

Daan Kolkman, Floris Bex, Nitin Narayan, Manuella van der Putt\*

DOI: 10.1177/20539517241255101

UDC: 34.096

### Justitia ex machina: The impact of an AI system on legal decision-making and discretionary authority

**Abstract:** Governments increasingly use algorithms to inform or supplant decision-making. Artificial Intelligence systems in particular are considered objective, consistent and efficient decision-makers, but have also been shown to be fallible. Furthermore, the adoption of artificial intelligence (AI) in government is fraught with challenges which are only partly understood and rarely studied in practice. In this paper, we draw on science and technology studies and human computer interaction and report on a critical case study of the development and use of an AI system for processing traffic violation appeal at a Dutch court. Although much empirical work on algorithms in practice is primarily observational in nature, we employ a canonical action research approach and actively participate in the development of the AI system. We draw on data collected in the form of interviews, observations, documents and a user-experiment. Based on this material we provide: 1. An in-depth empirical account of the tensions between street-level bureaucrats, screen-level bureaucrats and street-level algorithms; 2. An analysis of the differences between decisions made by, with and without the AI system and find that use of the AI systems impacts decisions made by legal experts; 3. A confirmation of earlier work that finds AI systems can best be applied in support of legal-decision making and demonstrate how the decision-making process of the traffic violation cases may mitigate some of the risks of algorithmic decision-making.

**Keywords:** legal prediction, discretionary authority, bureaucracy, artificial intelligence, work practices.

#### Introduction

Governments increasingly use algorithms to inform or supplant decision-making [32]. Artificial intelligence (AI) systems in particular are considered objective, consistent and efficient decision-makers [29] that can contribute to some of society's most pressing issues [21]. This techno-optimist narrative has also gained momentum in the legal sector, where organisations such as courts, public prosecutors and attorneys are experimenting with AI [63]. Recent advances with Large Language Models have sparked a renewed interest in AI techniques such as legal prediction, legal analytics, and legal language processing [7; 51].

However, the adoption of AI in government is fraught with challenges [61] and there are increasing concerns about the adverse impact of AI systems on government decision-making [24]. This discourse reflects 'algorithmic drama' [65] with stark polarisation in attitudes towards AI.

---

\* **Daan Kolkman** - Dr. Assistant Professor, Department of Information and Computing Sciences Utrecht University, the Netherlands. E-mail: d.a.kolkman@gmail.com

**Floris Bex** - Prof. Dr. Professor, Faculty of Law, Economics and Governance, Department of Information and Computing Sciences, Utrecht University, Netherlands. E-mail: f.j.bex@uu.nl

**Nitin Narayan** - Jheronimus Academy of Data Science, Eindhoven University of Technology, Eindhoven, Netherlands.

**Manuella van der Putt** - Tilburg Law School, Tilburg University, Tilburg, Netherlands.

Neyland [39] suggests that such drama distracts from the everyday impacts of algorithms and argues for inquiries into specific instances of AI systems<sup>1</sup>. Several scholars have taken up the gauntlet and demonstrated problematic nature of algorithmic transparency [28], identified strategies to resist adoption of algorithms in the workplace [16] and developed oversight strategies for algorithmic systems [62]. What emerges from such studies is not some grand algorithmic spectacle, but mundane accounts of people doing their jobs with – or despite - algorithms.

Building on such studies of algorithms in practice, this paper revolves around a critical case study of the development and use of an AI system for processing traffic violation appeals at a Dutch court. The judiciary presents an interesting background for this study because it has been particularly slow in the adoption of AI [7]. Owing to several high-profile incidents involving algorithms in the judiciary, questions have been raised about the desirability of automated or assisted legal decision-making [10; 41] and its impact on the discretionary authority that is seen as a pillar for individualised justice [14; 16].

While many studies of algorithms in practice are observational in nature, we employ a canonical action research (CAR) [18] approach and actively participate in the development of the AI system. In this paper, we draw on focus groups, interviews, observations and a user experiment conducted as part of the case study. Based on this material we make contributions to two strands of literature on the adoption and use of AI in government. First, we dissect the challenges involved in the adoption of AI systems in government, building on the work of Dwivedi et al. [19]. We show that, in the judiciary data availability and outsourcing of key information systems remain key hurdles to development of AI systems. Second, we explore and assess the tension between discretionary authority and algorithmic decision-making [3] and through the user-experiment demonstrate that the use of an AI system impacts legal decision-making.

### **Artificial intelligence for legal decision-making in the judiciary**

Applications of AI systems<sup>2</sup> in government range from video surveillance to fraud prevention [24; 61]. AI for the legal sector has been around since the rule-based, legal expert systems of the 1980s [57]. While these expert systems had well-recognised shortcomings, the rule-based approach is still in use in domains where the input is discrete and the rules clear<sup>3</sup>. Ultimately, expert systems' failure to live up to the hype led to disillusionment in the legal sector [30]. However, with a legal tech sector that is growing by the year and a new wave of machine learning and natural language processing techniques [7], expectations for legal AI systems are once again high.

Broadly speaking, AI for the legal sector focuses on: 1. Searching and organising information, for example, finding specific clauses in contracts or facts in case files [33]; 2. Summarising legal texts; data analysis and clustering, for example, finding similar cases in a large collection of cases [58] or discovering trends in judicial rulings [6]; 3. Prediction, for example, predicting the risk of recidivism [4], or the outcome of legal cases [36]. Many of these technologies are finding their way to the market, being used by law firms and legal publishers.

#### **Challenges in the adoption of artificial intelligence**

---

<sup>1</sup> Gebru et al. (2023) made a similar call to consider the 'very real and very present exploitative practices' of AI systems.

<sup>2</sup> The terms 'algorithms' and 'artificial intelligence' are often used interchangeably. Here, we look at AI system in the broadest sense, including both data driven machine learning algorithms as well as more classical rule-based (expert) systems (Surden, 2019)

<sup>3</sup> Consider, for example, tax-reporting software, which contains numerous (legal) rules to automatically compute taxes.

The legal sector, and the judiciary in particular, faces a number of challenges regarding the adoption of AI. We briefly discuss these, using Dwivedi et al.'s [19] categorisation of economic, social, technological, data, organisational, ethical and political challenges associated with the development and use of AI systems.

First, there are social and organisational challenges: members of the judiciary are generally conservative with regards to new technology [60], and the judiciary as an organisation is slow to adopt new technologies [7], being long organised around paper files [16] – or its digital equivalent of unstructured text files. In the Netherlands, the large central digitisation effort of the judiciary was cancelled because of a lack of progress, with an evaluation committee citing, among other things, legacy information systems (a technological challenge) and issues with digitising analogue data and processes (a data, technological and organisational challenge) as the main impediments [59]. The Dutch judiciary has been arguing about a lack of funds (an economic challenge) and its dependence on the national budget – whereas in theory the judiciary acts independently from the national government, in practice their ability to act is constrained by the budget as set by parliament (a political challenge).

## 2.2 Prediction, legal decision-making and discretionary authority

The algorithmic drama in the legal field often overlooks the above-mentioned practical challenges of digitisation, funding and data quality. Rather, it centres on ethical and social challenges and questions whether legal experts can be replaced by AI. The discussion about the ‘robot judge’ started when Aletras et al. [2] published their paper on predicting decisions of the European Court of Human Rights based on the verdict texts. This led to speculation that, if such algorithms could predict decisions of judges, they might eventually replace them<sup>4</sup>. Mayson [35] argues there is some merit to the techno-optimist idea that AI systems can arrive at fairer decisions than judges, arguing that judges may generalise to a greater extent, and with less grounding, than AI systems and may overemphasise factors that have particular salience to them. Others [8; 37] claim that decision predictors may improve the predictability and consistency of judicial decision making, which is demanded by the principle of equality [15]. However, there is also the fear that when judges’ decisions are informed by decision predictors, people will not be judged on the legal merits of their individual case but on the basis of general statistics [41].

Here, it should be noted that legal decision-making and algorithmic ‘legal prediction’ are two fundamentally different processes. Legal prediction merely predicts the decision of a case based on, for example, the text of the verdict. This decision, however, is but the final step in a legal decision-making process that puts emphasis on reasoning, motivation and explanation [46; 55]. Ahsmann [1] offers a decomposition of the legal decision-making process in civil law<sup>5</sup>:

1. The inventory phase: The judge<sup>6</sup> starts reading the case file to determine what it is about, what has been claimed, and what legal questions are involved. They determine the applicable rules and case law, list the relevant facts, and consider the propositions put forward by the claimant and the defence.

<sup>4</sup> For example, Artificial intelligence ‘judge’ developed by UCL computer scientists. The Guardian. 24 October, 16. <https://www.theguardian.com/technology/2016/oct/24/artificialintelligence-judge-university-college-london-computer-scientists>

<sup>5</sup> In common law systems the process will be similar, of course with a greater emphasis on case law and legal precedent in Phase 4.

<sup>6</sup> Note that the decision-maker can also be a paralegal (when preparing cases, cf. Diagnosis and action planning Section) or a jury (when considering questions of fact in many common law systems).

2. The selection phase: The judge establishes the facts and lists the points of dispute. They determine the basis for - and substantiation of - the claim that has been put forward.

3. The assessment phase: The judge analyses and assesses the dispute. They determine whether relevant facts have been stated and if these facts are sufficiently substantiated. If this is not the case, the claim will in principle be rejected.

4. The decision phase: The judge applies the relevant legal rules and legal precedents to the case and decides. If the case offers insufficient information to arrive at a decision, they may ask for additional information.

5. The editing phase: The judge motivates their decision and writes a verdict.

There are several legal prediction algorithms that perform part of the above legal decision-making process based on text [20]. While some of these perform tasks from earlier phases, such as identifying the important articles of law in a case (Phase 1), most legal prediction algorithms predict the decision of the court based on, for example, the part of the verdict text that discusses the facts and the law (Phase 4). Although it seems that this kind of algorithm performs some sort of case-based reasoning based on precedents - looking at the (factual, legal) aspects of previous cases to come to a decision in a new case - all they do is learn to identify statistical, possibly spurious, correlations between words in the text and the case decision. Hence, these algorithms cannot provide the legally relevant reasons for the outcome of a case [10; 41].

Even if new, more advanced legal prediction algorithms could provide legally relevant reasons for the (predicted) outcome of a case, there are still several shortcomings. First, there are issues with the ground truth the algorithm is trained on since there is not necessarily a 'correct' decision [25]: a correct prediction of a legally incorrect decision counts as a success for the predictive algorithm [10]. Furthermore, because algorithms are trained on datasets of previous cases, so they cannot quickly adapt to changing legal conditions [41]. Finally, because these algorithms tend to learn the most general rules (patterns, correlations) that appear in the dataset of previous cases, they do not allow for exceptions to a general rule in an individual new case [11]. In other words, the algorithm cannot exercise discretion, which Christin [16] defines as 'the autonomy to decide what should be done for each individual case', which aligns with the principle of individualised justice that is responsive to the individual offender, their background and the offense [5]. So, while legal prediction algorithms might lead to more consistency in decisions, discretion cannot be automated by legal prediction algorithms, since automation prevents individualised judgement [9; 44]. This is especially problematic since discretionary authority is an essential part of what it means to be a judge, it comes with the judge's status as legal experts. In the Netherlands, legal professionals (e.g., judges, prosecutors and lawyers) need to complete an academic master's degree and go through at least three years on the job training before they are sworn in. Legal experts are expected to possess a degree of knowledge about the law and thus are afforded a degree of liberty in how they apply the law. When judges exert their discretionary authority, they may make decisions that have a profound impact on people's lives. Judges are, in Bovens and Zouridis [13] sense 'street-level bureaucrats' that have 'substantial discretion in allocating facilities or imposing sanctions'. In the case of legal prediction algorithms, even if such 'street-level algorithms' are retrained when new data becomes available, their discretion is always retrospective [3].

These limitations of legal prediction have redirected efforts from mere decision prediction towards algorithms as support tools for summarising verdict texts or finding similar cases [64]. The use of algorithms to support bureaucratic processes rather than replace discretionary authority has

gained traction beyond the legal sector [52]. Proponents of this approach contend it may result in more consistent decision-making [8; 9] and call for further research to explore how and in which parts of the decision-making process (i.e., in which of the five phases) AI systems can be most valuable [40; 60]. However, the use of algorithms for decision-support in the legal decision-making process has also been criticised. It may change the division of labour in the legal decision-making process, leading some to speculate whether screen-level bureaucrats may be influenced by the algorithm, exercise non-expert discretion, and adversely affect the process in several ways [11]. The notion here is that such screen-level bureaucrats cannot - or may not be inclined to - offer 'tailormade solutions or exceptions to prevent disproportional negative outcomes' [43] as they are isolated from people and individual cases through AI systems. Furthermore, even though decision-support algorithms for earlier phases of the decision-making process do not directly take away human discretionary authority, they can influence the decision in phase 4 of the process, as this is directly based on what information (e.g., which summaries, similar cases) is presented in the earlier phases.

Following Alkhatib and Bernstein [3], we observe that there has been very little work that revolves around the tensions between street-level bureaucrats, screen-level bureaucrats and street-level algorithms in practice. Amongst other things, it remains unclear how - and where - in the legal decision-making process AI can be useful [40] and how AI system decisions compare to those of judges [35]. In our case study, we engage with these challenges and consider how the use of an AI system impacts the day-to-day work of legal experts in the judiciary and the decisions they make.

### **Methods**

To explore how legal experts use AI systems in practice, we rely on action research conducted in a critical case study. We follow a growing body of literature that studies algorithms in practice using ethnography or associated qualitative methods [16; 28; 52; 53; 62]. Such qualitative inquiry often builds on science and technology studies (STS) and is well suited for the study of work practices and development and use of artefacts [45], such as AI systems, because it permits a rich and detailed description of the research subject and its context [42]. AI systems in particular are not isolated from the broader socio-technical systems of 'regulations, procedures, instruments and institutions' they are part of [49]. Since applications of AI in the (legal) public sector are scarce, identification of an AI system in the initial stages of development and negotiating access to that process would be problematic. We thus opted to deviate from the STS approach in which researchers are embedded in organisations that use AI systems and 'scavenge' [53] material. Rather, we worked with stakeholders to iteratively develop an AI system akin to human computer interaction (HCI) studies. More specifically, our case study has a strong action research component, an approach that has been applied in HCI research [23], but also seeks to move beyond investigation of the intervention in isolation from its context [26]. A variety of action research frameworks exist [50], but we draw on Davison, Martinsons and Kock's [18] CAR framework because it overcomes the lack of methodological rigour in earlier frameworks [17]. CAR prescribes a series of six phases, which may be repeated throughout the research (see Figure 1).

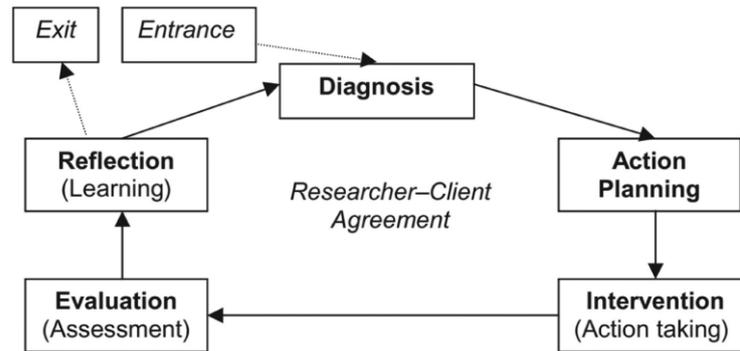


Figure 1. Canonical action research phases (Davison et al., [18]).

Our entrance to this subject was the literature on legal AI systems and a series of three focus groups. The outcome of these focus groups informed our selection of a case study in the legal domain: automation of traffic violation appeals. We then entered the iterative phases of the CAR framework, and over the course of nine months worked closely with paralegals to develop an AI system. In our initial iteration, we interviewed the paralegals, inspected several case files, and attended court proceedings as part of the diagnosis phase. This informed the action planning phase in which we identified user requirements for the AI system. These user requirements were then implemented in the first version of the AI system in the intervention phase. The AI system was then presented to the paralegals as part of an informal evaluation phase, which helped us to develop a better understanding of their needs in the reflection phase. These insights then fed back into a new diagnosis, restarting the process. In the next sections, we report on our case study by collapsing all iterations into a single sequence. We have discussed the relevant literature in Artificial intelligence for legal decision-making in the Judiciary section, Methods section will detail our entrance to the field, our diagnosis of the perceived issues, and Diagnosis and action planning section our action plan. In Development and intervention section, we discuss the development of the intervention in the form of an AI system, Evaluation section presents the results from the evaluation of the AI system, and Reflection and discussion section offers a reflection and discussion.

**Diagnosis and action planning**

We aimed to identify a case study that would help us better understand the challenges of AI system development in government and the impact of AI on discretionary authority. More specifically, we sought to engage with legal experts who routinely deal with cases either in court sessions - Lipsky’s [34] ‘street-level bureaucrats’ - or from behind their computer screen - Bovens and Zouridis’ [13] ‘screen-level bureaucrats’. We conducted three focus groups with twelve domain experts and data specialists. The focus groups were organised over the course of two months.<sup>7</sup> Each 3-hour focus group was moderated, and followed a similar setup in which the participants were invited to respond to a series of statements and questions.<sup>8</sup> The discussion points most salient for this paper include ‘What technologies hold promise for experiments with the judiciary?’ and ‘What areas of law would be most suitable for such an experiment?’ The focus groups were recorded and transcribed, after which we used an iterative coding scheme to distil topics. Using this approach, we

<sup>7</sup> See Appendix A in the online supplemental materials.

<sup>8</sup> See Appendix B in the online supplemental materials.

identified traffic violation appeal cases - colloquially called ‘Mulder-cases’ - as the most suitable focus for our case study.

#### Mulder cases

Minor traffic violations in the Netherlands are governed by the Administrative Enforcement of Traffic Regulations Act, also known as the Mulder Act. In the event of a minor traffic fine such as speeding or parking violations, the public prosecutor hands out an administrative fine, also referred to as the Mulder decision, to the person concerned. The organisation responsible for handling this fine is the Central Judicial Collection Agency. If the person concerned disagrees with the Mulder decision, an appeal can be made, and this appeal process is shown in Figure 2.

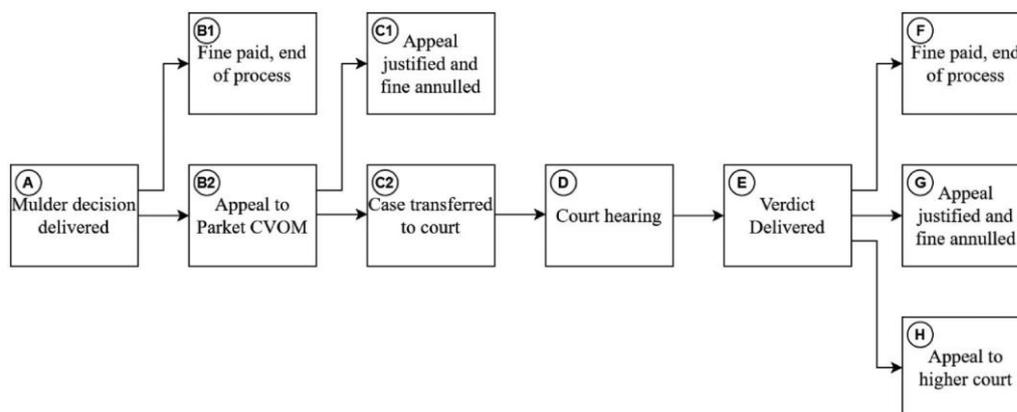


Figure 2. Workflow of the Mulder process. A. The individual receives the Mulder decision. B.1 If in agreement, they pay the fine. B.2 If not, they can appeal with the public prosecutor ‘Parket Centrale Verwerking Openbaar Ministerie’ (Parket CVOM), who then decides. C.1 If the public prosecutor deems the appeal valid, the fine is annulled, and the appeal withdrawn. C.2 If not, the case goes to the sub-district court. D. The court holds a hearing with both parties, concluding with an oral decision. E. The individual receives a written decision. F. If the judge upholds the fine and the individual agrees, the fine remains. The judge also has the discretion to adjust the fine’s amount. G. If the judge deems the fine unjust, it is adjusted or annulled. H. If the appellant disagrees with the verdict, they can appeal at a higher court.

The focus group participants mentioned several reasons for the suitability of Mulder cases for our study. Primarily, they suggest that Mulder cases are simple from a legal perspective. Traffic law stipulates clear fines for specific traffic violations, making the ultimate decision by the judge less dependent on context - and on discretionary authority - than in other areas of law. Moreover, the number of Mulder cases is considerable<sup>9</sup>, and the cases are available in digital format. The focus group participants suggested that this volume of cases would be helpful towards training an AI system. Other reasons mentioned by the participants included the societal relevance of Mulder cases and the workload associated with Mulder cases.

#### Identification of bottlenecks

After having settled on the development of an AI system for Mulder cases, we worked to improve our understanding of that area of law and diagnose the current work practices. We started with a series of meetings with the three paralegals at the court who handle Mulder cases and attended several court proceedings. During the meetings, we discussed our aim of exploring the potential and pitfalls of AI systems in the judiciary with the paralegals and asked them about their work. As

<sup>9</sup> In 2019, 40,200 Mulder cases were registered.

part of a participant validation scheme [12], we then wrote up our understanding of the problem context and the paralegals work and asked for feedback. Over the course of the following seven months, we updated that write-up as we iteratively developed the AI system in close collaboration with the paralegals.

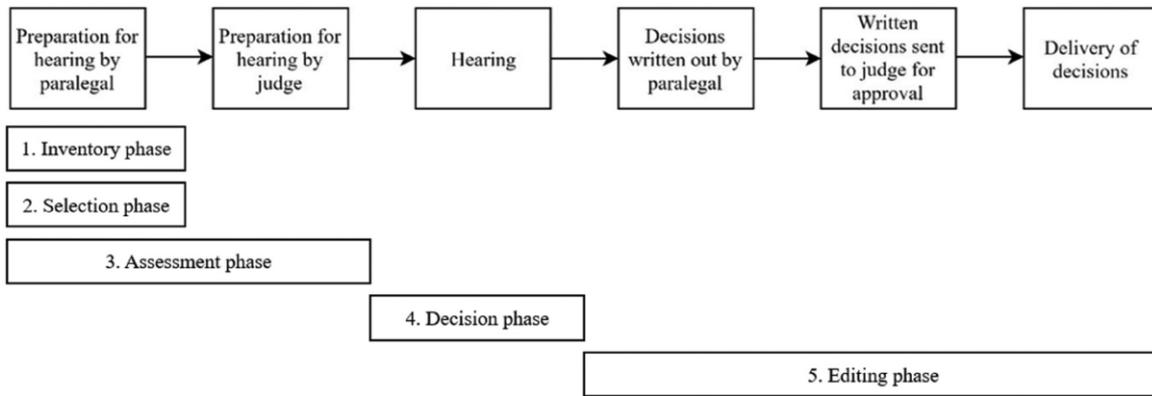


Figure 3. Workflow of the Mulder process at the court.

Figure 3 shows the Mulder case process once the case has been delivered at court at stage C.2 of Figure 2. The paralegals, who are in Bovens and Zouridis’ [13] terms ‘screen-level bureaucrats’, receive a single paper file per case from which they draw all their information. When preparing a case for a hearing, paralegals perform five tasks: checking formalities, extracting case details, studying the appellant’s motivation, identifying similar cases and providing a draft verdict. These five tasks, which we will discuss in more detail below, correspond to the inventory, selection and assessment phases of the legal decision-making process as presented in Prediction, legal decision making and discretionary authority section, and take up most of the time of the paralegals, particularly because of the high volume of cases and the number of pages they need to go over for each case. Below we identify opportunities to improve the effectiveness, efficiency and satisfaction of paralegals when conducting these tasks. We translate these opportunities to user requirements [29] to guide the development of functionalities for the AI system.

**Checking formalities**

The paralegals first check several formalities to see if the case is admissible. First, they look at the dates of the appeal at the OM and the appeal at the court and ensure that these fall within the maximum limit. Then the paralegals proceed to check if the fine has been paid and if this payment has been received on time. If any of these formalities are deemed not to be in order - and the appellant provides no convincing and valid motivation - then the case is considered inadmissible. These checks involve the paralegal skimming through pages of case files to identify the relevant information. The paralegals consider this a monotonous task because of the great number of cases that cross their desk:

A: Look, we often have to prepare dozens of these cases for a single court sitting and we must check all the formalities for each case. Obviously, this is not the most creative and fun part of our job, but you need to be very precise. Once you get to case thirty out of forty, flipping through pages and pages of a case file is plain out boring. So yes, having the system automate all that would be very welcome.

Based on this input by the paralegals, we identified the automated checking of formalities as a first opportunity - or user requirement - to assist the paralegals in their work.

Extracting case details. Once the paralegals have established the case is admissible, they proceed to collect details about it. The most vital details involve the type of traffic violation, the date and time of the violation, the location the violation took place, and the amount of the fine. The paralegals perceive the extraction of such details to be as tedious as the checking of formalities. As such, we have identified the automated extraction of case details as a second user requirement for the AI system.

Studying appellant's motivation. Once the paralegals have drafted an overview of the most important case details, they proceed to study the motivation of the appellant in detail. The paralegals consider this a more intellectually challenging and overall, more rewarding part of their daily work:

B: So with the Mulder cases, it is our job to prepare the case for the court hearing. For these cases, which goes beyond providing an overview, but we try to look at the case from the judge's point of view. Of course, there may be new information provided during the court hearing that changes the perspective on the case. However, in those cases where the appellant does not show up, or no new information is disclosed during the hearing, I think it is rewarding when the judge follows the draft verdict with no, or minor, modifications.

As for any type of court proceeding, the motivation provided by the appellant and any evidence supporting that perspective plays a vital role in the Mulder cases. A sound motivation with ample evidence may provide enough basis for the judge to overturn the decision made by the public prosecutor. To assist the paralegals in this part of the job, the third user requirement we identified was automatically extracting the appellant's motivation from the case file.

Identification of similar cases. In some situations, it may be straightforward for the paralegals to draw up a draft verdict for a Mulder case. Some types of traffic violations occur frequently and are appealed on similar grounds. For instance, speed limit violations observed by a speed trap occur frequently and are sometimes appealed on procedural or technical grounds. A common reason for appealing this type of violation is to doubt the accuracy of the speed trap measurement, one paralegal explains:

A: So for those frequently occurring situations, we really just use standard blocks of texts that refer to the validity of the measurement instrument. Some people will just try to appeal their traffic fine and hope that we cannot prove the accuracy of the speed camera. Usually this information is on hand, so there really is no reason to question the measurements.

However, there are also situations that occur less frequently, or the appellant may have a stringent reason for violating the speed limit. Such reasons may include driving at high speeds to get an injured spouse to the emergency room or violating the speed limit to allow an emergency vehicle to pass by. In such situations the verdict may be less clear-cut. The paralegals will then often look at the verdicts reached in similar cases in the past. Currently, the paralegals do not have a standard system in place to retrieve similar cases, but several paralegals reported working with a custom archive consisting of historical cases grouped by type in folders. As one paralegal explains:

A: So after a while I started organising all my cases a bit better. I have this folder structure where I save finalized cases. For instance, if a case had been about someone running a red light because they had to move out of the way for an emergency vehicle, I would save it as

‘running red light – emergency vehicle – well-founded.’ So, for new cases I could easily find what we decided in the past.

The fourth user requirement we identified was a possibility to help paralegals retrieve historical cases beyond the ones that they have encountered themselves.

Providing a draft verdict. Once the paralegal has worked through the previous steps and has developed a clear overview of the case, they will write a draft verdict in preparation for the court session. This verdict consists of a decision (appeal justified, inadmissible, appeal unjustified, verdict adjusted) and a motivation of the decision. As mentioned above, sometimes the writing of a draft verdict will involve copying and pasting parts from cases that the paralegal previously encountered. This is particularly true for a subset of cases that has been submitted by intermediaries on behalf of the appellant. Such intermediaries advertise online and suggest that about 40% of traffic violations they appeal are either overturned or adjusted. These businesses operate on a no-cure-no-pay basis. When they successfully appeal a case, they claim the reimbursed litigation costs. The paralegals suggest that cases submitted by these intermediaries seem to employ standard templates. The fifth and final user requirement was hence to recommend a draft verdict for paralegals.

### **Development and intervention**

As we began the fieldwork at the court, we set out to acquire the data needed to develop an AI system. At this stage, our understanding of the user requirements was still broad, and primarily based on the outcomes of the three focus groups. We determined early on that the data needed to be in a machine-readable format, and that 10,000 case files would be necessary to develop an AI system.

#### **Data collection**

Gathering such a large number of case files posed challenges. The number exceeded what any single court had handled since 2015, meaning we had to involve more than one court. With the help of the president of the court that had been involved from the start, we managed to convince two other courts to provide clearance for us to access their digital case files.

A key point here is that the courts do not themselves hold digital case files. Rather, they receive relevant paper case files from the public prosecutor’s office every week. It was assumed that the public prosecutor held digital copies of the case files, so we needed to contact the public prosecutor’s office (CVOM). When we first reached out to the public prosecutor’s office in January 2018, it was unclear if they could facilitate a one-time data transfer. The CVOM and the courts could not determine formal ownership of the digital case files: are they the property of the public prosecutor, since they own the database, or are they the property of the court that handles the case? Due to this uncertainty, terms for data transfer and use were established in several agreements covering terms such as privacy and security. Specifically, the option in the GDPR for use of data in the context of scientific research was invoked (Article 24 of the GDPR). Stringent security demands were laid out, which among other things prescribed storage of the hard disk containing the data in a locked safe, processing of the data on a computer with no internet access, and a system for logging access to both.

During the drafting of these agreements another challenge surfaced. Although the public prosecutor’s systems could access digital case files through purposefully developed software, this software did not offer the option to export case files in bulk. This meant we had to contact the consultancy firm that developed and maintained the software for the public prosecutor’s office. This consultancy firm was happy to assist by writing the required queries to export the required number of

cases. Given the high security standards under which the software was developed, this query would have to be run at the data-centre. The consultancy firm argued that this constituted a considerable effort not covered by the existing agreement between the public prosecutor's office and themselves, thus additional costs would be incurred by the public prosecutor's office.

This again sparked a discussion on the ownership of the data and more specifically on who exactly should incur the costs for the bulk export. There was considerable unease among all public parties involved about the fact that access to what were perceived as their case files was dependent on payment to a private sector party (see Prins C. [47]). To avoid further delay to the project, we decided that we would incur the costs.

#### Data exploration

Although the Mulder cases were recommended as one of the most digitised selections of case files in the focus groups, we encountered several challenges once we received the data. While a substantial proportion of the case files contained machine-readable documents – at least documents that could be made machine readable using off-the-shelf optical character recognition technology – this was not so for every case file. As mentioned, the case files can contain handwritten letters, notes and photos, or maps that provide more details about the context of the traffic violation. We resigned ourselves to the fact that we would not be able to use all this material, but upon analysis of the case files we found that the motivation provided by the appellant, the core of the Mulder cases, was often handwritten. We therefore decided to use the motivation of the public prosecutor's office instead. In this motivation, the public prosecutor's office often provides a rebuttal to the points made by the appellant and as such we considered it to be a useful proxy.<sup>10</sup>

#### Evaluation

We built a web-based application that supports paralegals working on Mulder cases. The AI system was developed as a standalone system, integration with an existing system was deemed infeasible and beyond the purpose of exploring the benefits and risks of an AI system in practice. The Upload interface (see Figure C.1 in Appendix C in the online supplemental materials) allowed the users to upload a case file for processing. The Search interface (Figure C.2 in Appendix C in the online supplemental materials) allowed users to search for cases using a free text search, as well as searching for similar cases by uploading a case file.

The five user requirements from Diagnosis and action planning section were each translated into separate functionalities, with varying success. Checking formalities and extracting case details were two requirements that were fully implemented and accessible via the Upload interface, where after uploading a case file the required checks are automatically done and together with the case details shown in the interface. Studying appellant's motivation was possible, but only by opening the original case file via the Upload interface – the interface itself only showed the prosecution's arguments. Identification of similar cases is possible through the search interface, where similar cases can be searched for using either free text search or search based on an uploaded case file. Finally, providing a draft verdict was implemented as a text-based prediction system (like Aletras et al. [2]), where based on the case file, a decision was predicted and shown in the Upload interface.

---

<sup>10</sup> There are several downsides to not using the appellant's motivation directly, most prominently this assumes that the prosecutor's account of the motivation is complete, accurate, and fair which may not always be the case.

Narayan [38] provides a comprehensive technical overview of the AI system. Here, we focus on the evaluation of the AI system and. After having iteratively developed the functionalities in close collaboration with the paralegals, we set up a user-experiment. In the user experiment we presented the paralegals with cases they were not familiar with and asked them to use the AI system to conduct all the tasks they need to do in preparation for a court session. The user-experiment was designed to replicate their daily routines and assess how much the new functionalities would help them.<sup>11</sup> We discuss feedback from paralegals gathered during the user-experiment and subsequent interviews.

#### Process automation over prediction?

All three paralegals welcomed the automation of tasks they consider tedious and monotonous: checking formalities and extracting case details. The paralegals remarked this process would normally take at least several minutes and could now be completed within seconds. The paralegals were also pleased with the functionality that automatically extracts the public prosecutor's motivation from the case file. Again, their primary feedback revolved around the efficiency gains that result from the automated extraction. However, during the user-testing they pointed out that this functionality was less useful to them since the tool included the motivation of the public prosecutor, but not the motivation of the appellant (see Development and intervention section).

The paralegals welcomed the decision prediction functionality during the CAR iterations. However, during and after the user-experiment, they expressed the concern that the underlying classifier was sensitive to certain keywords. For instance, one paralegal noted that the inclusion of words such as 'hospital' or 'emergency services' frequently led the AI system to suggest the public prosecutor's verdict should be overturned. Moreover, the paralegals noted that the use of these automated predictions would not necessarily result in less work for them:

B: Well, while having the system predicting a decision is nice, it does not really help us all that much. Sure, having the automated extraction of dates etcetera is helpful, but writing the draft verdict requires more work than just the decision itself. Even if the system would have been capable of writing entire motivated draft verdicts, it would never be perfect, and we would have to check it. So, then we are back to writing, editing, and copy-pasting and that kind of defeats the purpose.

#### Impact of the AI system on decisions

In the user-experiment, we examined the prediction functionality. Paralegals were instructed to make decisions for two hours without the AI system, and then for another two hours with the AI system. We provided cases such that approximately half of the decisions were made with the AI system (n=43) and half without it (n=46). Notably, eighteen cases were reviewed by two paralegals: one using the AI system and the other not using it. The objective of this setup was to explore the influence of the AI system on decision-making. The AI system achieved a weighted F1-score of 65% on its test set (4436 cases).<sup>12</sup> Eighty-nine random cases out of the test set were covered during the user-experiment. Out of those eighty-nine cases, the AI System correctly identified the judge's decision 46% of the time versus 57% for the paralegals. When assisted by the AI system, paralegals correctly identified the judge's decision 48% of the time (n = 43), compared to 67% of the time

<sup>11</sup> As noted in the methods section, several authors (see Christin, 2017; Kolkman, 2020) have stressed the importance of studying algorithms in practice.

<sup>12</sup> Eighty per cent of the historical case files (13306 cases) were used to train the classifier and 20% of the case files (4436 cases) were used as the test set.

when not assisted by the system ( $n = 46$ ), which suggests that use of the AI system affects the paralegals decision.

Assessment of the paired cases ( $n = 18$ ), in which one paralegal decided with and one paralegal decided without the AI system, shows paralegals agreed on the decision only in seven out of 18 cases. This further points towards some impact of use of the AI system on the paralegal's decision. However, the lack of agreement on paired cases may also reflect the discretionary authority of paralegals even in simple Mulder-cases. Moreover, whether the paralegal (or the algorithm) correctly 'predicts' the judge's decision does not speak to the 'correctness' of the prediction: as judges can make errors (see Prediction, legal decision making and discretionary authority section), and during the hearing new information might have come to light that influences the final decision. Yet, despite the exploratory nature of this study and the small sample size, it suggests use of the AI system impacts the paralegals' decision-making.

We further assess the decisions and motivations provided for the eighteen paired cases (see Appendix E in the online supplemental materials) qualitatively. Examination reveals that in seven of the eleven cases (cases: 2, 5, 8, 9, 10 and 11) where paralegals arrived at different decisions, they cited similar facts. For example, although both paralegals agreed that the public prosecutor was correct in declaring Case 5 inadmissible, their final decisions diverged. Paralegal A maintains the 'inadmissible' decision. Paralegal B felt that the appellant had a valid reason for not paying the guarantee, leading them to evaluate the case on its merits. This assessment resulted in a decision that the fine was justly imposed, thus marking the appeal as 'unjustified'. The discrepancy between the two paralegals' decisions lies in their interpretations of the reasons given by the appellant for non-payment of the guarantee.

This leaves four cases (1, 4, 6 and 12) with a different type of disagreement. In Case 1, Paralegal A argues the appeal is inadmissible due to a missing signature in the appeal letter. Conversely, Paralegal B believes the case is admissible, but argues that the motivation provided by the appellant for speeding ('the traffic signs were not visible') is not valid and reaches an 'appeal unjustified decision'. Cases 4 and 12 pertain to similar situations, in which one paralegal reaches an 'inadmissible' decision and the other arrives at 'appeal unjustified' when considering the motivation of the appellant. In Case 6, the paralegals demonstrate a different interpretation of the facts presented by the appellant. Paralegal C argues that the appellant convincingly argues that they leased their car to someone else and that the offense was registered during that period. Paralegal A argues that the appellant has not convincingly shown that they leased their car at the time of the offense.

All this points towards considerable subjectivity in the Mulder case decision-making process, which echoes earlier findings that courts have much freedom in terms of their decision-making which may improve individual justice but works against the unity of law (cf. Bex and Prakken [9]). When asked to reflect on the impact of the AI system on their decision-making, the paralegals suggested that the AI system helped them to look beyond the case files they previously handled themselves and could therefore result in more unified decision-making. Asked about the potential downsides of using the system, one paralegal remarked:

C: Well, of course the AI system may be biased but it is not like our current system is without its problems. Currently we only consider historical cases that we have worked on ourselves so there is a risk that different paralegals would arrive at different conclusions. I don't think using an AI system to prepare the concept verdicts would adversely affect this. In

any case, we prepare a concept verdict before the court hearing so it is not finalised yet and the judge would still look over it.

The paralegals' folder archiving system (refer to Section Identification of similar cases), as noted in the excerpt, promotes a path-dependent decision-making approach, anchored on past verdicts that an individual paralegal handled. This finding echoes earlier work which suggests legal experts give undue weight to their own experience [35].

### **Reflection and discussion**

With the increasing interest in - and use of - AI systems by government, growing concerns about the adverse impact of AI systems, and the particularly slow adoption of AI systems in the judiciary, we set out to study the development and use of an AI system. Our goal was to cut through the 'algorithmic drama' that has polarised much of the discussion surrounding AI by exploring the impact of an AI system in a practical setting.

To that end, we employed CAR to develop an AI system to support paralegals with the preparation of the most mundane of court cases: traffic violation appeals. These 'Mulder cases' were particularly suited as a critical case study not only because they are numerous and digitised, but also because they are simple from a legal perspective: we expected the impact of discretionary authority on traffic violation appeals is minimal in comparison to other proceedings.

#### Lessons on AI and discretionary authority

We took a qualitative approach to data collection and focused on the everyday work practices of the people we worked with to develop a rich account of the development and use of the AI system. Through close collaboration with the paralegals and judges involved in the so-called Mulder cases, we found that paralegals can be described as 'screen level bureaucrats' [13], while the judges - owing to their direct contact with appellants during the hearing - resembled 'street level bureaucrats' [34]. The AI system we developed, and the 'predict decision' functionality in particular can be considered a 'street-level algorithm' [3].

Through the user-experiment we learned that the AI system we developed was less capable of identifying the judge's decision than the paralegals. This finding seems to echo earlier suggestions that, often, AI systems do not work [48]. More precisely, we see that legal prediction algorithms do not and cannot accurately perform the complex legal reasoning that is required in the decision-making process (Prediction, legal decision making and discretionary authority section, Bex and Prakken [10]). Two further findings also show that using a legal prediction algorithm as a 'monitor' to improve the consistency of decision making [10] might not make sense. First, using the AI system does not make the paralegals' assessment cases more consistent with the judges' assessments (Impact of the AI system on decisions section). Second, the paralegals routinely ignored the predictions of the system (Section Impact of the AI system on decisions), because the system cannot provide any legally meaningful explanation for its predictions. We see that screen-level bureaucrats like paralegals want to remain in control, and do not follow recommendations blindly without an explanation, echoing other recent findings about police case workers as screen-level bureaucrats [54]. The above findings run contrary to the perceived dangers of legal prediction algorithms, particularly the expectation that screen-level bureaucrats engage in less individualised decision-making [11; 43]. Our observations and interviews show that paralegals actively seek tailor-made solutions for the appellants despite being isolated from them by a computer screen until the hearing. Moreover, the paralegal's decisions are but an intermediate step in the Mulder proceedings. A street-level

bureaucrat, the judge, will preside over the hearing in which they can directly engage with the appellant, and may ignore the paralegal's initial draft decision altogether.

So, fears and drama over automated decision making through legal prediction seem to have little footing in practice. Mere legal prediction does not seem to have any clear advantages and represents an intellectual dead-end, especially because it rests entirely on the assumption that judges' decision making is sound and that the law does not change - as Mayson [35] argues, this is a shaky assumption at best. This is of consequence to those seeking to develop AI systems for the legal field: efforts could be more fruitfully directed towards earlier phases in the legal decision-making process. Specifically, we have shown that an AI system can assist paralegals in 1. inventory, 2. selection and 3. assessment phases, by taking over simple tasks such as checking formalities and providing search functionalities that present the paralegal with a wide range of similar cases. We see that paralegals feel that the AI system might help them offer tailor-made solutions by offering them such a larger reference-set of cases to look at. In other words, a paralegal supported by an AI system may be able to perform the initial exploratory stages of the legal decision-making process better than a paralegal alone, as the AI system can perform a wider search lessening the danger that the paralegal might overemphasise factors that they have seen in previous cases [35]. This would offer a fruitful direction in response to Pääkkönen et al.'s [40] suggestion to explore the parts of the process where AI systems can be most valuable. It is probably not to make decisions, or perhaps even support judges directly, but may be useful to support paralegals.

#### Further lessons for adoption of AI systems in the judiciary

During our case study, we faced many challenges that have less to do with the oft-mentioned ethical and social challenges of 'robot judges'. The primary hurdle was accessing and ensuring data quality (data challenge). While the court was, in theory, the owner of the desired data, it did not have the technical infrastructure (technical challenge), nor the means (organisational challenge), or political power (political challenge) to secure access to its own data. This meant we were dependent on external consultants and IT providers to secure an export of the data we needed to develop the AI system. The external parties charged for their support in the development of the tool (economic challenge). Furthermore, to negotiate access to the data, we had to convince several gatekeepers at different organisations. This was not solely a bureaucratic exercise in which we had to prove GDPR compliance (legal challenge), but we also had to convince individual people of the value of this project (political challenge). The problem of access to data might be solved if the data (i.e., the Mulder cases) would have been handled and stored digitally at the court itself in the first place.

#### Concluding remarks

The polarisation in attitudes towards AI sometimes referred to as 'algorithmic drama' cannot not be tackled from a purely theoretical perspective. It requires in-depth qualitative research that sheds light on the people who work on - and with - these systems in practice. In our case study, we have seen that concerns associated with the 'robot judge' in the judiciary are exaggerated at best: legal prediction algorithms cannot accurately perform the complex legal reasoning that is required in the decision-making process. Furthermore, even though recent breakthroughs with Large Language Models [27] may result in legal reasoning that looks convincing, these models still suffer from the same core problems: the absence of clear ground-truths, the changing nature of the legal system, and the inability to make exceptions to a general rule in a new case. Even if such algorithms would be implemented to help, for example, generate draft verdict texts, it seems that the legal deci-

sion-making process and the professional opinion and expertise of paralegals and judges offers some safeguards the discretionary authority some perceive to be at risk.

Research and development of Artificial Intelligence can be more fruitfully directed towards supporting legal experts. The exact influence of such decision-support systems on decision making in the judiciary should be examined in more detail. To enable this however, the judiciary (and more broadly the legal field) should step away from the algorithmic drama and tackle important general, more practical challenges.

### References

1. Ahsmann M. (2011) *De weg naar het civiele vonnis*. Amsterdam: Boom Uitgevers.
2. Aletras N., Tsarapatsanis D, Preoțiu-Pietro D., et al. (2016) Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science* 2(e93): e93.
3. Alkhatib A., Bernstein M. (2019) Street-Level algorithms: A theory at the gaps between policy and decisions. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*, Glasgow, Scotland, 4–9 May 2019, pp.1–13. New York, NY: CHI '19 Paper 530, Association for Computing Machinery.
4. Angwin J., Larson J., Mattu S. and Kirchner L. (2016) Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. In: *Pro Publica*. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
5. Anthony T., Bartels L. and Hopkins A. (2015) Lessons lost in sentencing: Welding individualised justice to Indigenous justice. *Melbourne University Law Review*, 39, 47-73.
6. Ash E. and Chen D.L. (2018) What kind of judge is Brett Kavanaugh. *Cardozo L. Rev. De-Novo HeinOnline*: 70.
7. Ashley K.D. (2017) *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge: Cambridge University Press.
8. Babic B., Chen D.L., Evgeniou T., et al. (2021) A better way to onboard AI. *Post-Print. HAL*. Forthcoming 2021.
9. Bex F., Prakken H. (2021a) Can predictive justice improve the predictability and consistency of judicial decision-making? *Legal knowledge and information systems*. In: *JURIX 2021: The Thirty Fourth Annual Conference: Frontiers in artificial intelligence and applications*, Vilnius, Lithuania pp. 207-214. IOS Press.
10. Bex F., Prakken H. (2021b) On the relevance of algorithmic decision predictors for judicial decision making. In: *Proceedings of the eighteenth international conference on artificial intelligence and law*, Sao Paulo, Brazil, 21-25 June 2021. New York, NY, USA: ICAIL '21. Association for Computing Machinery, 175-179.
11. Binns R. (2020) *Human Judgment in algorithmic loops: Individual justice and automated decision-making*. Regulation & governance (rego.12358). Wiley. Forthcoming October 7, 2020. DOI: 10.1111/rego.12358.
12. Bloor M. (1978) On the analysis of observational data: A discussion of the worth and uses of inductive techniques and respondent validation. *Sociology* 12(3): 545-552.

13. Bovens M., Zouridis S. (2002) From street-level to system level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review* 62(2): 174-184.
14. Brayne S., Christin A. (2020) Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems* 68(3): 608-624.
15. CEPEJ (2018) European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment. Council of Europe Strasbourg. Forthcoming 2018.
16. Christin A. (2017) Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4(2): 2053951717718855.
17. Cohen L., Manion L., Morrison K. (2017) Action research. In: *Research Methods in Education*. London: Routledge, 440-456.
18. Davison R., Martinsons M.G., Kock N. (2004) Principles of canonical action research. *Information Systems Journal* 14(1): 65-86.
19. Dwivedi Y.K., Hughes L., Ismagilova E., et al. (2021) Artificial intelligence (AI): Multi-disciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 57: 101994.
20. Feng Y., Li C., Ng V. (2022) Legal judgment prediction: A survey of the state of the art. In: *Proceedings of the thirty first international joint conference on artificial intelligence, California, July 2022*. International Joint Conferences on Artificial Intelligence Organization. Available at: <http://dx.doi.org/10.24963/ijcai.2022/765>.
21. Floridi L., Taddeo M. (2016) What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083): 1-5.
22. Gebru T., Bender E.M., McMillan-Major A., et al. (2023) Statement from the Listed Authors of Stochastic Parrots on the “AI Pause” Letter.
23. Hayes G.R. (2011) The relationship of action research to human computer interaction. *ACM Transactions on Computer-Human Interaction* 18(3): 1-20.
24. Janssen M., Kuk G. (2016) The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly* 33(3): 371-377.
25. Kang E.B. (2023) Ground truth tracings (GTT): On the epistemic limits of machine learning. *Big Data & Society* 10(1): 205395172211461.
26. Kashfi P., Feldt R., Nilsson A. (2019) Integrating UX principles and practices into software development organisations: A case study of influencing events. *The Journal of Systems and Software* 154: 37-58.
27. Katz D.M., Bommarito M.J., Gao S., et al. (2024) GPT-4 passes the bar exam. *Philosophical Transactions A* 382: 1–35. DOI: <https://doi.org/10.1098/rsta.2023.0254>.
28. Kolkman D. (2020) The (in)credibility of algorithmic models to non-experts. *Information, Communication and Society* 4462: 93-109.
29. Kujala S., Kauppinen M., Rekola S. (2001) Bridging the gap between user needs and user requirements. Citeseer. Available at: <https://citeseerx.ist.psu.edu/viewdoc/download?repid=rep1&type=pdf&doi=10.1.1.103.808>.
30. Lee M.K. (2018) Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5(1): 205395171875668.
31. Leith P. (2016) The rise and fall of the legal expert system. *International Review of Law, Computers & Technology* 30(3): 94-106.

32. Lepage-Richer T., McKelvey F. (2022) States of computing: On government organization and artificial intelligence in Canada. *Big Data & Society* 9(2): 205395172211233.
33. Lippi M., Pałka P., Contissa G., et al. (2019) CLAUDETTE: An automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law* 27(2): 117-139.
34. Lipsky M. (1969) *Toward a Theory of Street-Level Bureaucracy*. Madison, WI: Institute for Research on Poverty, University of Wisconsin.
35. Mayson S.G. (2019) Bias in, Bias out. *The Yale Law Journal* 128(8): 2218–2300.
36. Medvedeva M., Vols M., Wieling M. (2018) Judicial decisions of the European court of human rights: Looking into the crystal ball. In: *Proceedings of the Conference on Empirical Legal Studies, 2018*. martijnwieling.nl. Available at: <https://www.martijnwieling.nl/files/Medvedeva-submitted.pdf>.
37. Muhlenbach F., Sayn I. (2019) Artificial intelligence and law: what do people really want?: example of a french multidisciplinary working group. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, New York, NY, USA, 17 June 2019*, pp.224-228. ICAIL '19. Association for Computing Machinery.
38. Narayan N. (2020) A decision support system for the Court of East Brabant. *Research.tue.nl*. Forthcoming January 6, 2020.
39. Neyland D. (2019) *The Everyday Life of an Algorithm*. London: Palgrave Pivot.
40. Pääkkönen J., Nelimarkka M., Haapoja J., et al. (2020) Bureaucracy as a Lens for analyzing and designing algorithmic systems. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, Hawaii, 25-30 April 2020*, pp.1-14. New York, NY: CHI '20. Association for Computing Machinery.
41. Pasquale F., Cashwell G. (2018) Prediction, persuasion, and the jurisprudence of behaviourism. *University of Toronto Law Journal* 68(Suppl. 1): 63-81.
42. Pedersen M.A. (2023) Editorial introduction: Towards a machinic anthropology. *Big Data & Society* 10(1): 205395172311538.
43. Peeters R. (2020) The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity* 25: 1-16.
44. Petersen A.C.M., Christensen L.R., Hildebrandt T.T. (2020) The role of discretion in the age of automation. *Computer Supported Cooperative Work: CSCW: An International Journal* 29(3): 303-333.
45. Pinch T., Bijker W.E. (1987) *The Social Construction of Technological Systems, Anniversary Edition: New Directions in the Sociology and History of Technology*. Cambridge: MIT Press.
46. Posner R.A. (2010) *How Judges Think*. Cambridge: Harvard University Press.
47. Prins C. (2019) Beschikbaarheid van digitale rechtspraakdata *Juristenblad, Nederlands* 94(38): 2859.
48. Raji I.D., Kumar I.E., Horowitz A., Selbst A. (2022) The fallacy of AI functionality. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21-24 June 2022*, pp. 959-972. ACM
49. Ratner H.F. and Elmholt K. (2023) Algorithmic constructions of risk: Anticipating uncertain futures in child protection services. *Big Data & Society* 10(2): 20539517231186120.
50. Reason P. and Bradbury H. (2008) Concluding reflections: Whither action research. In: *Handbook of Action Research*. London: Sage Publications Ltd, 695-707.

51. Reiling A.D. (2020) Courts and artificial intelligence. *International Journal for Court Administration* 8.
52. Saxena D., Badillo-Urquiola K., Wisniewski P.J., et al. (2021) A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2): 1-41.
53. Seaver N. (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 205395171773810.
54. Soares C., Grimmelikhuijsen S., Meijer A. (2023) Screen-level bureaucrats in the age of algorithms. An ethnographic study of algorithmically supported public service workers in the Netherlands Police. *Information Polity*. DOI: 10.3233/ip-220070.
55. Stobbs N., Hunter D., Bagaric M. (2017) Can sentencing be enhanced by the use of artificial intelligence? *Criminal Law Journal* 41(5): 261-277.
56. Surden H. (2019) Artificial intelligence and law: An overview. *Georgia State University law review*. papers.ssrn.com. forthcoming 2019.
57. Susskind R.E. (1991) *Expert Systems In Law: A Jurisprudential Inquiry*. elibrary.ru. forthcoming 1991.
58. Tran V., Nguyen M.L., Satoh K. (2019) Building Legal Case Retrieval Systems with Lexical Matching and Summarization using A Pre-Trained Phrase Scoring Model. In: *Proceedings of the seventeenth international conference on artificial intelligence and law, New York, NY, USA, 17 June 2019*, pp.275-282. ICAIL '19. Association for Computing Machinery.
59. TRConsult (2018) Quick scan review kei. review op risicobeheersing en basis succescondities voor grote ictra-5 jecten. Available at: <https://www.rechtspraak.nl/SiteCollectionDocuments/2018-rapport-review-commissie>.
60. van der Put M. (2022) Kunstmatige intelligentie bij rechterlijke oordeelsvorming, theoretische analyse en praktische implementatie. *research.tilburguniversity.edu*. forthcoming 2022.
61. Wirtz B.W., Weyerer J.C., Geyer C. (2019) Artificial intelligence and the public sector - Applications and challenges. *International Journal of Public Administration* 42(7): 596-615.
62. Young M., Katell M., Krafft P.M. (2019) Municipal surveillance regulation and algorithmic accountability. *Big Data & Society* 6(2): 205395171986849.
63. Završnik A. (2021) Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology* 18(5): 623-642.
64. Zhong H., Xiao C., Tu C., et al. (2020) How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. *arXiv [cs.CL]*. Available at: <http://arxiv.org/abs/2004.12158>.
65. Ziewitz M. (2016) Governing algorithms: Myth, mess, and methods. *Science Technology and Human Values* 41(1): 3-16.
- Original source – Daan Kolkman, Floris Bex, Nitin Narayan, and Manuella van der Putl. - *Big Data & Society*: <https://journals.sagepub.com/doi/epub/10.1177/20539517241255101>.

*The article was submitted: 2025 November 04*

*Accepted for publication: 2025 November 14*

**Daan Kolkman, Floris Bex, Nitin Narayan, Manuella van der Putl\***

DOI: 10.1177/20539517241255101

UOT: 34.096

**Justitia ex machina: Süni intellekt sisteminin hüquqi qərar qəbuletmə və diskresion səlahiyyətlərə təsiri**

**Xülasə:** Hökumətlər qərar qəbuletmə prosesinə məlumat vermək və ya onu əvəz etmək üçün alqoritmlərdən getdikcə daha çox istifadə edirlər. Xüsusilə süni intellekt sistemləri obyektiv, ardıcıl və səmərəli qərar qəbul edənlər hesab olunur, lakin eyni zamanda səhv edə biləcəyi də sübut edilmişdir. Bundan əlavə, süni intellektin (Sİ) hökumətdə tətbiqi yalnız qismən başa düşülən və praktikada nadir hallarda öyrənilən çətinliklərlə doludur. Bu məqalədə biz elm və texnologiya tədqiqatlarına və insan-kompüter qarşılıqlı əlaqəsinə əsaslanırıq və Hollandiya məhkəməsində yol hərəkəti qaydalarının pozulması ilə bağlı apelyasiya şikayətinin işlənməsi üçün Sİ sisteminin inkişafı və istifadəsinin kritik bir iş araşdırması haqqında məlumat veririk. Praktikada alqoritmlər üzərində aparılan çoxlu empirik iş əsasən müşahidə xarakteri daşısa da, kanonik fəaliyyət tədqiqat yanaşmasından istifadə edirik və Sİ sisteminin inkişafında fəal iştirak edirik. Müsahibələr, müşahidələr, sənədlər və istifadəçi təcrübəsi şəklində toplanan məlumatlardan istifadə edirik. Bu materiala əsasən aşağıdakıları təqdim edirik: 1. Küçə səviyyəli bürokratlar, ekran səviyyəli bürokratlar və küçə səviyyəli alqoritmlər arasındakı gərginliyin dərin empirik hesabatı; 2. Sİ sistemi ilə, onunla və onunsuz verilən qərarlar arasındakı fərqlərin təhlili və Sİ sistemlərinin istifadəsinin hüquq mütəxəssisləri tərəfindən verilən qərarlara təsir etdiyini aşkar etmək; 3. Süni intellekt sistemlərinin hüquqi qərar qəbuletməni dəstəkləmək üçün ən yaxşı şəkildə tətbiq oluna biləcəyini və yol hərəkəti qaydalarının pozulması halları ilə bağlı qərar qəbuletmə prosesinin alqoritmik qərar qəbuletmənin bəzi risklərini necə azalda biləcəyini nümayiş etdirən əvvəlki işlərin təsdiqi.

**Açar sözlər:** hüquqi proqnozlaşdırma, diskresion səlahiyyət, bürokratiya, süni intellekt, iş təcrübələri.

Orijinal mənbə – Daan Kolkman, Floris Bex, Nitin Narayan və Manuella van der Putl. - Big Data & Society: <https://journals.sagepub.com/doi/epub/10.1177/2053951724125510>

*Məqalə daxil olmuşdur: 04 noyabr 2025-ci il*

*Çapa qəbul edilmişdir: 14 noyabr 2025-ci il*

\* **Daan Kolkman** - Dr. Dosent Utrext Universitetinin İnformasiya və Hesablama Elmləri kafedrası (Hollandiya). E-poçt: [d.a.kolkman@gmail.com](mailto:d.a.kolkman@gmail.com)

**Floris Bex** - Prof. Dr. Professor, Hüquq, İqtisadiyyat və İdarəetmə Fakültəsi, İnformasiya və Hesablama Elmləri kafedrası, Utrext Universiteti (Hollandiya). E-poçt: [f.j.bex@uu.nl](mailto:f.j.bex@uu.nl)

**Nitin Narayan** - Jheronimus Məlumat Elmləri Akademiyası, Eindhoven Texnologiya Universiteti, Eindhoven, Hollandiya

**Manuella van der Putl** - Tilburq Hüquq Məktəbi, Tilburq Universiteti, Tilburq, Hollandiya

Даан Колкман, Флорис Бекс, Нитин  
Нараян, Мануэла ван дер Путл♦

DOI: 10.1177/20539517241255101

UOT: 34.096

### **Justitia ex machina: влияние системы искусственного интеллекта на принятие юридических решений и дискреционные полномочия**

**Аннотация:** Правительства все чаще используют алгоритмы для информирования или подмены процесса принятия решений. Системы искусственного интеллекта, в частности, считаются объективными, последовательными и эффективными средствами принятия решений, но они также подвержены ошибкам. Более того, внедрение искусственного интеллекта (ИИ) в правительство сопряжено с проблемами, которые лишь отчасти поняты и редко изучаются на практике. В этой статье мы опираемся на исследования науки и технологий и взаимодействия человека с компьютером и сообщаем о критическом исследовании случая разработки и использования системы ИИ для обработки апелляций по нарушению правил дорожного движения в голландском суде. Хотя большая часть эмпирической работы над алгоритмами на практике носит в основном наблюдательный характер, мы применяем канонический подход к исследованию действий и активно участвуем в разработке системы ИИ. Мы опираемся на данные, собранные в форме интервью, наблюдений, документов и пользовательского эксперимента. На основе этого материала мы предоставляем: 1. Глубокий эмпирический отчет о напряженных отношениях между уличным уровнем бюрократов, экранном уровнем бюрократов и уличным уровнем алгоритмов; 2. Анализ различий между решениями, принятыми с использованием и без использования системы ИИ, и вывод о том, что использование систем ИИ влияет на решения, принимаемые экспертами-юристами; 3. Подтверждение более ранней работы, которая показывает, что системы ИИ могут наилучшим образом применяться для поддержки принятия юридических решений, и демонстрация того, как процесс принятия решений по делам о нарушении правил дорожного движения может снизить некоторые риски алгоритмического принятия решений.

**Ключевые слова:** правовое прогнозирование; дискреционные полномочия; бюрократия; искусственный интеллект; трудовые практики.

Первоисточник – Даан Колкман, Флорис Бекс, Нитин Нараян и Мануэлла ван дер Путл.  
- Big Data & Society: <https://journals.sagepub.com/doi/epub/10.1177/20539517241255101>.

*Дата поступления: 04 ноября 2025 г.*

*Дата принятия в печать: 14 ноября 2025 г.*

♦ Даан Колкман - доктор наук, доцент кафедры информационных и вычислительных наук Утрехтского университета, Нидерланды. E-mail: d.a.kolkman@gmail.com

Флорис Бекс - профессор, доктор наук, профессор факультета права, экономики и управления, кафедры информационных и компьютерных наук Утрехтского университета, Нидерланды. E-mail: f.j.bex@uu.nl

Нитин Нараян - Академия наук о данных имени Иеронима, Эйндховенский технический университет, Эйндховен, Нидерланды

Мануэла ван дер Путл - школа права, Тилбургский университет, Тилбург, Нидерланды