ASHLEY 1987.

precedents, each of which is described or annotated with its outcome and a set of factors in favour of and against that outcome. For instance, if the defendant was aware of the plaintiff's activities and certain conditions were met, that would support a trade secret infringement conclusion. Conversely, other factors would support a no-infringement conclusion, like the plaintiff's communication during negotiations or information obtainable through reverse engineering. With these factors, the system uses analogical reasoning to predict the possible outcome of a new case. Essentially, it compares the factors of new cases to precedents where similar factors led to certain outcomes and forecasts the related outcome for the new case.

A case-based system is inherently aimed at predicting future cases: by drawing on a database of past cases and employing case-based reasoning techniques, it can anticipate how future cases will be decided using analogy[17]. At the same time, this task is particularly challenging, not unlike the creation of a rule-based system. First, consistently assigning all relevant factors to a large set of cases is a time-consuming process and can easily reflect the biases of the experts doing the task. Additionally, the case representation may not capture all aspects that influence decisions. To overcome these limitations, natural language processing (NLP) technologies are being experimented with the aim of automatically assigning factors to cases[18].

In conclusion, expert systems relying on human representations of knowledge, whether through rules and concepts or case-related factors, did not significantly succeed in predicting court decisions. They require translating the complexity of legal issues and the nuances of legal knowledge into a formalised representation of rules and concepts. This process is difficult and produces partial results that reflect the choices and interpretations of the person doing the formalisation and the time at which it was done. Legal reasoning involves applying rules and concepts but also

---

17 BRÜNINGHAUS-ASHLEY 2003.

18 ASHLEY-BRÜNINGHAUS 2009 and BRANTING *et al.* 2021.

requires the ability to find appropriate solutions, grasp analogies between cases, and consider social and human aspects of applying the law. Rule-based systems can be seen more as tools for understanding and applying the law, assuming that the law is viewed through the perspective of those who created the system's knowledge base.

2.1. Machine learning methods

Research in AI has made a great leap forward when it started to focus on learning implicit knowledge from large masses of data. In machine learning, the knowledge used by the system is no longer provided by the expert but rather is inferred by the system itself based on the data it has access to[19]. This direction has led to a large number of successful applications in many areas, from machine translation, industrial optimisation, marketing, robotic vision, motion control, and more.

We can identify three main approaches in machine learning, which may be seen as different ways of learning itself[20].

The first, which is currently the most commonly used, is supervised learning. In supervised learning, the system learns through supervision, i.e., through an instructor who provides the system with a large set of examples (s.c. training set) containing correct solutions to particular cases. Based on input pairs, the machine learning systems infer a general model that relates the input's different features to the possible output and apply that model to new instances, partially different from those in the training set.

More precisely, the system receives a set of pairs, each of which links the description of a problem with the correct answer to it. For example, a

---

[19] The idea of machine learning was anticipated by TURING (1950). Turing imagines a machine capable of learning can operate in ways that its creators and trainers did not anticipate, even without them knowing the details of the machine's inner workings.

[20] For an introduction to machine learning, see ALPAYDIN 2020.

system designed to recognise objects (e.g., to classify animals) may be provided with an image (the problem) and the label of that image, e.g. dog, cat, rabbit (the solution to the problem). Similarly, in an automated recruiting system, the instructor will provide a description of past candidates (age, experience, studies, etc.) linked to the outcome of the selection process (or to the assessment of the applicant's job performance) if the applicant has been hired. Similarly, in e-commerce recommendation systems, past consumers' information is associated with the objects purchased by them, etc.

In predictive justice, the examples would consist of past cases, and each example would be associated with the description of a case and the response to it, i.e., the judge's decision in that case. The purpose of systems intended to "predict" the decision of judges is thus not to provide the correct decision of the case but rather to anticipate the decisions that judges might, in fact, take.

Reinforcement learning is similar to supervised learning, involving training through examples. However, in this case, the system learns from the outcomes of its own actions, i.e., from the rewards or penalties (points earned or lost) associated with those actions' results. For example, in the case of a system designed to play a game (such as chess), rewards may be linked to victories and penalties to defeats; in a system learning to trade in the stock market, rewards can be linked to gains achieved and penalties to losses; in a system learning to send targeted advertising messages, rewards can be linked to user clicks. In any case, the reinforcement learning system will observe the results of its actions and self-administer the appropriate rewards and penalties. The system will learn to take actions most likely to lead to outcomes associated with rewards (wins, gains, clicks) and avoid actions most likely to lead to outcomes associated with penalties (defeats, losses, no clicks).

In the field of predictive justice, reinforcement learning approaches are hardly ever adopted. This is related to the fact that the programmer should set an objective of the system in advance so as to evaluate the system's decisions as positive and negative in terms of how they are able to maximise that objective. However, the objectives of justice cannot be precisely defined in advance: those objectives themselves and their relative importance are controversial. If the objective being pursued just consists of accurately anticipating future judicial decisions, then the system will just learn to mimic what judges on average do. If it consists of achieving some further social outcomes (e.g., diminishing crime), adverse side effects should also be considered (e.g., impacts on incarceration on offenders and their families). Moreover, defining positive and negative outcomes of judicial decisions requires considering not only the immediate case disposition but also the long-term impacts on individuals and society.

Finally, in unsupervised learning, the system learns without receiving instructions from external sources (supervised learning) or from the results of its own activities (reinforcement learning). Unsupervised learning techniques are particularly used for clustering, which involves grouping sets of objects that exhibit relevant similarities or connections (documents related to the same objects or issues, people with similar characteristics, words with similar meanings or functions).

For example, in an investigation, it can be useful to identify clusters in available electronic documents to pinpoint those related to the case at hand. Or, in a study of past criminal judgements aimed at adopting new criminal policies, the system could be asked to gather similar cases and examine, for instance, the connections between certain crimes and the use of drugs or weapons. Therefore, unsupervised learning can be employed, based on the techniques currently available, only for auxiliary tasks related to the goal of predicting judicial decisions.

A most recent approach in machine learning is to use pre-trained large models, particularly in the context of natural language processing[21]. The training of these large language models involves exposing them to a huge dataset containing vast amounts of text from diverse sources such as books, articles, websites, and other textual materials. The model learns to predict

[21] CHALKIDIS *et al.* 2020.

the next word in a sentence based on the preceding words. These models can be used for various tasks, such as text generation, machine translation, sentence completion and question-answering. When given a text input (a s.c. "prompt"), the model uses the information learnt during training to understand the context and generate a coherent response. The largest and best-performing family of such models so far is the GPT family, developed by Open AI, the latest releases of which are GPT3.5 and GPT4 (both embedded in ChatGPT). Other large language models have recently been produced, an example being Google's Bard.

Regardless of the approach used, all machine learning systems consist of two fundamental components: the *learning algorithm*, also known as the training algorithm, and the *learned algorithm* or the model. The learning algorithm uses a training dataset to construct the model, which can be perceived as a mathematical function generalising the connection between inputs and outputs in the data. For instance, it might associate images of animals with corresponding species names or legal case descriptions with related decisions. These models generalise from the training examples and can handle new cases that differ from the training data. The learned algorithm is then employed by the system to provide accurate responses to new cases based on similarities to previous examples. If the new case resembles past examples, the system proposes a response that is similar to the ones provided for those similar cases, ensuring the application of learned knowledge to real-world situations.

Machine learning systems may employ various learning methods, including decision trees, statistical regression, support vector machines, evolutionary algorithms, neural networks, and more. These methods differ not only in predictive performance but also in their ability to provide explanations. In fact, there is often a trade-off between the two objectives: systems that provide the most accurate predictions are often less capable of justifying their decisions.

For example, the learned algorithm may consist of a decision tree. This can be seen as a representation of a decision-making process, where an algorithm recursively splits the data into subsets based on the most significant attributes or features, ultimately leading to a decision or prediction. A decision tree can be understood by following the sequence of tests or decisions it makes to arrive at a particular result.

Some learned models, however, are not based on sequential operations, connecting meaningful premises to conclusions. Rather, these models involve complex calculations aimed at reproducing statistical correlations between input features and the outcomes to be predicted. Today, the most influential models are probably represented by neural networks, i.e., computer systems consisting of nodes (called neurons) connected by links with assigned numerical weights.

Opposite to expert systems, neural networks represent the most notable application of the s.c. "sub-symbolic" approach in AI[22]. A network does not process symbols – such as linguistic entities that express concepts and refer to certain types of objects, like words in human language – but rather vectors of numbers. For instance, when a neural network processes a sentence, it does not deal with the words themselves or their meanings in the way we do. Instead, it represents words as numerical vectors, often based on their frequency or context within a vast corpus of text.

Unlike a decision tree, a neural network does not offer easily comprehensible explanations. It is possible to trace back how the system arrived at a specific decision, by examining how the input influenced the

[22] As noted by SMOLENSKY, the sub-symbolic approach in AI expresses the idea of an intellectual activity performing at an intermediate level between the human brain's symbolic level and the pure neural level. In sub-symbolic AI, knowledge (concepts and rules) is not explicitly represented by symbols but implicitly learned from data through operations expressed in mathematical operations. Unlike in symbolic AI, the sub-symbolic hypothesis holds that it is impossible to give a complete representation of mental processes at the level of symbols. Instead, it assumes that such processes should be represented beneath the conceptual level, that of neuron-like nodes and synapses-like edges. The intelligent behaviour of sub-symbolic computer systems cannot be broken down into single logical operations but is the inseparable product of the operations performed by the neuronal units within the network (1988, 12).

activation of particular neurons and how the latter, in turn, influenced the activation of other neurons. However, this information fails to convey the reasons why the system gave a certain response to a particular case in a way that is understandable or meaningful to the human mind. The challenge in understanding neural networks amplifies with the increasing number of nodes, their organisational levels, and the intricacy of their connections (s.c. deep learning networks). Like other models that defy straightforward explanations, neural networks are often termed "opaque" or even "black boxes"[23].

Various research endeavours are currently dedicated to developing techniques to explain neural networks[24]. However, the achievements of such research remain considerably limited. As a consequence, when selecting the appropriate predictive system for a given context, a balance should be struck between efficiency, often linked to prediction accuracy and explainability. In various contexts like robotics, medical diagnosis, or maintenance, prioritising efficiency is common and may represent the wisest choice. Choosing a system that minimises errors, even at the cost of lacking explanatory capabilities, can be justifiable in such scenarios. In public functions, especially when multiple interests are at stake, and there is a significant need for control and oversight, a distinct balance might be needed when employing predictive systems. This need becomes all the more evident in the judiciary, where explanations for decisions are crucial. Explanation serves not only the parties engaged in the proceedings, including the judge, but also society as a whole.

3. A PHENOMENOLOGY OF PREDICTIVE JUSTICE: TYPES AND APPLICATIONS

---

23 PASQUALE 2015.

24 GUIDOTTI 2018.

We will now consider how automated systems can be used to afford legal prediction systems. This entails answering a theoretical and an empirical question. The first relates to the concrete possibilities for legal prediction given the current AI systems and those expected to emerge in the near future. The second deals with the actual practice of legal prediction in the judiciary, that is, to what extent predictive justice applications that are possible in theory are actually used in judicial decision-making.

3.1. Different types of legal predictions

To understand the possibility of delivering automated "predictive justice" functions, we need to look closer at the systems we have today emerging from state-of-the-art research in legal predictions.
A clarification regarding the word "prediction" is needed: in machine learning, the term prediction is understood as broadly referring to any inference aimed at expanding available information on a given problem. These inferences may pertain not only to the future (when the prediction is really a forecast) but also to the past and present. Compare, for instance, the case in which a medical system makes a prognosis (anticipating the future development of a particular pathology) and the case in which it makes a diagnosis (determining what pathology is currently affecting the patient). We would say that both are predictions but only the former involves a forecast. Similarly, in the legal domain, compare the forecast that an offender will recidivate (in the future) and the predictive assessment that an accused person is likely to have committed the crime (in the past).

Object of prediction

In the legal and judicial world, different kinds of predictions *latu sensu* matter. For example, a prediction is not necessarily related to the final outcome of a case may refer to (i) legal sources that could be relevant to establish that outcome, (ii) other events or circumstances that may contribute to that outcome, (e.g., the future conduct of a party, such as recidivism of the convicted person, escape of the accused, default of the debtor, etc.); (iii) collateral aspects of the decision, such as the quantification of litigation fees or the amount of compensating damages.

Predictive systems can be used to "predict" what, given a certain fact, is the most relevant legal source that should guide the interpretation or decision on that fact. For example, legal recommendation systems have been developed to suggest the most relevant decision criteria, statutes, or case laws that judges or lawyers should consider based on the specific facts presented in a case[25]. These systems would work by analysing vast amounts of legal data and employing association algorithms to link certain stereotypical facts to recurrent citations or legal arguments.

Other predictive systems do not anticipate legal meaning but predict future events that may condition the legal decision of a case. An example of this type of legal prediction is the COMPAS decision support system, an actuarial risk and need-assessment instrument widely used in the United States[26]. COMPAS is deployed in the criminal justice system for evaluating defendants' risk profiles: risk of recidivism, risk of violence, and risk of failure to appear in court. The assessments made by COMPAS are taken into account by judges in deciding whether to grant the benefit of parole/probation. Systems, such as COMPAS, have raised important ethical or legal concerns with reference to the accuracy of their predictions, fairness and impartiality, and the respect of due process rights, such as the right to an individual decision[27]. These concerns have greatly influenced the debate

---

[25] WINKELS *et al.* 2014.

[26] LARSON *et. al.* 2016. The system was brought into the spotlight by the famous case Loomis v. Wisconsin. Eric Loomis was charged with driving a stolen vehicle he used in a shooting and fleeing from police. Before deciding the case, the Circuit Court of Wisconsin ordered a pre-sentencing investigation in part based on the COMPAS assessment. As a result, Loomis was classified as being at high risk of reoffending and was sentenced to 6 years of imprisonment and 5 years of extended supervision.

[27] LAGIOIA, ROVATTI, SARTOR 2023.

on predictive justice beyond criminal risk assessment systems and are now associated with most AI applications in judicial decision-making, although not necessarily pertinent to some systems.

Finally, some predictions may relate to collateral aspects of the judgement, such as the amount of compensation for damages[28] or shared division in the assets of divorced spouses[29].

Past vs. future events

As observed above, predictions may relate to (i) a past event (e.g., the result of a case already decided) or (ii) a future event (e.g., the decision of a new case). In the first group, systems may "predict" the outcome of a decision that was already adopted based, for example, on parts of the same decision. For example, a system can be trained to "predict" the outcome of a judgement based on grounds contained in the motivation or based on the facts of the case. In this case, the "prediction" is best understood as an "outcome-based categorisation" task[30], namely categorising court judgements based on their outcome by using textual or any other information published with the final judgement but excluding (references to) the verdict in the judgement.

[28] DAL PONT *et al.* 2023.

[29] An example is the Split-up systems, one of the first legal neural networks developed to divide assets between ex-spouses following a divorce. In the Split-up network, input neurons represent the relevant factors in the decision of past cases, and output neurons indicate all possible divisions of the assets. The only input data accepted by the network (thus the only relevant factors in determining the division) are how the spouses contributed to the asset formation (lower levels of the asset lead to more egalitarian divisions). After being trained on several hundred previous cases, the network reproduces the decisions of past cases, thus being able to acquire the ability to "predict" with sufficient accuracy the decisions of future cases. See ZELEZNIKOW, STRANIERI 1995. For other methods and systems used, see the review contained in AL MUREDEN, ROVATTI 2020, 225 ff.

[30] MEDVEDEVA, WIELING, VOLS 2023.

This kind of "prediction" was realised in a study on the decisions of the European Court of Human Rights[31]. Based on the portions of the judgement containing the case facts, the system indicates whether the plea of violation is upheld or rejected. This is to say that the system does not anticipate a future event (the acceptance or upholding of the plea by the Court) but only classifies based on the outcome of already-written decisions according to the certain portion contained therein.

Two aspects deserve to be considered. First of all, although not actually predicting future cases, these systems may be useful in identifying "predictors", i.e., facts, arguments, judges, etc., of court decisions which may be recurrently associated with certain outcomes. Obviously, to avoid the system identifying the outcome within the text of the judgement and in order for it to learn new information, any references to the verdict need to be removed. At the same time, to be useful, it is essential that the model's features and their weight in the categorisation exercise can emerge. This problem relates to explainability v. opaque systems.

Second, the predictive models obtained in outcome-based categorisation may still be useful for future developments in the direction of actual prediction. This would occur when data on a case, available prior to the decision (e.g., the party briefs or the description of the facts of the case as provided by the parties, or even the entire first instance decision with respect to the outcome of the appeal) are input in the systems and used to actually predict the future decision. However, this transition bears a significant assumption, namely that the data used to learn the model (i.e., past cases) may effectively encapsulate the complexities of the legal disputes. In that case, it is not obvious that the model, having learned from past cases, can adapt to diverse, real-time inputs from parties involved in ongoing or new legal proceedings.

---

[31] See MEDVEDEVA, VOLS, WIELING 2018 and MEDVEDEVA, VOLS, WIELING 2020.

Anticipation vs. explanation

Third, predictions may be used either (i) to anticipate the legal treatment of a case or (ii) to offer some sort of guidance to the decision-makers themselves. Whether a legal prediction assumes the former or the latter meaning may depend on the interested party, namely the potential users of such prediction.

A lawyer may be interested in a prediction of the most likely outcome, and this prediction may influence his or her stance in the proceeding by using innovative arguments or avoiding filing the lawsuit in the first place. Indeed, many AI systems already exist that lawyers can use to forecast the possible results of a legal claim. For example, AI-powered legal platforms like "Case Law Analytics"[32] or "Predictice"[33] analyse vast databases of historical cases and extract patterns. By inputting details of a current case, such as the nature of the claim, relevant legal precedents, and key facts, these platforms may provide a prediction of the likely outcome of the case. Lawyers can use this information to strategise their approach, anticipate potential challenges, and advise their clients more effectively, ultimately enhancing the overall efficiency and accuracy of the legal process."

As we shall say in the next paragraph, such systems do not necessarily use machine learning functionalities, as they may use statistics with similar cases grouped around the same legal issues. Other systems are more advanced and use NLP functionalities to provide a legal assessment of a case. An example is provided by the CLAUDETTE system, which "predicts" potentially unfair clauses in consumer contracts based on consumer protection law. The system draws this prediction after learning from a set of 150 contracts in which unfair terms have been manually identified by legal experts, also considering previous court decisions[34]. It uses linguistic

[32] Case Law Analytics, <https://www.caselawanalytics.com>.
[33] Predictive, <https://predictice.com>.
[34] LIPPI *et al.* 2019.

and syntactical similarity between the sentences labelled as unfair in the training set and those found in the new contracts submitted to the system. In this case, the systems provide an assessment of an average legal expert, which may represent a proxy for the future judge's decision. Obviously, such a system will only be able to provide sufficiently reliable indications of the outcome of possible litigation to the extent that the information in the training set (the classification of terms as unfair) is actually based on previous court decisions on the matter.

Predictive systems may lead to problems in the relationship between lawyers and clients, for example, leading the lawyer to misjudge whether or not to bring a case. They could, in the long run, lead to a kind of automatic filtering of certain litigation claims, resulting in a jurisprudential stasis of potentially evolving issues. For the rest, however, this use could have a beneficial effect, namely, to discourage specious or implausible litigation when there is a reasonable certainty of the result.

Judges can use automated predictions for various purposes, each of which may bear different consequences.

The most inappropriate way of using prediction would be to determine the final decision of the case. As we shall see in the next section, assuming that legal predictions are equivalent to the decision of the final case, thus fearing a replacement of judges by the automated system, is an unfounded and unreasonable premise.

A similarly unfortunate scenario would be to look at prediction as binding or quasi-binding suggestions, namely, to mandate on judges the duty to follow the predicted outcomes, possibly when the score of this outcome is higher than certain thresholds. Judges should also not be obliged to provide specific justifications when they decide not to follow the outcome provided by the system. In this case, a predictive justice system would risk amounting to a tool for controlling judges and interfering with judicial independence.

Additionally, predictions could be used to determine how much a case is controversial. In this case, if the probability that a specific case will receive a particular outcome is very high, the judge may conclude that the case is not controversial, which means the activity aimed at its decision may be simpler. Vice versa, if the score is low, it means that the case can be solved either way and thus be particularly contentious. This information could be used, for example, by judicial offices or tribunal managers to allocate resources according to this evaluation.

Predictions could also be used as an indication of the prevailing view of other judges who have decided similar cases, or, in other terms, how an average judge would decide the case. In this case, prediction might be regarded as a synthesis of jurisprudential wisdom, amalgamating the collective experience and reasoning of the judicial community into a comprehensible form. This wisdom may be taken into consideration by the judge as much as all other elements linked to the fact at hand.

Finally, a prediction can be considered a rationale for deciding in a particular way insofar as it is accompanied by an explanation[35]. In this case, the collective judicial wisdom would not simply be passed over to the judge but also be explained by indicating the factual or legal factors, as well as their combinations, that, in most cases, led to a certain outcome. In this case, the judge would be allowed to compare the explained factors to those present in the case at hand and see how they relate to each other, namely whether there is a sufficient similarity between them or not. If the cases are sufficiently similar, then the judge may align himself or herself with the "prediction"; otherwise, he or she may distance from it. In any case, the decision will have to be properly motivated, and it may be discussed whether a sufficiently explained prediction could simply be referred to in the motivation of the case. At the same time, it must be noted that complementing a prediction with an explanation assumes two key features: the "predictors" of the system are legally or factually relevant, and the system is technically explainable.

Legal vs. extra-legal factors

[35] BEX, PRAKKEN 2021a.

As anticipated above, the prediction may be based on (i) normatively relevant elements of the case, which the judge should or can take into account in making or justifying his or her decision or (ii) normatively irrelevant elements, which the decision-maker generally does, or even should, not take into account in the final decision (identity of the parties, their social and economic status, identity of their lawyers, the judge, etc.).

As an example of the second case, predictions have been made on the decisions of the judges of the US Supreme Court on the basis of normatively irrelevant information, such as the procedure that led to the decision, the political orientation and professional background of the judges[36]. Similarly, a commercial system has been built to predict the outcome of patent cases on the basis of the characteristics of the parties, the lawyers and the judges[37]. Systems of this type may be useful for interested parties to anticipate the outcome of a dispute. However, they are not able to explain in a legally meaningful way the treatment of a case, since they use features which are not related to the merit of the case.

An illustrative example of an explanation of this system would be, "I predict that the Court of Appeal will overturn the decision of the Court of First Instance as this is what generally happens in commercial-related cases debated on Monday when Judge Doe sits as the president of the Court". This explanation may be interesting for legal scholars with a broader perspective on legal practice and its interaction with various socio-political contexts. However, it will not provide any useful knowledge to those interested in providing a legal explanation to the case, such as those carrying out the adjudicating function.

---

[36] See, KATZ, BOMMARITO, BLACKMAN 2017, which developed a time-evolving random forest classifier that leverages unique feature engineering to predict more than 240,000 justice votes of the US Supreme Court and 28,000 cases outcomes over nearly two centuries (1816-2015).

[37] SURDEANU *et al.* 2011.

Conversely, these systems may have a relevant impact in detecting forms of disparate judicial treatment on protected groups. This may occur, for example, when the system identifies statistical patterns within a set of past cases between extra-legal factors associated with protected categories (such as gender, ethnicity, religious belief, profession, etc.) and a particular outcome that favours or penalises individuals associated pertaining to that group[38].

Textual vs. non-textual data

Another distinction pertains to whether predictions are based (i) on the textual content of the judgement or other court decision or (ii) on non-textual data (e.g., keywords or factors associated with the text, information about the facts of the case and the characteristics of the parties).
In the first case, predictions are prominently based on natural language processing techniques. NLP techniques are used to provide a mathematical representation of the text, which can then be learned by machine learning systems by detecting relevant correlations.
These correlations may manifest at various levels, including the lexical, syntactical, or semantic layers of content. For instance, in our domain, a lexical-based correlation may involve detecting recurring keywords or phrases that tend to be associated with specific case outcomes. At the syntactical level, the system may recognise patterns in sentence structures or argument organisation that influence the outcome of the judgements. Finally, semantic correlations delve into the text's deeper meanings and contextual nuances, uncovering subtleties that can sway the outcome of a legal dispute. The detection of semantic correlation is rarely performed only

---

[38] For example, CHEN (2019) shows that low predictive accuracy may identify cases of judicial "indifference," where case characteristics (interacting with judicial attributes) do not strongly dispose a judge in favour of one or another outcome. In such cases, biases may hold greater sway, implicating the fairness of the legal system.

by using ML approaches, as they generally fail to understand the meaning of words, including legal concepts. Different approaches have been used in this domain to include explicit forms of legal knowledge representations, such as thesauri or ontologies.

As seen above, a notable area for the deployment in NLP pertains to the domain of LLMs, also called "foundational models"[39]. LLMs have also been used to "predict" (aka categorise, see above) judgements via legal prompt engineering, namely by designing natural language questions to generate pertinent responses by the model[40]. An already-written decision pertaining to the ECHR, where the verdict is obscured, is presented to the model with the subsequent question as to whether that specific case can be associated with a violation or not. The system appears to be able to correctly classify the decision.

Predicting outcomes via NLP and LLMs models by detecting a correlation between the text of the decision and the outcome may be particularly valuable for empirical research on judicial drafting methods. However, it rarely provides the case for genuine prediction. In fact, correlations of this type can be detected only after the decision has been written, which in turn implies that the outcome is already known. The fact that the outcome is already known may greatly influence the way the decision is drafted in its various sections, as the court's reasoning and the finding generally hint at the outcome. Similarly, the summary of the facts of the case, as well as the parties' claims and arguments, are generally influenced by the outcome of the decision.

---

[39] See BOMMASANI, HUDSON *et al.* (2022) who provide a report on opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations).

[40] TRAUTMANN, PETROVA, SCHILDER 2022.

This means, on the one hand, that predictive systems which find a relevant correlation between the linguistics and syntax of the motivation and the outcome may be biased because they might be considering words or sentences that already anticipate the outcome itself. On the other hand, we can conclude that a genuinely predictive system can be realised only if the correlation is detected between the outcome of a decision and textual information about a court case available before that decision is made public. This textual information can, for instance, be submissions by the parties or information (including judgements) provided by lower courts, that can be used to predict the outcome of the appeal.

Predictions can also be based on non-textual data, i.e., meta-data, manually or automatically associated with legal cases. This may include various extra-legal factors, such as the location of the court and the judges involved, but also legal factors related to the matter of the case. The second case might encompass the type of legal issues at hand, the specific statutes or regulations involved, and the parties' characteristics. By analysing this multifaceted metadata, machine learning models, such as decision trees or support vector machines, can predict how a given case will likely unfold within the legal system.

However, even in this case, these systems are not immune to bias. The selection of features is carried out by a human expert in the field, who, despite their expertise, may inadvertently introduce biases during the feature selection process. For example, the expert might unknowingly include certain features that are correlated with protected characteristics like race, gender, or socioeconomic status. This unintentional bias could lead to skewed predictions, favouring or disadvantaging certain groups of people based on these protected traits.

Furthermore, the bias can emerge from historical data. If the training data used to develop these predictive models contains biases from past legal decisions, the algorithms might learn and perpetuate those biases. For instance, if discriminatory practices influenced previous judgements, the predictive models might inadvertently incorporate these biases into their predictions, reinforcing unfair outcomes.

Additionally, biases can also arise due to the interpretation of meta-data. Different legal professionals might interpret the relevance of certain factors differently, leading to variations in how meta-data is labelled and categorised. These variations in interpretation can introduce subjectivity, potentially leading to biased predictions.

Explainable vs. opaque

Finally, the system may (i) be capable of providing an explanation for its decisions, citing reasons understandable to humans, or (ii) remain opaque to the request for explanation.

As seen above, in the first case, the system provides its prediction along with the indication of the relevant aspects it relies on. In Section 2.2, we have said that, for instance, a decision tree-based system can offer such explanations, by retracing the path from the tree's root to the decision. Similarly, a system based on association rules can explain the result by presenting the rules used to arrive at it.

Conversely, the second scenario applies to neural networks (especially deep ones) where a meaningful explanation is lacking. The backward chaining in neural networks only consists of retracing how the activation of a neuron situated in a subsequent layer has resulted from complex mathematical calculations triggered by neurons of previous layers. While still being a kind of explanation, this information cannot adequately be used as a justification for judicial predictions, as it bears no legal significance. We will further consider the problem of the black box in Section 4.5.

3.2. Current applications in predictive justice: a focus on the Italian legal system

While everything discussed above represents the state-of-the-art in predictive justice research, concrete, real-world applications are still limited. On this point, the European Commission for the Efficiency of Justice (CEPEJ) acknowledged in the *Ethical Charter on the use of artificial intelligence*

(AI) in judicial systems and their environment[41] that the employment of AI in the justice sector still primarily constitutes a commercial endeavour within the private sector. Initiatives are prominently directed towards entities such as insurance companies, legal departments, practising lawyers, and individual users. Conversely, the Charter highlighted that no Council of Europe member state has incorporated predictive software into their judiciaries. Although there have been localised trials to assess the potential of these applications, they have yet to witness widespread adoption on a substantial scale.

Five years later, the situation is not much different. Notably, the CEPEJ has launched an initiative to collect data and information on current applications in Europe[42]. This initiative has resulted in an open database where useful information can be found, e.g., the year the system became functional, a short description including the underlying technology, a link to an official public source/reference of the system, the status of the system (i.e. whether the system is currently in function or is still in piloting phase). The database includes applications, not only strictly dedicated to domestic courts but also lawyers and law firms. Different categories of users have been identified, such as court users (general public), court management, judges, lawyers and prosecutors.

Comparatively, Italy seems well positioned in the development of predictive justice applications. Although no actual use of AI in the judiciary is made at present, many projects are in the piloting phase.

Some of them represent individual initiatives of some Italian Courts. For example, the website of the Court of Appeal of Bari contains a section titled

---

[41] European Commission for the Efficiency of Justice (CEPEJ), European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018), available at <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018>.

[42] Cfr. Resource Centre on Cyberjustice and AI, https://www.coe.int/en/web/cepej/resource-centre-on-cyberjustice-and-ai.

"Progetto Prevedibilità delle Decisioni"[43]. The project started in 2016 from the initiative of the President of the Third Civil Section and, so far, has led to the manual creation of summaries of established case law on recurring subjects and common cases. The website provides access to such summaries, which are categorised by topic. While the initiative may provide valuable input for predictive justice applications, it has so far not involved the use of automated systems.

The Court of Appeal of Brescia has also been leading a project called "Predictive Justice" since 2018, involving the Tribunal and the University of Brescia. The aim is to provide legal professionals with a prediction of the duration of proceedings on specific matters (e.g., corporate and industrial matters, bankruptcy and insolvency proceedings, social security etc.). The research results are available on the University's website in a section named "Predictive Justice System of the Court of Appeal of Brescia and the Tribunal of Brescia"[44]. The website allows users to identify judicial cases most similar to their legal inquiries through a guided path consisting of essential linguistic and graphical formulas. This project, too, while being beneficial, is currently applying no AI techniques: it provides a simple classification of cases and related statistics on the average decision time of such cases.

More advanced techniques are being developed in the "Predictive Justice" Project, led by the Lider-Lab at the Dirpolis Institute of the Sant'Anna School of Advanced Studies in Pisa, and involving the Tribunals of Genoa and Pisa[45]. This initiative involves the annotation of judgements, primarily in the areas of personal injury and separation and divorce alimony. Based

[43] The results of the project are available at the Courts' website https://www.corteappello.bari.it/buone_prassi_4.aspx.

[44] The system is available on the Court's website, https://giustiziapredittiva.unibs.it/. See also CASTELLI-PIANA 2018.

[45] See, MAGLIONE 2021. The platform is available at the following https://www.predictivejurisprudence.eu/.

on such annotation, tasks are being developed pertaining to the automated identification of statistics and to case-flow management, i.e., recommending tribunals' managers how to efficiently allocate cases among the various sections. The declared long-term objective would be to develop predictive models that can assist judges or help evaluate the timing or outcome of a legal process.

Through Italian Next Generation EU funds, the Ministry of Economy and Finance and the Council of the Presidency of Tax Justice have funded the PRO.DI.GI.T project for predictive justice in the tax field[46]. Over one million judgements from the (now repealed with the recent reform Law 130/2022) regional and provincial tax commissions from the last five/six years have been made available to a team of computer scientists and lawyers from several Italian universities, such as University of Bologna, to develop AI techniques for different kinds of tasks. One particularly relevant outcome of the project has recently been the use of Large Language Models (such as GPT) to automatically extract summaries from cases[47]. Other works currently involve the automated extraction of summaries and the development of case outcome predictions.

The University of Bologna has also been involved in two other projects focused on predictive justice. The first is the 2017 PRIN LAILA Project involving the University of Turin, the University of Pavia and the University of Napoli. The project's objective is to 1) apply, refine and develop technologies for legal analytics to legal information (Italian legislation and contracts); 2) provide methodological analyses and guidelines for the efficient and ethical use of the aforementioned technologies; 3) expand the

[46] Dipartimento delle Finanze, Presentazione del progetto Pro.Di.Gi.T, https://www.finanze.gov.it/it/Progetti-europei/PRO.DI.GI.T/Presentazione/. Cfr. Ione Ferranti, Prodigit, come funziona il progetto per la giustizia tributaria digitale, in Agenda Digitale, 10 Marzo 2023, < https://www.agendadigitale.eu/documenti/giustizia-digitale/prodigit-come-funziona-il-progetto-per-la-giustizia-tributaria-digitale/>.

[47] DAL PONT *et al.* 2023.

understanding of the structure, logic and dynamics of Italian law in connection with EU law.

The second project is called ADELE (Analytics for DEcision of LEgal cases), which is coordinated by Bologna in partnership with the University of Luxembourg, the University of Turin, the European University Institute, Apis Europe, the Union of Bulgarian Jurists and the Bulgarian LIBRe foundation. The project aims to analyse Italian and Bulgarian case law on intellectual property and tax law for the extraction and modelling of knowledge to be used for predictive purposes. The pilot tool developed in the project is available on the Internet and, among others, provides the possibility of automatically extracting arguments from judicial texts[48], as well as gaining a score on the potential outcome of a specific request input by the user on matters related to the legal focuses of the projects[49].

4. ANALYSING CRITIQUES ON PREDICTIVE JUSTICE

In this section, we address the main critiques and concerns that have so far been raised about using predictive techniques in the judiciary. We will discuss issues based on our experiences and what we have seen so far about the status of predictive justice. We map the critiques into three distinct groups. The first two are critiques based on assumptions that, in our opinion, are unlikely to make practical sense. The third and fourth pertain to substantial challenges that could be mitigated through practical or technical remedies. The last three critiques relate to fundamental issues that require further investigation.

4.1. Substitution effect

---

[48] GRUNDLER *et al.* 2022, SANTIN *et al.* 2023.
[49] GALLI *et al.* 2022.

As we mentioned in the introduction, the most resonating concern around predictive justice relates to the fear of machines substituting human judges, decisions being entirely handed over to machines, or a new form of normative information taking control, being this dystopian future commonly exemplified by the picture of a "robot-judge".
It seems to us that the scenario of a full replacement of judicial functions by an AI system is completely unrealistic. We argue that this is the only conclusion that can be drawn based on the current and expected developments in predictive justice that we have provided in the previous sections.
Besides this simple empirical explanation, we can provide another one based on social and institutional reasons.
As made clear by the theory of speech acts[50], the same sentence or utterance has different values and meanings depending on the person delivering it, the context in which it is said, and the intention behind the speech act. For example, the sentence "Giovanni is right" has a different meaning if it is uttered i) by a friend at a bar over a drink, ii) by a student or a professor in an academic class, iii) by a lawyer in conversation with her client or iv) by a judge in the exercise of her duties. Only in the last case will that sentence be a decision in a dispute because it will be uttered by one who, according to the established legal and social rules, is in the position of, and is entitled to, "speak the law", namely, to decide about the case involving Giovanni. This person has such power because, following a public selection, she has acquired the position of judge and, according to specific organisational and procedural rules, is called upon to decide that specific case. The effect is that she can pronounce the sentence "Giovanni is right" with the force of an authoritative and binding decision for the parties. All other utterances will be, the first, a friend's opinion; the second, a didactic exercise on a

[50] Literature on this subject is overwhelming. We limit our reference to AUSTIN 1962, which represents the text underlying the theory of speech acts.

hypothetical case; the third, the mere prediction of a lawyer in the interest of her client.

The prediction of an automated system stating that "Giovanni is right" and, therefore, will win the case should be put in such a perspective. First, the sentence is uttered by an entity which has not been given the power to make legal judgements by either a law (i.e., by providing for the legal efficacy of automatic prediction) or a social norm (by passively adapting to such prediction). Second, the sentence should be related to the context in which it was generated. In judicial decision-making, the context is provided by specific procedures and norms that the judge must follow to come to the verdict, which implies the examination of facts and evidence, hearing the parties, due process, and motivating the decision. These can be seen as consecutive facts of judicial decision-making so that the above-mentioned statement by an automated system may be regarded as incomplete or ontologically inadequate in a legal context. Finally, there is no intention behind the statement. When an AI system produces a statement like "Giovanni is right", it is not driven by a purpose or ethical commitment to that decision and its potential consequences, but rather follows statistical correlations and patterns found in the training data.

In conclusion, machines do not and will not decide anything because, in order to decide the law, one must be entitled to do so. We can expect that no government, neither at present nor in the near future, would ever delegate adjudicating functions over the law to predictive machines, nor will the people start to consider the transferal of such power an acceptable option.

To this, we can add that there is no actual possibility of an objective "ground truth"[51], as the target to be predicted is precisely the human decision in a

[51] In machine learning domain, the term "ground truth" is used to refer to the correct outcome, as identified through standards external to the system, as opposed to the outcome that is proposed by the system. This expression, often used in machine learning apparently derives from cartography, and opposes the representation on a geographical map to the

dispute, and there is no objective verification of the correctness of the decision. As seen above, there is also the possibility of developing systems that do not merely predict a single solution but can formulate several alternative hypotheses, with the effect of enhancing the judge's ability to choose and speeding up the evolution of the law (only prototypical realisations of these techniques exist at present).

There is one exception where a substantial replacement by machines may happen and may even be wise. In the case of online dispute resolution[52], having an automated system that merely "indicates" which litigation solution is most closely aligned with past decisions in similar cases may be a useful guide for the parties who may decide whether to adhere to it or not. This may be a wise solution, firstly, because the system does not decide anything, as it limits itself only to indicating what could be the state of the art at the time of the decision. Secondly, the decision to adhere or not to the "indication" of the machine (machine utterance) is discretionary and relies on negotiation between the parties without any authoritative value (other than the implicit and consequent value of entering into a contract).

4.2. Rule of code

According to another concern, once the judicial function is handed over to machines, much of the existing law and legal institution as we know it today will be replaced by fully computational mechanisms. According to this idea, if the decision is completely automated, there may be no need for human lawyers to defend their clients, as the decisions would be made solely based on algorithms and data. Similarly, it would make no sense to allow for appeals, as automated decisions are deterministic and thus final and not

real situation on the ground, which provides the undisputable standard to determine whether the map is correct or wrong.

52 For an overview, see LODDER, ZELEZNIKOW, 2012.

subject to any other interpretation. If machines take on the judicial role entirely, there is a possibility that they might not just stop at judging but also start influencing or even creating laws[53]. This could mean that the automation of legal decisions extends beyond the courtroom into the legislative process.

This idea is based on a completely unrealistic view of the capabilities of machines and a mechanistic view of law and its operation that must be rejected.

With regard to the first vision, it is completely fallacious to argue that delegating certain judicial functions to AI systems means necessarily accepting the view of the mechanistic and unitary application of the law to the extent that whichever way one looks at the same legal problem, the answer is always the same. As we have seen in the previous section, different techniques may be used to build predictive systems, starting from completely different theoretical assumptions. AI encompasses a wide range of approaches and methodologies, including machine learning, neural networks, rule-based systems, and natural language processing, among others. Each of these approaches can be used to build predictive systems, but they do so in distinct ways. For example, machine learning models can identify patterns in data, but the interpretation of those patterns can vary based on the specific algorithm and training data used. This diversity in AI approaches allows for different perspectives and interpretations of legal issues. Rather, what is important is to grasp how legal norms and principles, with their interpretation and assumption, are translated into datasets or rule-based codes[54].

Also, keeping the technical perspective, AI systems can continuously learn and adapt based on new data and feedback. This ability to evolve and

---

[53] This development is what DIVER has called "digisprudence", i.e., the idea that a *rule by* or even a *rule of* computer code, taking its normative force seriously while raising issues with the normative framework of the rule of law (2021).

[54] HIDELBRANDT, 2023.

improve over time aligns with the dynamic nature of the law. As legal precedents evolve and societal norms change, AI systems can abstractly incorporate these shifts into their decision-making processes, offering flexibility and responsiveness to legal developments.

Also, it must be recalled that if AI systems can provide an explanation for their decisions, this may actually broaden the scope of action of judicial decision-making rather than conflate to only one possible interpretation. Explainable AI can shed light on the factors and data points that influenced a particular decision, allowing judges to understand and scrutinise the reasoning behind AI-generated predictions. This means that either judge will conform to the precedent thanks to an enhanced comprehension of previous cases and after detecting similarity with the case at hand or distance from it by providing novel interpretations.

Regardless of these technical aspects, we argue that predictive justice does not mean that the whole legal experience can coherently represent itself with mathematical and "mechanical" precision. In fact, the law faces several barriers when it comes to being represented in a computable form.

First, the law does not always provide clear and exhaustive rules nor straightforward canons of interpretation. It often uses vague terms, and interpretations can change with the evolution of social norms, particularly through jurisprudence. Codifying political and social opinions is not straightforward because it always reflects a political compromise. Moreover, it is impossible to anticipate all possible scenarios. Second, to represent the changing nature of legal norms, computational systems must be designed to account for these changes, including developments in jurisprudence. The law is not static; it evolves over time as new cases are decided, and societal values shift. An effective computational representation of the law must accommodate this dynamic aspect, which adds complexity to the task. Finally, the law involves a delicate balance between legal certainty, fairness, contextualisation, and justice. This balance is not always easy to strike, as it depends on the specific circumstances of each case and the broader legal and societal context. Achieving this balance is a nuanced and complex task that may not always lend itself to straightforward computational representation.

4.3. Herd effect

Let us address two significant critiques that warrant attention, which, however, can –in our view– be effectively addressed through practical or technical solutions.

Some authors fear that the development of predictive systems, if not leading to replacement, will result in an inexorable impoverishment of judges, who would rather adapt to what the system suggests than engage with the case and develop new solutions[55]. As a matter of fact, predictions may wield significant influence on decisions. When an algorithm suggests a likely solution, it can markedly sway a judge's decision-making process. Especially, the higher the probability assigned to a specific outcome, the greater the impact. Judges, urged to decide quickly to move on to the next case and unwilling to stand against the majority of their colleagues, might well align their decisions with automated predictions. All this may lead to a "herd effect", the situation where individual judges tend to blindly follow the actions of the majority rather than making independent decisions based on their own knowledge.

In our view, however, the risk of a herd effect is not a distinct feature of predictive justice systems, as any collection of precedents, even just made on paper, may lead to such a result. In these cases, judges may naturally be influenced by the decisions made by their fellow judges or colleagues. Especially in countries with stare decisis principles, past decisions, especially famous cases, can set legal precedents, indicating how a particular issue

---

[55] This is what GARAPON has called "*effet moutonnier*" (2018). See also, the interview to the Author released in the Italian Journal *Questione Giustizia*, edited by FRONZA, CARUSO 2018.

should be decided based on previous cases. Also, judges may often hold their colleagues in high regard and may respect their legal expertise. If a respected colleague has made a decision in a certain way, it can carry significant weight in influencing the judge's own decision. In other cases, some judges may be risk-averse when making decisions and may prefer to align with previous decisions to avoid criticism, controversy, or potential appeals. So, similar issues may arise, as judges might not independently analyse the legal issues at hand. Instead, they may rely heavily on established precedents as guiding principles, effectively following the lead of their colleagues without critically evaluating the unique aspects of the current case.

Henceforth, we are faced with a dichotomy: either we eliminate any form of documentary evidence and even past professional knowledge, as these may exert profound influence, or we accept –actually we support– the idea that judicial decision must be influenced by past knowledge, either when represented in isolation (as a jurisprudence maxim) or aggregated in a statistical form.

The key is to ensure that judges possess the education and training, as well as the means, to understand the statistics provided by the systems. On the one hand, this requires programmes to explain how these systems work and should be fostered. On the other hand, efforts should be devoted to building explainable predictive systems. Indeed, if the system is opaque and cannot explain the prediction, it can lead to a dependency of the judges on the machine, whose conclusions are accepted without understanding the reasons for them. Similarly, the judge would conform to the system even when the output of the system is presented as an indication to follow the decision at hand.

The only way to prevent a herd effect would be to look at and present the prediction of the systems as a mere suggestion that has to be proved with the facts provided in the case at hand. The judge would not be able to simply refer to the output of the systems but should be pushed towards the verification of all the factors in the present case on which the explained prediction is based. Moreover, an important remedy to the herd effect may come from the design of systems that do not merely provide a suggestion on the most likely outcome but also indicate the existence of different trends within precedents case law, possibly associating each of them with the different legal reasons supporting the different attitudes. It will then be up to the judge, once alerted of the existence of possible conflicting interpretations, to decide whether to align himself with the prevailing trend or instead with a minority orientation.

4.4. Data

Among the critiques to predictive justice, some are related to the data-driven nature of prediction.

Scholars have argued that a prediction system that is based only on judicial precedents would be incompatible with a civil law system, where decisions should be based on law and not on precedents[56]. This critique seems to us to be only partially accurate. First, traditional clear-cut distinctions between civil law and common law systems can no longer be given a priori[57]. Both cultural and legislative changes have shown that, also in civil law systems, precedents, especially those on points of law, have become a 'strong' rule supporting a certain decision. De facto, precedents have become endowed with that capacity to become a projection of the outcome of future cases to which they are applied.

Another problem concerning the use of data relates to the inherent bias of AI systems trained on previous judgements, as large as this collection may be. As existing experiences show, trained systems generally have access to a limited amount of data that cannot fully represent the whole existing law.

While this criticism makes a point, it seems to us that it does not actually concern the use of AI in the judicial systems. It rather relates to the efforts

---

56 See, among others, FILIPPELLI 2019.

57 CAPPELLETTI 2018.

that exist or should exist in rendering the largest possible number of judgements, even those that are extremely old, in digital form. Likewise, although we recognise the importance of pluralism in legal databases, one crucial aspect to overcome the situation is to fix the current fragmentation of case law into multiple databases, most of which are possibly closed.

A further aspect that deserves to be considered concerns whether a system trained on textual precedents is in itself adequate to ensure reliable and complete legal predictions. In particular, the question arises as to whether the text of a judgement is always and necessarily representative of the factual and legal reality it represents. For example, it is common knowledge that the facts of a case described in a decision are presented as appraised by the judge for the purpose of deciding the case and are not represented as they may manifest in the parties' submissions. This is the reason why, often, first-instance judgements are challenged for misrepresentation of facts or erroneous legal qualification. Likewise, legal issues included in the motivation of a decision may not be complete. With respect to this last point, lawyers often claim the tendency of judges to articulate their reasons in a non-exhaustive manner and to disregard certain arguments put forward. As a result, new legal questions and novel interpretations proposed by lawyers, which might actually deserve to be included in the circuit of legal knowledge, often go completely unnoticed. These two issues underline the need for a more explicit recognition of the potential limitations and opportunities associated with the incorporation of precedents in civil law systems.

On closer inspection, this problem relates to the very nature of judicial decisions. These are not narrative documents that document the facts and arguments in a proceeding but represent the institution through which judges bring justice to a case. What is described in a judgement should not be regarded as "true" or "full" information describing the process. For this purpose, other documents may be needed, such as parties' submissions, briefs reporting meetings between lawyers and the judge, evidence and expert reports, and referrals between judges of different instances.

Obviously, one could work in the direction of enriching existing databases with judges' decisions, trial documents, and all the documents of the lawyers in the respective trials. Obviously, such a direction would require very careful consideration of the privacy and confidentiality aspects of forensic strategies. More broadly, one could rethink the role of the judgment in the broader circuit of trial documents. They could bring back the essential logical-legal structure of judicial reasoning with immediate and direct references to lawyers' questions. The new generation of measures could thus become digital native files, capable of collecting structured data and capable of automatically capturing the parts of the lawyers' defences that lawyers themselves will have indicated as essential (a sort of abstract). In this way, the new collections of case law will consist of the decision-makers but will also contain the essential and potentially valuable parts of the parties' pleadings, selected (it is worth repeating) by the lawyers themselves.

4.5. Black box

In Section 2.2, we observed how a neural network, while being particularly effective in simulating decision-making, cannot provide explanations regarding how it arrived at a particular decision. On the other hand, explainable AI systems, such as those case-based, can provide insights into the factors and reasoning that determine a specific outcome, but at the same time are limited and scalable as they require a demanding formalisation exercise. Striking a balance between these two aspects, namely efficiency (accuracy in predictions) and explainability, represents a key issue.
This balance may differ depending on the context and on applications[58]. For example, in medical diagnoses, accuracy may be more important than explanation, as precise and reliable diagnoses are critical for patient treatment and well-being, especially if they require immediate decisions and

[58] BELL *et al.* 2022.

interventions. Also, patients trust healthcare professionals to make accurate diagnoses and provide appropriate treatments without necessarily understanding their explanations.

When fundamental rights are at play, and the need for control is paramount, as in the case of judicial decision-making, the need to afford explanations becomes fundamental. At the same time, different applications may require different levels of explainability. For example, systems predicting legally relevant sources or qualification of facts may not necessarily be explainable; what is important is that they can succeed in their objective accurately and reliably. In legal contexts, especially in complex cases where extensive data analysis is involved, the focus may be on the system's ability to correctly identify relevant legal precedents, statutes, or factual evidence without necessarily providing a detailed explanation of every step taken. Conversely, as said above, outcome prediction systems are greatly useless if they are not capable of providing an explanation for their predictions. In domains such as criminal justice or financial forecasting, understanding the rationale behind a particular prediction is essential. Without a clear explanation, stakeholders, including judges, lawyers, or financial analysts, may lack the necessary confidence in the system's predictions, hindering their ability to make informed decisions based on the system's output.

The balance may also entail attempting to merge the two approaches. Currently, there are several research endeavours to use knowledge-based methods to provide an explanation of black box systems[59]. For example, logic as constraint provides methods supporting the creation of predictive models –possibly including or involving some black box component– whose behaviour is constrained by a number of symbolic and intelligible rules usually expressed in terms of (some subset of) first-order logic, so to build explainability by design. Other techniques are used to explain ex-post the behaviour of predictive systems by somehow manipulating some poorly interpretable pre-existing system. For example, feature relevance methods

[59] CALEGARI-CIATTO-OMICINI 2020.

focus on how a model works internally by assigning a relevance score to each of its features, thus revealing their importance for the model in the output.

In this complex landscape, computational argumentation assumes a distinct role[60]. By integrating logical reasoning and computational capabilities, argumentation frameworks can provide a structured approach to explaining AI-driven decisions. These frameworks allow for the systematic analysis of evidence, legal principles, and inference processes, enabling a comprehensive understanding of the decision-making rationale. In essence, computational argumentation may bridge the gap between the intricacies of legal reasoning and the computational power of AI, offering a promising avenue for enhancing both accuracy and transparency in predictive justice systems.

While the choice of explainable AI systems should always prevail over, or possibly be combined with, black box techniques, we also think that the problem should not overestimated.

4.6. Intellectual property

In Section 3.1, we mentioned the case of Mr. Loomis, who received a significant increase in the overtime penalty due to the predicted risk of re-offending without being informed about the underlying reasons. All the concerns raised pointed towards a most serious problem: the legal inability of Mr. Loomis to get to know how the algorithm worked due to the copyright protection on the programme.

Assuming that the system is not a black box, the possibility of studying the code would mitigate many issues that were raised in the COMPAS case. For instance, biases resulting in disparate treatment for protected groups could be monitored and corrected by amending the code or revising the dataset.

---

[60] ROTOLO, SARTOR 2023.

Similarly, once it is possible to know and correct potential biases in the systems, one could ask whether using automated systems in sentencing is not more equitable than leaving the decision exclusively to humans. Human minds also have prejudice, but unlike machines, they cannot be forced by the law to be inspected to detect biases.

Copyright, or better, the restriction of its exercise, is, therefore, one crucial aspect to consider[61]. Transparency requirements could be considered to explain the functioning of predictive justice algorithms, the data they are trained on, and how decisions are reached. Also, predictive justice systems may be required to undergo an auditing process, which would allow code and dataset inspection to be carried out by expert people and within a protected environment. At the same time, national judiciaries could promote the development and use of predictive justice algorithms, allowing public discussion and scrutiny of the code. For example, consider this hypothetical scenario: the Italian Ministry of Justice provides Italian tribunals with a COMPAS-like system, whose logic and architecture are, however, known in advance, discussed and approved by bar associations and judges, and verified and corrected based on applications.

A similar development has taken place in the United States with the introduction of the public PATTERN system (Prisoner Assessment Tool Targeting Estimated Risk and Needs)[62]. This marks a significant shift towards transparency. Not only is the new algorithmic system not owned by private companies and thus not covered by intellectual property rules, but it must also undergo checks by independent entities appointed by the Department of Justice (DOJ).

4.7. Status quo effect

---

[61] PASQUALE 2015. Also, on copyright law as an additional layer of "remediable opacity" on algorithms, see BURREL 2016.

[62] HAMILTON *et al.* 2022.

As previously discussed, in the context of supervised learning, an AI system learns from a collection of past legal decisions. The system typically suggests a solution that aligns with the majority of these historical judgments. However, if judges were to consistently follow the system's direction, it could stagnate the evolution of the law. This concept was critiqued by legal scholar Roscoe Pound, who referred to it as "mechanical jurisprudence" that would lead to a "petrified" legal system, unable to adapt to contemporary challenges and societal perspectives. Additionally, relying solely on past decisions could overlook legislative changes, potentially misleading judges in future cases and contradicting the principle of adhering to current law.

Several technological solutions have been proposed to address this issue. One approach is to give more weight to recent examples within the training data rather than historical ones. Another involves linking past cases to the legislative norms they applied, considering the possibility of subsequent norm changes. However, these solutions are not foolproof. They must be complemented by the user's knowledge and expertise. In the context of cognitive computing, users should view the system as a valuable information tool rather than a complete substitute for their decision-making role.

Some scholars, notably Massimo Luciani, have raised concerns about the type of legal doctrine and methodology incorporated into AI systems, how they manage relationships with legal precedents, and how they resolve conflicts between norms. It's worth noting that, to date, there is not a system capable of handling these complexities, and attempting to replicate it within AI may not be resource-effective[63]. These nuanced aspects of legal reasoning are best left within the purview of human expertise.

Consider the scenario of a judge who, suspecting a conflict with a constitutional rule, may refer a legal statute to the Constitutional Court for review. In this context, it is noteworthy that the rule in question has been

[63] LUCIANI 2018.

consistently applied on numerous occasions, including by the Supreme Court, without controversy. However, as the judge prepares to apply this rule for the 1001st time, a sudden realisation dawns, either through independent contemplation or under the counsel of a defence attorney, that the rule may indeed conflict with the Constitution, supported by compelling yet previously unnoticed reasons. In this instance, we posit this as an illustrative case of a distinct facet of human intellectual capability, one that would pose a formidable challenge for an AI system to replicate. It underscores the imperative for judges to nurture and develop this cognitive ability.

5. CONCLUSION: TOWARDS COGNITIVE-ENHANCING LEGAL PREDICTIONS

In this paper, we have delved into the realm of predictive justice, examining its current approach in AI applications.

We have scrutinised the existing methods employed in AI to develop predictive justice applications.

Building on this empirical knowledge, we have explored predictive justice phenomenology, shedding light on the diverse legal predictions achievable through contemporary AI systems. At the same time, we have reviewed current AI projects and pilot tools that are being developed in Italy, specifically for the judiciary.

Moving forward, we have critically analysed the recurrent fears and critiques surrounding predictive justice. We discerned between unreasonable critiques, those reasonable concerns that could be addressed through technical measures, and issues that should be taken seriously while necessitating further investigation. By dissecting these critiques, we not only highlighted the challenges but also paved the way for potential solutions, thereby contributing to the ongoing discourse on predictive justice.

From this discussion, it can be concluded that legal prediction can have a place in judicial practice to the extent that it is possible to obtain explicable

systems, i.e., that enable the judge to understand the factors that led previous judges to decide previous cases in a particular way. At present, however, most of the AI systems used for prediction are not explainable and act as a black box. Caution must, therefore, be exercised when planning to embed these in judicial practice and decision-making.
Simultaneously, we emphasised that it is meaningful to look not just at the predictive system alone but rather at its interaction with the judge. How should a judge use the result delivered by a predictive system? To what extent can the result of the system constrain the judge's assessment and decision? We have explained why judges cannot and should not be obliged to follow the result proposed by the system. The only acceptable solution is to look at the automated prediction as a heuristic technique, i.e., as one of the many facts and knowledge that the judge may resort to when making a decision, such as the facts of the case, personal knowledge and notes provided by staff members.
In the future, lawyers, computer scientists and policymakers should take a broader look at the potential applications of AI beyond the narrow scope of outcome prediction. The richness of AI technologies lies in their ability to support and augment human legal decision-making processes in multifaceted ways.
Through advanced techniques such as retrieval, AI systems can swiftly gather vast amounts of legal data, ensuring judges have comprehensive access to relevant precedents. Extraction of rules, principles, and factors aids in distilling legal complexities into essential elements, enhancing judges' ability to focus on crucial aspects of cases. Clustering algorithms enable the organisation of this data into coherent patterns, aiding judges in discerning complex legal relationships and trends. Summarisation tools distil lengthy legal documents into concise, digestible insights, facilitating quicker comprehension of intricate cases. Argument-mining capabilities empower judges by presenting nuanced perspectives from past cases, providing a foundation for robust and well-informed decision-making. Moreover, technologies like drafting support not only expedite the creation of legal documents but also ensure their accuracy and coherence. Identifying similar

cases and trends enables judges to consider a broader context, fostering a more holistic understanding of legal issues. Furthermore, AI-driven bias detection mechanisms serve as vigilant guardians, flagging potential biases and ensuring fairness in legal proceedings. By employing targeted statistics, judges can make data-driven decisions, enhancing the objectivity and reliability of their judgements. Workflow optimisation tools streamline administrative tasks, allowing judges to dedicate more time to substantive legal analysis. Embracing these diverse opportunities, the integration of AI technologies in legal systems promises a revolution, making justice not just predictive but also profoundly insightful and equitable.

Amidst the evolving landscape of AI and justice, a transformative path could, therefore, emerge, which we may call "*cognitive-enhancing predictive justice*". This paradigm recognises predictive systems as invaluable tools not merely for forecasting outcomes but as instruments that can enrich the cognitive capacities of legal professionals. In this vision, AI technologies serve as collaborators, empowering judges to delve deeper into the complexities of legal cases. This approach emphasises not just the augmentation of judicial reasoning but the elevation of legal cognition. Judges, armed with these enhanced cognitive tools, can unravel intricate legal puzzles, ensuring that every decision is well-informed, transparent, and coherent. By embracing cognitively enhancing predictive justice, we may enter a new era where technology and judicial expertise harmonise, fostering a legal system that is not only predictive but profoundly intelligent, insightful, and fairer.

REFERENCES

AL MUREDEN, Enrico, ROVATTI, Riccardo, *Gli assegni di mantenimento tra disciplina legale e intelligenza artificiale*, Turin, 2020.

ALPAYDIN, Ethem, *Introduction to machine learning*, Boston, 2020.

ASHLEY, Kevin D., *A brief history of the changing roles of case prediction in AI and law*, 93-112, in *Law Context: A Socio-Legal Journal*, 36, 1/2019.

ASHLEY, Kevin D., BRÜNINGHAUS, Stefanie, *Automatically classifying case texts and predicting outcomes*, 125-165, in *Artificial Intelligence and Law*, 17, 2009.

AUSTIN, John Langshaw, *How to do things with words*, Oxford, 1962.

BELL, Andrew *et al.*, *It's just not that simple: an empirical study of the accuracy-explainability trade-off in machine learning for public policy*, 248-266, in *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, June 2022.

BERGOLD, Amanda Nicholson *et al.*, *Bias in the law: A definitive look at racial prejudice in the US criminal justice system*, Lanham Maryland, 2020.

BEX, Floris, PRAKKEN, Henry (a), *Can predictive justice improve the predictability and consistency of judicial decision-making?*, 207-214, in *Proceeding of The Thirty-fourth Annual Conference (JURIX2021)*, Vilnius, Lithuania, 8-10 December 2021.

BEX, Floris, PRAKKEN, Henry (b), *On the relevance of algorithmic decision predictors for judicial decision making*, 175-179, in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law (ICAIL2021)*, São Paulo Brazil, 21-25 June, 2021.

BICHI, Roberto, *Intelligenza digitale, giurimetria, giustizia predittiva e algoritmo decisorio*, 423–447, in RUFFOLO, Ugo (ed.), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020.

BOMMASANI, Rishi, HUDSON, Drew A. *et al.*, *On the Opportunities and Risks of Foundation Models*, in *arXiv preprint*, 2022, arXiv:2108.07258

BRANTING, Karl *et al.*, *Scalable and explainable legal prediction*, 213-238, in *Artificial Intelligence and Law* 29, 2023.

BRÜNINGHAUS, Stefanie, ASHLEY, Kevin D., *Combining case-based and model-based reasoning for predicting the outcome of legal cases*, 65-79, in *Case-Base Reasoning Research and Development: 5th International Conference 30 on Case-Based Reasoning*, ICCBR 2003 Trondheim, Norway, 23-26 June.

BURRELL, Jenna, *How the machine 'thinks': Understanding opacity in machine learning algorithms,* in *Big data & society*, 3, 1/2016.

CALEGARI, Roberta, CIATTO, Giovanni, OMICINI, Andrea, *On the integration of symbolic and sub-symbolic techniques for XAI: A survey*, 7-32, in *Intelligenza Artificiale* 14, 1/2020.

CAPPELLETTI, Mauro, *The doctrine of stare decisis and the civil law: a fundamental difference – or no difference at all?*, 381-393, in *Festschrift für Konrad Zweigert zum 70. Geburstag*, Tübingen, 1981.

CASTELLI, Claudio, PIANA, Daniela, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione Giustizia*, 2018 (<https: / / www questionegiustizia . it / articolo / giustizia - predittiva - la - qualita-della-giustizia-in-due-tempi_15-05-2018.php.>).

CHALKIDIS, Ilias *et al.*, *LEGAL-BERT: The Muppets straight out of Law School*, 2898-2904, in *Findings of the Association for Computational Linguistics: EMNLP 2020*.

CHEN, Daniel L., *Machine learning and the rule of law* in *TSE Working Papers* 18-975, Toulouse School of Economics (TSE).

CONTISSA, Giuseppe, *Rule-based systems per il diritto. Modelli e sistemi per la gestione e applicazione di regole*, Rome, 2015.

DAL PONT, Thiago Raulino *et al.*, *Regression applied to legal judgments to predict compensation for immaterial damage*, in *PeerJ Computer Science*, 2023.

– *Legal Summarisation through LLMs: The PRODIGIT Project*, in *arXiv preprint,* 2023, arXiv:2308.04416.

DIVER, Lawrence, *Digisprudence: Code as Law Rebooted,* Edinburgh, 2022.

EMANUALE Fronza, CORRADO Caruso, *Ti faresti giudicare da un algoritmo? Intervista ad Antoine Garapon*, 196-199, in *Questione giustizia*, 4, 2018.

FERRANTI, Ione, *Prodigit, come funziona il progetto per la giustizia tributaria digitale*, in *Agenda Digitale*, 10 Marzo 2023, (< https://www.agendadigitale.eu/documenti/giustizia-digitale/prodigit-come-funziona-il-progetto-per-la-giustizia-tributaria-digitale/>).

FERRARI, Maurizio, *Predizione algoritmica, intelligenza artificiale generativa e rischi di cristallizzazione dell'ermeneutica giurisprudenziale*, 117-122, in *Il Foro italiano*, 3, 5/2023.

FILIPPELLI, Michele, *L'incoerente parallelismo dei sistemi giuridici di civil law e di common law, tra giustizia predittiva e giurisprudenza prevedibile*, in *La Nuova Procedura Civile*, 2019, 23 July 2019.

GALLI *et al.*, *Predicting outcomes of Italian VAT decisions*, 188-193, in *35th International Conference on Legal Knowledge and Information Systems (JURIX 2022)*, 14-16 December 14-16, 2022, Saarbrücken, Germany.

GRUNDLER, Giulia *et al.*, *Detecting arguments in CJEU decisions on fiscal state aid*, 143-157, in *ArgMining 2022: 9th Worbshop on Argument Mining*, 2020.

GUIDOTTI, Riccardo *et al.*, *A survey of methods for explaining black box models*, 1-42, in *ACM Computing Surveys (CSUR)*, 51, 5/2018.

HAMILTON, Zachary *et al.*, *Tailoring to a mandate: The development and validation of the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN)*, 1129-1155, in *Justice Quarterly* 39, 6/2022.

HILDEBRANDT, Mireille, *Algorithmic regulation and the rule of law*, in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2128/2018.

– *Grounding Computational 'Law' in Legal Education and Professional Legal Training*, in BROŻEK, Bartos *et al.* (eds.), *Elgar Handbook on Law and Technology*, forthcoming, 2023.

KATZ, Daniel M., BOMMARITO, Michael J., BLACKMAN, Josh, *A general approach for predicting the behavior of the Supreme Court of the United States*, in *PLOS ONE* 12, 4/2017.

LAGIOIA, Francesca, ROVATTI, Riccardo, SARTOR, Giovanni, *Algorithmic fairness through group parities? The case of COMPAS-SAPMOC*, 459-478, in *AI & SOCIETY* 38, 2/2023.

LARSON, Jeff *et al.*, *How we analysed the COMPAS recidivism algorithm*, in *ProPublica*, May 23, 2016.

LASSÈGUE, Jean, GARAPON, Antonie, *Justice digitale: Révolution graphique et rupture anthropologique*, PUF, 2018.

LEIBNIZ, Gottfried Wilhelm Freiherr von, *Leibnizens mathematische schriften* [1646-1716], I, Berlin, 1849.

LIPPI, Marco *et al.*, *CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service*, 117-139, in *Artificial Intelligence and Law*, 27, 2/2019.

LODDER, Arno R., ZELEZNIKOW, John, *Artificial intelligence and online dispute resolution*, 73-94, in Rainey *et al.* (eds), *Online Dispute Resolution: Theory and Practice A Treatise on Technology and Dispute Resolution*, Den Haag, 2012.

LUCIANI, Massimo, *La decisione giudiziaria robotica*, in *Nuovo Diritto Civile*, 3, 1/2018.

MAGLIONE, Valentina, *Pisa allena l'algoritmo che «prevede» le sentenze*, in *Norme e Tributi Plus*, 2021 (<https://ntplusdiritto.ilsole24ore.com/art/pisa-allena-l-algoritmo-che-prevede-sentenze-AERGw6L>)

MARTORANA, Marco, SAVELLA, Roberta, *Giustizia 4.0, come il PNRR supporta la roadmap verso l'innovazione dei tribunali*, in *Agenda Digitale*, 7 giugno 2021 (< https://www.agendadigitale.eu/documenti/giustizia-digitale/giustizia-4-0-come-il-pnrr-supporta-la-roadmap-verso-linnovazione-dei-tribunali/>).

MCCARTY, L. Thorne, *Artificial intelligence and law: How to get there from here*, 189-200, in *Ratio Juris*, 3, 2/1990.

MEDVEDEVA, Masha, VOLS, Michel, WIELING, Martijn, *Judicial decisions of the European Court of Human Rights: Looking into the crystal ball*, in *Proceedings of the conference on empirical legal studies*, 2018.

– *Using machine learning to predict decisions of the European Court of Human Rights*, 237-266, in *Artificial Intelligence and Law*, 28, 2/2020.

– *Rethinking the field of automatic prediction of court decisions*, 195–212, in *Artificial Intelligence and Law*, 31, 1/2023.

MORISON, John, HARKENS, Adam, *Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making*, 618-635, in *Legal Studies*, 39, 4/2019.

PAJNO, Alessandro, *L'uso dell'intelligenza artificiale nel processo tra problemi nuovi e questioni antiche*, 205-223, in *BioLaw Journal – Rivista di BioDiritto*, 1/2022.

PASQUALE, Frank, *The black box society: The secret algorithms that control money and information*, Harvard, 2015.

PASQUALE, Frank, CASHWELL, Glyn, *Prediction, persuasion, and the jurisprudence of behaviourism*, 63-81, in *University of Toronto Law Journal*, 68, 1/2018.

POUND, Roscoe, *Mechanical Jurisprudence*, 605-623, in *Columbia Law Review*, 8, 8/1908.

RISSLAND, Edwina L, ASHLEY, Kevin D., *A case-based system for trade secrets law*, 60-66, in *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, 1987.

ROTOLO, Antonino, SARTOR, Giovanni, *Argumentation and explanation in the law*, in *Frontiers in Artificial Intelligence*, 6, 1/2023.

SANTIN *et al.*, *Argumentation structure prediction in CJEU decisions fiscal state aids*, in *Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023)*, 19-23 June, 2023, Braga, Portugal.

SMOLENSKY, Paul, *On the proper treatment of connectionism*, 1-23, in *Behavioral and Brain Sciences*, 11, 1/1988.

SURDEANU, Mihai *et al.*, *Risk analysis for intellectual property litigation*, 116–120, in *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, 2011.

SUSSKIND, Richard, *Expert systems in law*, Oxford, 1987.

TRAUTMANN, Dietrich, PETROVA, Alina, SCHILDER, Frank, *Legal prompt engineering for multilingual legal judgement prediction*, in *arXiv preprint*, 2022, arXiv:2212.02199.

TURING, Alan M., *Computing machinery and intelligence*, 433-460, in *Mind*, 59, 236/1950.

VAN HARMELEN, Frank, LIFSCHITZ, Vladimir, PORTER, Bruce (eds.), *Handbook of Knowledge Representation*, Amsterdam, 2008.

WINKELS, Radboud *et al.*, *Towards a Legal Recommender System*, 169-178, in *Legal Knowledge and Information Systems, JURIX 2013: The Twenty-Sixth Annual Conference*, 11-13, 2013, Bologna, Italy.

ZEKOS, Georgios I., *Advanced Artificial Intelligence and Robo-Justice*, Cham, 2022.

ZELEZNIKOW, John, STRANIERI, Andrew, *The split-up system: integrating neural networks and rule-based reasoning in the legal domain*, 185-194, in *Proceedings of the 5th International Conference on Artificial intelligence and Law*, College Park Maryland USA, 21-24 May, 1995.