# Master Thesis – Criminal Injustice

## Dutch citizens' preferences on AI versus human judicial decision-making

Supervisor: Dr. A. Poama

Author: Wessel S. van Alphen (3958191)

Public Administration (Public Management and Leadership)

Governance and Global Affairs, Leiden University (Campus The Hague)

June 7th, 2024

Word count: 9,984

# Table of contents

Universiteit Leiden

# 1. Introduction

Advancements in computer and big data technologies have increased the application of artificial intelligence (AI) (Duan et al., 2019). AI can bring various changes and opportunities to the private and public sectors (Wirtz et al., 2018). Considering the potential, the private sector, particularly technology companies such as Google and Microsoft, makes significant investments in AI to explore its application in decision-making processes (Liu, 2021; Wirtz et al., 2018). AI finds its application in various private sector domains, such as dating apps, social media feeds, bank loans, misinformation detection, and problematic social media behaviour (Bambauer & Risch, 2021; Liu, 2021). Consequently, AI has also found its way into the public sector, as the governments of China and the United States recognise its potential for public services (Wirtz et al., 2018). The application of AI in the public sector, particularly in the judiciary, is becoming increasingly significant in terms of legal disputes and the criminal justice system (Aini, 2020; Barysė & Sarel, 2022; Završník, 2020). AI finds its application in the judiciary as a support tool for sentencing and document generation, while also impacting the judicial decision-making process itself (Aini, 2020; Barysė & Sarel, 2022). AI serves as a support tool in legal applications to enhance document screening, online legal services, and even algorithm prediction (Buocz, 2018; Susskind, 2023). Consequently, AI's influence extends beyond the courtroom's physical realm to impact the judicial decision-making process itself (Barysė & Sarel, 2022).

While still in its early stages, the era of AI legal-solvers is anticipated to arrive soon, driven by ongoing advancements in technology (Sourdin, 2018; Susskind, 2013). In China, internet courts already facilitate online dispute resolution with the aid of AI. Similarly, in the state of Wisconsin (United States), judges employ AI to generate recommendations for criminal sentencing (Barysė & Sarel, 2022). The Netherlands is a frontrunner in digitalisation, making it expected that AI will eventually find its way into the Dutch judicial decision-making process (Van Gelder, 2018; Van Wingerden, 2020). Nevertheless, despite the availability of innovations like AI, they have not been implemented on a large scale within the judiciary (Barendrecht, 2019). As mentioned by Van Wingerden (2020), AI has entered the Dutch judiciary, albeit outside of the judicial decision-making processes itself. AI is currently utilised as a risk prediction consultation tool in cases of temporary custody and as a database for consistent sentencing in the Netherlands. Furthermore, AI applications based on data from previous cases and verdicts to estimate sentencing are explored, although these are still far from implementation.

Judges stand to benefit from the implementation of AI in the judiciary (Barysė & Sarel, 2022). Automation can reduce effort and increase efficiency in searching documents, retrieving legal data, and accurately applying the law (Aini, 2020; Barysė & Sarel, 2022; Kulk & Deursen, 2020). Moreover, one of the key benefits of AI lies in its ability to comprehensively review evidence according to the set standards and generate predictions of judgement results, tasks that humans may struggle with due to the complexity of the cases (Aini, 2020; Barysė & Sarel, 2022). This is especially true in complex cases where identifying patterns may be limited by human capacity (Barysė & Sarel, 2022). Additionally, this is particularly important given instances where judges may struggle to maintain consistent sentencing due to external factors like fatigue and emotional instability, influenced by their characteristics (Buocz, 2018; Van Wingerden, 2020). Furthermore, AI can enhance individual legal protection and reduce discrimination by analysing individual characteristics, leading to fairer treatment of cases with varying degrees of inequality (Kulk & Deursen, 2020).

On the contrary, in recent times, many legal and computer science scholars have extensively discussed and criticised the integration of AI in decision-making processes, including the judiciary (Bambauer & Risch, 2021; Barysė & Sarel, 2022). Many have raised normative objections, viewing the algorithms of AI as fraught with risks of discrimination and errors (Bambauer & Risch, 2021). According to Susskind (2023), judicial decision-making, particularly in complex cases involving principles, policy, and morality, is far beyond the capabilities of current computers. Moreover, discrimination can arise from biases embedded by AI programmers. AI's self-learning algorithms with non-representative data can lead to discriminatory outcomes based on societal stigmatisation, stereotyping, and bias related to race, ethnicity, or age (Barysė & Sarel, 2022; Kulk & Deursen, 2020). Objections have been raised regarding how algorithms weigh data and the automation of decision-making itself. Discrepancies in outcomes for similar cases may violate dignity interests, or algorithms may treat individuals similarly despite facing vastly different socioeconomic circumstances (Bambauer & Risch, 2021). Additionally, AI applications may dimmish individual autonomy and legal protection by restricting their options and operating in a manner not easily understood by all, resulting in a ‘black box’ scenario that obscures the stages of judicial decision-making (Bambauer & Risch, 2021; Kulk & Deursen, 2020; Završník, 2020).

## 1.1 Problem statement

While scholars debate AI’s role in criminal sentencing, they generally agree that its implementation significantly impacts the common perception of the legal process (Susskind, 2013; Završník, 2020). Nevertheless, many judicial tasks still rely on human intelligence, as AI cannot replace them or engage with individuals empathetically and responsively (Sourdin, 2018). Critics advocate various reforms to tackle these issues, including improving transparency, imposing fiduciary duties, regulating algorithms, and even prohibiting certain processing methods altogether (Bambauer & Risch, 2021). Similar to other fields, AI can learn to interpret and apply legal principles by analysing legislation and case law, making its application to factual circumstances viable. With advancements in AI and increased investment, the development of more sophisticated AI applications within judicial decision-making is probable within the next decade (Sourdin, 2018). Implementing AI could reduce congestion in civil courts and manage the growing volume and complexity of cases, preventing less-complex cases from being sidelined due to capacity limitations (Yalcin et al., 2022). Access limitations to the judiciary create significant challenges for citizens seeking resolutions. The judiciary’s distance, complexity, and high costs can hinder individuals from seeking justice, ultimately undermining access to justice and creating a serious social problem (O’Donnell, 2004; Susskind, 2023; Yalcin et al., 2022). Failure to report incidents undermines the judiciary’s ability to fulfil its duties, leaving offenders unidentified and unpunished. Affected are less likely to recover losses or receive necessary support (Bosick et al., 2012). Consequently, it is predictable that courts view AI as a potential solution to this issue (Yalcin et al., 2022).

Perceptions of AI’s role in judicial decision-making and trust in the justice system are vital indicators of effective public administration (Yalcin et al., 2022). If citizens believe that judges give unfair sentences or take into account unsuitable considerations, they might start to doubt the integrity of the public administrative system as a whole (Roberts & Plesničar, 2014). Governments and international organisations are actively discussing policies regarding the integration of AI applications in courts, which is pivotal for upholding the rule of law (Yalcin et al., 2022). Public trust in governmental institutions, including the judiciary, influences citizen's willingness to report crimes, highlighting the significance of considering public opinion in social policy decisions, irrespective of regulatory actions (Bambauer & Risch, 2021; Yalcin et al., 2022). Nevertheless, legal scholarship frequently overlooks the examination of citizen's reactions to decision systems involving AI. Yet, understanding these responses is crucial. Failure to do so will ultimately undermine the legitimacy of legal reforms and hinder their implementation (Bambauer & Risch, 2021; Yalcin et al., 2022). As mentioned by Yalcin et al. (2022), public preferences provide information about AI and human errors, helping policymakers understand how the public views the processes involved in decision-making. Opinions regarding the decision-maker (AI versus humans) vary depending on the circumstances. In decision-making processes with limited consequences, anomalies do not have significant effects, leading citizens to be more accepting of AI decision-makers (Filiz et al., 2023). However, a poor decision in certain situations, such as judicial proceedings, can have serious consequences, making citizens less inclined to trust AI with high-stakes decisions (Filiz et al., 2023; Mahmud et al., 2022; Susskind, 2023). This highlights the need for policymakers to approach high-stakes decision-making processes differently from those with lower stakes (Bambauer & Risch, 2021).

Considering the Netherlands leading position in digitalisation, it is expected that AI will eventually integrate into the Dutch judicial decision-making process (Van Gelder, 2018; Van Wingerden, 2020). Helberg et al. (2020) conducted a survey experiment with a Dutch sample to investigate human perceptions of algorithmic decision-making applications. However, their research was broad and did not focus on any specific sector. In contrast, this study focuses on the Dutch judiciary, examining not just perceived fairness, a part of preferences, but preferences as a whole (Yalcin et al., 2022). Building on existing literature and addressing the potential for criminal injustice, it is crucial to investigate whether trust in AI and human judges varies with the complexity (stakes at risk) of legal cases.

This inquiry results in the following research question:

*'Do Dutch citizens preferences for utilising AI versus human decision-making in the judiciary change when the stakes at risk vary in terms of incident gravity and severity of punishment?'*

## 1.2 Scientific relevance

This study contributes to the increasing amount of literature on public involvement in scientific research, which has been stimulated by ongoing technological advancements as more people use online applications (Lotfian et al., 2021). It offers empirical insights into how citizens (laypeople) respond when presented with a choice between AI and human judicial decision-makers. While prior studies by Yalcin et al. (2022) and Bambauer and Risch (2021) examined citizen preferences regarding AI in courtrooms using a sample of US citizens, their findings might not directly apply to the Dutch population. As mentioned by Yalcin et al. (2022), as cross-country characteristics may influence general trust in the judiciary, potential differences may impact public trust in judges. For instance, AI judicial decision-making may be trusted more in nations with low judicial trust and vice versa. Consequently, there are notable differences in the general level of trust and acceptance of AI systems between nations. Western-European nations with lower acceptance rates of AI systems include the Netherlands (18%) (Gillespie et al., 2023). Additionally, factors such as the case itself, legal culture, presence of attorneys, and prior experience influence public perception of the judiciary (Yalcin et al., 2022). This research seeks to address this gap by investigating citizen perceptions of AI and human judicial decision-making within the context of the Dutch jurisdiction. Nevertheless, this study aligns with prior research by Yalcin et al. (2022), Bambauer and Risch (2021), and Sela (2018), which generally suggests less citizen favourability towards AI decision-making. This research contributes to the ongoing discourse regarding the application of AI in judicial decision-making (Osborne, 2010). Particularly, this study contributes to the scientific discourse in the Netherlands regarding the possible integration of AI into the judicial decision-making process, as empirical research-based discourse is a fundamental component of scientific inquiry (Hardy et al., 2010).

## 1.3 Societal relevance

As outlined in paragraph 1.2, this study offers empirical insights into the preferences of Dutch citizens regarding judgments delivered by AI versus human judges. However, the significance of this research extends beyond academic discourse; it holds crucial relevance in societal contexts, especially for public administrators and software engineers. Notably, the focus here is on the Netherlands, a country that has received limited attention in prior studies. Public administrators can use the findings of this study to optimise the integration of AI within the judicial system, build public trust, raise awareness of court digitalisation, and conduct further investigations involving stakeholders (Yalcin et al., 2022). They can utilise survey data to collect factual information, including subjective evaluations of administrative services, such as those within the judiciary (Stipak, 1980). This is essential for considering public opinion in social policy decisions and, consequently, the legitimacy of reforms (Bambauer & Risch, 2021; Yalcin et al., 2022). As detailed in paragraph 1.1, less-complex cases are increasingly sidelined from court proceedings due to capacity constraints, leading to underreporting and compromising the efficiency of the criminal justice system, thereby exacerbating challenges for individuals seeking resolutions (Bosick et al., 2012; Yalcin et al., 2022). In the Netherlands, accessibility to the judiciary is declining due to rising registry expenses, lengthy waiting times, and complex procedures. Consequently, there is a growing interest in implementing IT solutions, such as AI, to decrease these burdens on the judicial system for citizens (Van Gelder, 2018). However, before incorporating social policy, such as potentially introducing AI into the Dutch judiciary, it is essential to consider the opinions of citizens, regardless of whether policymakers ultimately heed or override them. Failure to do so will undermine the legitimacy of legal reforms and obstruct their implementation. This research will contribute to this topic by providing a dataset on the extent of consumer preferences regarding how citizens perceive AI in the judiciary (Bambauer & Risch, 2021). Consequently, software developers can utilise this research to understand the elements that citizens consider problematic or beneficial, in what types of cases, and conduct their studies to determine how to develop their products, as this research only focuses on preferences rather than specific application characteristics (Yalcin et al., 2022).

# 2. Theoretical framework

In this chapter, the different theories utilised in this research will be examined. The theories will be discussed separately, later the theories will be discussed with each other. This section will serve as the basis for answering the research question and the hypotheses.

## 2.1 AI sentencing

In the coming decades, the court system will undergo significant transformation, altering how legal services are provided. Despite the legal profession's slow adoption of new technology, it is steadily progressing (Susskind, 2013). As emphasised by Sourdin (2018), legal technology can be classified into three approaches. Firstly, it assists and advises individuals within the judiciary. Secondly, it substitutes tasks traditionally performed by humans with technology. Finally, it introduces disruptive technology that fundamentally challenges and modifies long-standing practices in the judiciary, thereby fundamentally altering the nature of the work that judges do (Sourdin, 2018; Susskind, 2013). This includes utilising predictive analysis to redefine the adjudicative role. AI is a crucial element that currently typically falls outside of the sentencing process and usually falls into the first two categories of legal technology. For instance, first-level legal technology implementations may involve the online provision of justice-related information and processes (Sourdin, 2018). Some technologies may advance to the third-level, implementing AI sentencing by fully replacing the human decision-making process. As emphasised by Ryberg & Roberts (2022), the court is an important public institution, and punishment needs to be carefully considered. The state's power over the lives of citizens reaches its peak during the sentencing process. Legal punishment encompasses various forms of deprivation of property and liberty that would be considered wrongful treatment under normal circumstances. These potential consequences necessitate a compelling justification.

The complexity of AI-based sentencing arises from the ambiguity surrounding the scope of its technological application, whether it involves replacing human decision-makers or serving as a substitute (Ryberg & Roberts, 2022; Sourdin, 2018). Firstly, the term AI lacks a precise definition, as there is no universally accepted definition (Nowotko, 2021; Ryberg & Roberts, 2022). Some interpret AI as technology capable of replicating human intelligence accurately, prompting philosophical inquiries about whether such systems can achieve human-like thought processes and consciousness. Conversely, AI in a weaker sense refers to systems that execute tasks without emulating human reasoning. This framework includes a wide range of computer-driven algorithms, from simple to complex machine-learning models (Ryberg & Roberts, 2022). As mentioned by Jussupow et al. (n.d.), the algorithmic approach underlying AI should be described as mathematical models integrated with problem-solving methods employed by computers. However, other terms are also utilised to refer to this concept, namely the mathematical approach, automated system, algorithm, computer program software, recommender system, machine, AI, and supercomputer. The algorithmic approach should be understood as a computer agent that employs rule-based or non-rule-based (machine-learning) methods to generate an output.

Employing AI in sentencing encompasses diverse applications (Ryberg & Roberts, 2022). AI sentencing could operate as an (online) service utilising vast amounts of structured and unstructured legal data. It would understand legal issues communicated in natural language, identify patterns in these issues, make deductions, provide legal advice, and deliver this advice through a computer-generated voice (Susskind, 2013). As emphasised by Ryberg and Roberts (2022), AI can serve as a tool to assist judges during sentencing, or it may entirely supplant human decision-making. Alternatively, various forms of collaboration between human judges and algorithms can be foreseen. AI could handle decisions in minor offences, while human judges retain authority over sentences in more serious or intricate cases. At one end of the spectrum, AI utilisation in sentencing might involve a basic algorithm aiding a judge in considering a single sentencing factor. On the opposite end of the spectrum, Araujo et al. (2020) and Ryberg & Roberts (2022) state that it might involve a completely automated judge delivering verdicts via technological processes, utilising an expanding array of personal data, which algorithms subsequently analyse to make independent, data-driven decisions, without human intervention.

This study will centre on the implementation of AI in sentencing procedures within the Dutch judicial system. AI sentencing, as referred to here, entails a fully automated robotic judge delivering judgments through technological means. This process involves the utilisation of an increasingly diverse set of personal data, which algorithms subsequently analyse to make impartial, data-driven decisions (sentences) independently of human intervention (Araujo et al., 2020; Ryberg & Roberts, 2022). This relies on an AI system designed with human-like cognitive processes and awareness (Ryberg & Roberts, 2022). This AI system incorporates a non-rule-based algorithmic machine learning model, derived from historical judicial rulings (Jussupow et al., n.d.). Such AI sentencing constitutes a distinct category of legal technology, marking a significant shift in the role of judges and potentially altering the fundamental nature of justice itself (Sourdin, 2018).

## 2.2 Algorithm aversion

In numerous fields, both experts and citizens exhibit resistance to adopting algorithm-driven AI applications (Dietvorst et al., 2015; Jussupow et al., n.d.). Even though AI is increasingly the better choice in terms of output quality in terms of speed and accuracy, they frequently prefer to rely on less competent human decision-makers (Dietvorst et al., 2015; Filiz et al., 2023; Jussupow et al., n.d.). Early-stage disruptive technologies like AI sentencing are often dismissed as superficial and unlikely to succeed (Susskind, 2023). Research shows resistance from citizens and decision-makers to using algorithms in decision-making, preferring to assign responsibilities to human specialists or handle themselves (Araujo et al., 2020; Filiz et al., 2023). This behaviour illustrates a common phenomenon referred to as algorithm aversion, indicating a tendency to reject automated systems, such as AI decision-makers (Alon-Barkat & Busuioc, 2022; Dietvorst et al., 2015). Nevertheless, this is not always the case, as research by Alon-Barkat & Busuioc (2022) did not find consistent evidence for a universal tendency toward automation bias. Moreover, some decision-makers appreciate algorithmic support as they prefer to adjust to algorithmic judgment instead of human judgment (Jussupow et al., n.d.). As emphasised by Filiz et al. (2023), in decision-making processes with limited consequences, anomalies do not have significant effects, and individuals are more likely to opt for algorithmic decision-makers. In low-stakes scenarios like using a dating app or getting a weather forecast, the worst outcomes are minor, such as meeting an unsuitable candidate or wearing inappropriate clothing. While some prefer algorithms for certain tasks, they may avoid them in high-stakes decisions like judicial decision-making, where errors can have serious consequences (Filiz et al., 2023; Mahmud et al., 2022). Therefore, understanding the drivers of algorithm aversion behind such a decision is of great importance in maximising the merits of algorithms (Mahmud

Universiteit Leiden

et al., 2022). Jussupow et al. (n.d) mentioned that algorithm aversion stems from comparing computational methods to intuitive approaches. Despite lacking a universal definition, it describes the reluctance to use algorithms once flaws are noticed. However, some individuals exhibit aversion even before noticing imperfections. A broader definition sees algorithm aversion as a biased assessment, leading to negative attitudes and behaviours toward algorithms compared to human agents. Consequently, Filz et al. (2023) state that it is only possible to speak about algorithm aversion once the algorithm provides the highest quality result or probability of success. Others consider algorithm aversion present as soon as subjects exhibit fundamental disapproval of an algorithm despite its possible superiority.

Mahmud et al.'s (2022) algorithm aversion framework comprehensively examines various factors shaping individuals' attitudes and behaviours regarding algorithmic decision-making. These factors span psychological, personality, familiarity, demographic, algorithmic, task-related, and high-level dimensions. This research focuses on task-related factors, such as decision domain and experience. In domains characterised by risk and volatility, individuals often prefer human decision-making, even when highly effective algorithms are available, as the context of the situation can significantly influence these decisions (Filiz et al., 2023; Mahmud et al., 2022). As said the phenomenon of algorithm aversion is influenced by various factors, particularly in the context of complex and less complex decision-making scenarios and the stakes at risk (Mahmud et al., 2022). In complex cases with high stakes and serious consequences at risk, individuals tend to exhibit greater aversion to algorithmic decision-making, even though algorithms have the potential to provide more accurate and effective solutions (Filiz et al., 2023; Mahmud et al., 2022; Susskind, 2023). The framing of a case significantly influences the choice between an AI or human decision-maker, as citizens tend to be more risk-averse when losses are emphasised (Levy, 1992). As the perceived stakes rise, citizens prefer human decision-making, likely due to feelings of uncertainty about the results and potential consequences (Bambauer & Risch, 2021; Mahmud et al., 2022). Citizens are therefore less likely to trust AI with making important decisions, potentially stemming from a bias toward losses, as they prioritise avoiding losses over the potential but unknown gains of a superior algorithmic decision-maker (Levy, 1992; Mahmud et al., 2022). Mahmud et al. (2022) indicate that in high-risk finance, demand forecasting, and medical decision-making, individuals generally prefer human advisors over algorithms. Research by Yalcin et al. (2022) shows that trust in algorithm judges is lower in complex legal cases than in straightforward ones. Conversely, studies regarding less risky decisions, such as selecting media feeds, have shown a preference for algorithms perceived as less risky in this context (Araujo et al., 2020).

Mahmud et al. (2022) noted that the degree of algorithm aversion is influenced by demographic factors like age, gender, and educational attainment. Different age groups of citizens have different perspectives on algorithms. For instance, whereas younger generations accept and trust algorithms more, older generations believe algorithms are less useful and therefore less reliable (Gillespie et al., 2023; Mahmud et al., 2022). Furthermore, studies in a variety of fields, most notably justice, have shown that women are more averse to making arbitrary decisions, supporting the notion that gender plays a role in algorithm aversion (Mahmud et al., 2022). Moreover, citizens who achieve lower levels of education also value algorithms less because educated citizens are more likely to trust and accept algorithms than people without a university degree (Gillespie et al., 2023; Mahmud et al., 2022).

To implement algorithm-driven AI judicial decision-making in the Dutch legal system, this study attempts to investigate algorithm aversion (Dietvorst et al., 2015; Jussupow et al., n.d.). Even though algorithms may be objectively superior, algorithm aversion is the tendency to distrust or disapprove of them, which results in negative attitudes and behaviours toward them, especially when compared to human decision-makers (Filiz et al., 2023). In particular, since citizens typically prefer human decision-making when the stakes are higher and choose algorithms in less risky scenarios, this research specifically investigates whether people exhibit algorithm aversion in AI sentencing (Araujo et al., 2020; Bambauer & Risch, 2021).

## 2.3 Hypotheses

The theory discussed above has led to the following two hypotheses:

*$H_1$: Citizens tend to favour AI decision-making over human judgment when the stakes at risk are low in judicial cases in the Netherlands.*

*$H_2$: Citizens tend to favour human decision-making over AI judgment when the stakes at risk are high in judicial cases in the Netherlands.*

# 3. Research design

The purpose of this research is to investigate whether there exists a variation in the preference among the Dutch population for legal cases to be adjudicated by either an algorithmic decision-maker, AI sentencing, or a human decision-maker, depending on the fluctuation of stakes concerning the case gravity and punishment severity (collectively, the stakes at risk). Therefore, this study employs the following research question:

*'Do Dutch citizens preferences for utilising AI versus human decision-making in the judiciary change when the stakes at risk vary in terms of incident gravity and severity of punishment?'*

To address the research question, this study employs quantitative methods in conjunction with a between-subjects design and an experimental (qualitative) vignette survey. A binomial logistic regression model is employed to determine the relationship between Dutch citizens' preferences for an AI or human judicial decision-maker. Additionally, chi-square tests are utilised to identify any significant associations between demographic variables and the choices made by participants in this study. This chapter will explain the research methods and design in more detail.

## 3.1 Methodology

### 3.1.1 Experimental vignette survey

This study employs an experimental vignette survey to address the research question. Utilising the vignette methodology instead of a conventional questionnaire prevents bias and unreliable self-reports from participants, making the research suitable for studying human attitudes and behaviour (Alexander & Becker, 1978). To elicit responses or gather information on specific phenomena, vignette studies involve providing participants with concise, detailed descriptions of individuals, behaviours, situations, or events. These descriptions are usually provided in written or visual formats (Törrönen, 2018). A vignette study consists of two main components: the vignette experiment, the focal point, and a conventional survey to measure respondent demographics. These demographic factors serve as covariates in the analysis (Atzmüller & Steiner, 2010). Vignette studies offer participants carefully constructed and real-world scenarios, enabling the evaluation of dependent variables like intentions, attitudes, and behaviours. Which improves experimental realism and increases researchers' ability to control and modify independent variables more successfully (Aguinis & Bradley, 2014). Vignette studies are useful for analysing Dutch citizens' preferences for AI or human decision-making in judicial sentencing. By adjusting the stakes (case gravity and punishment severity), four scenarios can be created. In a between-subjects design, each respondent is exposed to only one scenario, ensuring they receive the same vignette with differing independent variables. With these designs, causal claims are obtained by comparing the behaviour of individuals in one experimental condition with that of individuals in another (Charness et al., 2012).

### 3.1.2 Vignette randomisation

The survey software tool for this research, Qualtrics, was employed to measure citizen preferences regarding the utilisation of AI in sentencing. By utilising Qualtrics' randomisation feature, the application ensured that the four scenarios were randomly and equally distributed, employing varying stakes at risk to detect any shifts in citizen preference (Bambauer & Risch, 2021). Opting to present respondents with just one scenario in a study, rather than multiple scenarios, strategically prevents participants from guessing the study's purpose and potentially altering their responses to conform with social norms rather than expressing genuine opinions. This is vital for preserving the research's integrity, ensuring that responses reflect individual viewpoints rather than external influences (Aguinis & Bradley, 2014). By doing so, it becomes possible to make causal claims about the effects of the vignette factors on the outcome variable (Steiner et al., 2017). While this approach may lead to lower response rates per vignette, compared to exposing participants to all vignettes, it mitigates the risk of making the treatment too conspicuous, which could bias the results (Aguinis & Bradley, 2014).

### 3.1.3 Vignette design

In survey studies, the wording of vignettes is crucial for eliciting accurate responses from participants. Vignettes, which consist of varying scenarios, serve as brief descriptions to gather respondent's judgments (Atzmüller & Steiner, 2010). Given the sample's focus on Dutch citizens, conducting the survey in Dutch was essential. These vignettes must be clearly written to ensure participants can easily comprehend the scenarios presented to them, as a lack of disciplinary knowledge and language proficiency among respondents can obstruct communication (Chereni et al., 2020). As indicated by Hupe. (2019), a vignette's language and presentation should be straightforward. This helps reduce ambiguity and ensures that respondents accurately interpret and respond to the situations. Following the vignette, respondents are asked to indicate their preference for which adjudicator they would favour in their specific hypothetical scenario.

As stated in paragraph 3.1.2, respondents receive a vignette with one of the two options for each independent variable. The combinations of the independent variables are listed in Table 3.1. Due to its length, only the first scenario is presented below; the other combinations of independent variables can be found in Appendix 1 (English) and Appendix 2 (Dutch).

*Table 3.1 – Independent variable combinations (scenario's)*

| | | **Case gravity** | |
|---|---|---|---|
| | | **High-gravity accident** | **Low-gravity accident** |
| **Punishment severity** | **High-severity punishment** | Scenario 1 | Scenario 3 |
| | **Low-severity punishment** | Scenario 2 | Scenario 4 |

<u>*Scenario 1:*</u>

*'Please review the following information carefully:*

*You were on your way home from work during rush hour. Due to the traffic jam, you wanted to inform your family about your late arrival. You began to grab your phone and write a message to them. Unfortunately, the text message distracted you from paying attention to the road, resulting in your car colliding with someone else's vehicle and causing severe injuries to the driver of that car. This leads to the conclusion that the accident occurred due to your fault, resulting in a relatively high-level punishment: a potential civil traffic fine of €420 (for using your phone while driving), along with a two-month prison sentence and a one-year suspension from driving (for causing a serious traffic accident due to negligence). Following the court's decision, you have been convicted of 'Causing a serious traffic accident with negligence.' After your conviction, you are informed that a new program has been implemented offering you two options for selecting your sentence: human decision-making by a traffic court judge or algorithmic decision-making by an AI system incorporating a machine learning model based on past decisions in similar cases.*

*What option do you choose?'*

The differences between the independent variables have been clarified by emphasising the case severity (high-gravity versus low-gravity) and the punishment severity (high-severity versus low-severity). In the hypothetical case of a high-gravity accident, an individual caused a serious traffic accident due to negligence while using their phone, resulting in severe injuries to the driver of the other car. According to the Ministerie van Justitie en Veiligheid (2023), the 'Wegenverkeerswet 1994 (WVW 1994)' (the road traffic law) outlines the guidelines for prosecuting traffic accidents and dangerous traffic behaviour. The directive of the Dutch Public Prosecution Service follows this prosecution policy. The corresponding sanctions for each of the four scenarios are listed in Table 3.2.

*Table 3.2 – Corresponding sanctions for each vignette design, independent variable combination, according to the Dutch Public Prosecution Service* (Ministerie van Justitie en Veiligheid, 2023)

| **Scenarios (S)** | | **Corresponding sanctions for each scenario** |
|---|---|---|
| **High-gravity accident** | **S1: High-severity punishment** | 420 EUR fine for phone usage while driving<br>Two-month prison sentence<br>One-year suspension from driving |
| | **S2: Low-severity punishment** | 420 EUR fine for phone usage while driving<br>1,000 EUR for causing a serious traffic accident due to negligence |
| **Low-Gravity accident** | **S3: High-severity punishment** | 420 EUR fine for phone usage while driving<br>100 EUR fine for causing a minor accident<br>Two-month suspension from driving |
| | **S4: Low-severity punishment** | 420 EUR fine for phone usage while driving<br>100 EUR fine for causing a minor accident |

### 3.1.4 Individual consent and demographics

Before presenting the vignette, respondents were asked to consent to the use of their provided data for research purposes (Appendix 1). This is in line with researchers' ethical and professional obligations to safeguard respondent privacy and ensure informed consent. These measures encompass data linkage, passive data collection, and the archiving of replication data (Plutzer, 2019). Because anonymous surveys tend to reveal more socially inappropriate attitudes, beliefs, and behaviours, it is noteworthy that the questions are structured to prevent data from being linked to individuals, thereby enhancing respondent honesty (Lelkes et al., 2012). A country's sample should be nationally representative in terms of gender, age, and education (Gillespie et al., 2023). As stated by Hughes et al. (2016), following the vignette, respondents were asked demographic questions to encourage inclusivity, improve sample description for clarity, and facilitate generalisation and possible replication of findings. While demographics may provide insights into behaviour, it is essential to recognise that identity does not dictate actions. By collecting demographic information, researchers can verify if the survey reached the intended participants (Dutch citizens) and assess if the sample comprehensively represents the target population. The demographic questions are outlined in Appendix 1.

To address both $H_1$ and $H_2$, the investigation following the vignette examines the preferences of Dutch citizens regarding their preferred adjudicator in hypothetical scenarios. This enables the observation of shifts in their perceptions based on the stakes involved. As theorised in Chapter 2, individuals tend to display algorithmic aversion in AI sentencing, particularly as the stakes at risk increase. Studies suggest that citizens typically favour human decision-making as risks increase, while showing a preference for algorithms in less critical situations (Araujo et al., 2020; Bambauer & Risch, 2021).

### 3.2 Respondents

The survey exclusively targeted Dutch nationals with no additional participation criteria other than being part of the Dutch population. The survey was initiated on May the 2nd 2024 and remained active until May the 6th 2024. It was distributed through various channels, including direct and anonymous channels to personal contacts and students across different institutions. Participants were urged to share the survey link with their contacts . Upon closure, a total of N = 98 respondents completed the survey, each presented with the same hypothetical vignette but with different variable options (scenarios). However, 2 respondents were excluded for failing whether they hold the Dutch nationality, resulting in a final analysis pool of N = 96 respondents. The distribution of the survey link in Qualtrics was randomised for each respondent. As stated by (Sandelowski, 1995), to guarantee that the sampling strategy utilised in qualitative research is adequate, a large enough sample size is crucial. Claims of achieving theoretical saturation or information redundancy are not supported by small-sized samples.

This study employed a convenience sample, where individuals meeting the study criteria were identified in any feasible manner, alongside snowball sampling, where individuals were encouraged to share the research with their contacts (Emerson, 2015). Since these methods primarily resulted in educated and young respondents, purposive sampling was also employed. This involved targeting less educated respondents to achieve a more representative sample of the Dutch population (Campbell et al., 2020). However, while the convenience and snowball sampling approaches facilitated the attainment of a sufficient number of respondents, they may have influenced the outcome. Utilising these sampling methods introduces unexpected uncontrolled variables, such as individuals from similar geographical areas, socio-economic statuses, or ethnic backgrounds, resulting in a less representative sample of the full Dutch population (Emerson, 2015). While these sampling types have limitations in drawing statistically significant conclusions, they provide valuable insights into a range of attitudes and opinions, including consumer perceptions of algorithm use in the judiciary (Galloway, 2005). A significant proportion of respondents completed a form of higher education, either at a university (WO) or university of applied sciences (HBO), with 60,83% holding Bachelor's and Master's degrees combined. This percentage is notably higher than the national average, as of 2023, where 36.4% of the population had higher education qualifications (Ministerie van Onderwijs, Cultuur en Wetenschap, 2024). The respondent group skewed towards a younger demographic, with 58,33% respondents falling within the 18-24 age group. This demographic representation is not reflective of the Dutch population, where, as of 2024, the 18-24 age group constituted only 9% (Centraal Bureau voor de Statistiek, n.d.). Additionally, the sample exhibited a slight gender imbalance, with 42.71% men and 56.25% women. As of 2021, the Dutch population consisted of 49.7% men and 50.3% women (VZinfo, n.d.). The obtained data regarding the demographics can be found on the next page (Table 3.3).

*Table 3.3 – Demographics of survey respondents*

| **Demographic factors** | | | |
|---|---|---|---|
| **Age** | **Percentage %** | **Frequency** | **Sample size (N)** |
| Under 18 | 1.04% | 1 | 96 |
| 18-24 | 58.33% | 56 | |
| 25-34 | 9.38% | 9 | |
| 35-44 | 8.33% | 8 | |
| 45-54 | 11.46% | 11 | |
| 55-64 | 5.21% | 5 | |
| 65-74 | 4.17% | 4 | |
| 75-84 | 2.08% | 2 | |
| Over 84 | 0.00% | 0 | |
| | | | |
| **Gender** | **Percentage %** | **Frequency** | **Sample size (N)** |
| Male | 42.71% | 41 | 96 |
| Female | 56.25% | 54 | |
| Prefer not to say | 0.00% | 0 | |
| Other | 1.04% | 1 | |
| | | | |
| **Highest educationa** | **Percetage %** | **Frequency** | **Sample size (N)** |
| None | 0.00% | 0 | 96 |
| Primary education | 5.21% | 5 | |
| Secondary education | 31.25% | 30 | |
| Bachelor's degree | 53.13% | 51 | |
| Master's degree | 8.33% | 8 | |
| Ph.D. | 0.00% | 0 | |
| Doctorate | 0.00% | 0 | |
| Prefer not to say | 2.08% | 2 | |

## 3.3 Data analysis

This study employs a binomial logistic regression model to analyse the vignette survey data. This model predicts whether an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables, which can be either continuous or categorical. Specifically, $H_1$ and $H_2$ are tested to evaluate whether the independent variables, namely, accident severity and punishment gravity, affect the single independent variable: Dutch citizen's preference between AI and human judicial decision-makers, as outlined in paragraph 4.1. The binomial logistic regression test has been performed in the software application of SPSS (UCLA, n.d.). Moreover, the study utilises the chi-square test to assess independence across two categorical variables, which has been performed in the software application of

Microsoft Excel (Franke et al., 2011). As mentioned by Mahmud et al. (2022), age, gender, and level of education are examples of demographic factors that influence algorithm aversion. Generating the necessity to investigate the potential relationship between the three distinct categorical variables of age, gender, and education and the categorical dependent variable that represents Dutch citizens' preferences for AI judges as opposed to human judges.

### 3.4. Validity and reliability

Validity refers to accurately measuring the intended concept (Fitzner, 2007). Another critical aspect, reliability, assesses whether a research instrument consistently yields the same results under similar conditions across multiple uses (Heale & Twycross, 2015). Ensuring internal validity is crucial for this study due to its use of vignettes as a data collection method. This involves crafting narratives that are authentic, plausible, and reflective of real-world social scenarios, thus enhancing realism compared to traditional surveys (Jenkins et al., 2020; Steiner et al., 2017). For instance, using a real-world problem like a traffic dispute, along with clear scenarios of incident severity and punishment severity, establishes the stakes involved, thereby ensuring high internal validity (Steiner et al., 2017). This increased internal validity contributes to construct validity by aligning theoretical concepts with specific measurement procedures (Fitzner, 2007). Consequently, the vignette design produces desired outcomes, as the high internal validity supports valid inferences regarding the causal relationships between presented stimuli and respondents' reactions to them (Steiner et al., 2017). However, the external validity, which concerns the extent to which findings can be applied to other groups or settings, remains vulnerable (Fitzner, 2007). As mentioned in section 3.2, the sample utilised is not representative of the entire Dutch population, given significant disparities in education level, age, and gender compared to the real-world population, thereby lowering internal validity.

## 4. Results

This chapter will present the results of the statistical testing of the hypotheses and the potential association of demographic factors with Dutch citizens' preferences regarding their choice between human and AI judicial decision-makers. The hypotheses will be tested using the binomial logistic regression model (paragraph 4.1). The association of demographic factors with Dutch citizens' preferences for judicial decision-makers will be examined using the chi-square test to explore the potential associations of age, gender, and education with the preferred option of judicial decision-making (human versus AI) (paragraph 4.2).

### 4.1 Effects of varying stakes at risk

This section will focus on testing the following two hypotheses:

*$H_1$: 'Citizens tend to favour AI decision-making over human judgment when the stakes at risk are low in judicial cases in the Netherlands.'*

*$H_2$: 'Citizens tend to favour human decision-making over AI judgment when the stakes at risk are high in judicial cases in the Netherlands.'*

Dutch citizens were asked to provide their reactions to vignettes with varying independent variables (accident gravity and punishment severity), resulting in the utilisation of a between-subject design. For each vignette with varying variables, respondents were asked, 'What option do you choose?' Figure 4.1 presents the output data from the between-subject design survey, with the dependent variable being the type of judicial decision-maker (AI versus human). Each case reflects varying independent variables and the associated stakes at risk.

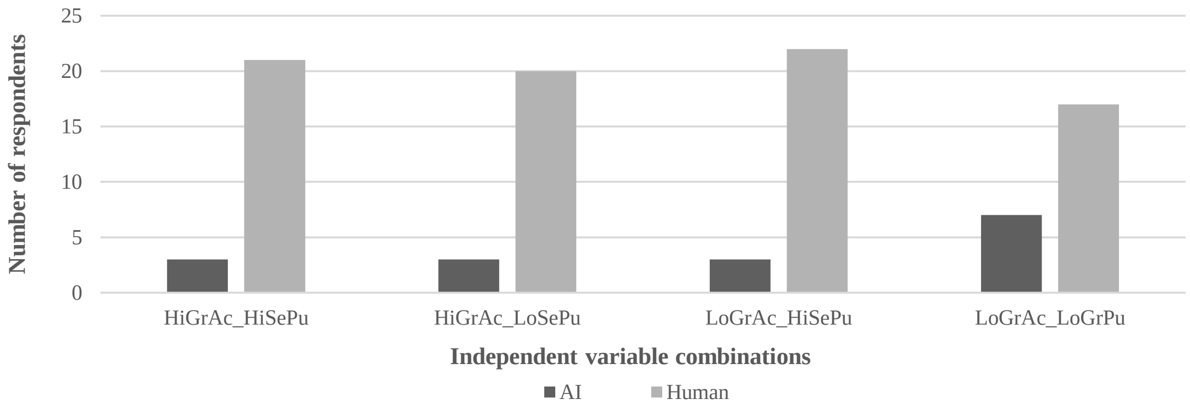


*Figure 4.1 – Preferred judicial decision-maker (AI versus human) for each independent variable combination*

The y-axis of figure 4.1. shows the number of respondents for each type of decision-maker for each combination of independent variables. The x-axis labels represent the combinations: HiGrAc_HiSePu (High-gravity accident, High-severity punishment), HiGrAc_LoSePu (High-gravity accident, Low-severity punishment), LoGrAc_HiSePu (Low-gravity accident, High-severity punishment), and LoGrAc_LoSePu (Low-gravity accident, Low-severity punishment).

A binomial logistic regression model was required to determine whether different independent variables, namely accident gravity and punishment severity, affect a single categorical independent variable (Dutch citizens' preference between AI and human judicial decision-makers). The preferred judge type is a dichotomous dependent variable (a nominal variable with only two values). This model predicts the likelihood that an observation will fall into one of two categories based on one or more categorical independent variables. SPSS Statistics was utilised to perform the binomial logistic regression test in this study. SPSS generates multiple output tables during the binomial logistic regression analysis, which are presented in Appendix 3 The following paragraphs will clarify the two primary tables (4.1 and 4.2) to aid in understanding the statistical test results.

*Table 4.1 – Model summary, SPSS output*

Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 83.22 | 0.03 | 0.06 |

The first step involved determining how much of the variance in the dependent variable, Dutch citizens' preferences for AI or human judges, could be explained by the model. This amount of variance is represented by the $R^2$ coefficient in multiple regression. Table 4.1 (Model Summary) should be consulted to accomplish this. Cox & Snell $R^2$ and Nagelkerke $R^2$ are two techniques used to measure variation. Nonetheless, it is better to use Nagelkerke $R^2$ for this purpose because it is an adaptation of Cox & Snell $R^2$. The Nagelkerke $R^2$ of .06 indicates that approximately 6% of the variability in the dependent variable is accounted for by the independent variables (the scenarios presented to the respondents with varying independent variables). In other words, the model only explains a small portion of the variability in the choice between options A and B (AI and human judicial decision-maker) based on the situation. Thus, these factors are unable to explain the remaining 94%. Mahmud et al. (2022) have proposed a framework that encompasses various variables such as psychological, personality, familiarity, demographic, algorithmic, task-related, and high-level dimensions. These variables can impact an individual's attitude toward the utilisation of AI.

*Table 4.2 – Binomial logistic regression: Accident gravity and punishment severity, affect Dutch citizen preference*

Variables in the Equation

| | | | | | | | 95% CI for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | B | *S.E.* | Wald | *df* | *p* | Exp (B) | Lower | Upper |
| LoGrAc_LoSePu | | | 3.42 | 3 | 0.332 | | | |
| HiGrAc_HiSePu | -1.06 | 0.76 | 1.92 | 1 | 0.165 | 0.35 | 0.08 | 1.55 |
| HiGrAc_LoSePu | -1.01 | 0.76 | 1.74 | 1 | 0.187 | 0.36 | 0.08 | 1.63 |
| LoGrAc_HiSePu | -1.11 | 0.76 | 2.1 | 1 | 0.147 | 0.33 | 0.07 | 1.47 |
| Constant | -0.89 | 0.45 | 3.9 | 1 | 0.048 | 0.41 | | |

Note. Number of participants = 96; Cl = confidence interval; Lower = lower limit; Upper = upper limit

The results of the binomial logistic regression model are presented in Table 4.2. The initial column represents the combinations of independent variables. The binomial regression assesses the likelihood of an event occurring. As shown in the *p*-column, which indicates potential significance, there is no significant variation in the choices Dutch citizens make between human and AI judicial decision-makers, as all scores exceed the *p*-value threshold of .05. ($p > .05$). Consequently, both $H_1$ and $H_2$ are statistically rejected since, in the utilised sample, the independent variables (accident gravity and punishment severity, stakes at risk collectively) do not significantly influence the dependent variable (Dutch citizens' preference between AI and human judicial decision-makers). The respective *p*-values are: LoGrAc_LoSePu: $p = .332$, HiGrAc_HiSePu: $p = .165$, HiGrAc_LoSePu: $p = .187$, and LoGrAc_HiSePu $p = .147$, all of which exceed the threshold of $p > .05$. However, it is noteworthy that the *p*-value for the constant term is $*p = .048$, indicating significance as it is less than .05. This suggests an initial preference for the human judicial decision-maker when no independent variables are influencing the decision.

## 4.2 Demographic effects

Chi-square needs to be utilised to highlight the potential influence of demographics, such as age, education, and gender, on the preference for either AI or human judicial decision-makers (Franke et al., 2011; Mahmud et al., 2022). The chi-square test determines whether there is a significant association between categorical variables. In this case the dependent variable of Dutch citizen's preference between AI and human judicial decision-makers and the categorical variables, the demographics, age, gender, and education. For each chi-square test, it is needed to determine a null hypothesis and an alternative hypothesis.

### 4.2.1 Association between age and judicator

As stated by Mahmud et al. (2022), the perception of AI decision-makers varies among different age groups. Younger individuals are more inclined to accept AI applications compared to older individuals (Gillespie et al., 2023; Mahmud et al., 2022). This underscores the necessity of determining whether there is a significant association between Dutch citizens' preferences for AI or human judicial decision-makers and the age of respondents.

The null hypothesis and alternative hypothesis:

*$H_0$: 'The age of the respondents is not associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*$H_1$: 'The age of the respondents is associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*Table 4.3 – Chi-square of the association between age and preferred judicial decision-maker*

| Observed (O) | | | |
|---|---|---|---|
| | AI | Human | |
| Under 18 | 0 | 1 | 1 |
| 18-24 | 11 | 45 | 56 |
| 25-34 | 1 | 8 | 9 |
| 35-44 | 1 | 7 | 8 |
| 45-54 | 3 | 8 | 11 |
| 55-64 | 0 | 5 | 5 |
| 65-74 | 0 | 4 | 4 |
| 75-84 | 0 | 2 | 2 |
| | 16 | 80 | 96 |

| Expected (E) | | |
|---|---|---|
| | AI | Human |
| Under 18 | 0.167 | 0.833 |
| 18-24 | 9.333 | 46.667 |
| 25-34 | 1.500 | 7.500 |
| 35-44 | 1.333 | 6.667 |
| 45-54 | 1.833 | 9.167 |
| 55-64 | 0.833 | 4.167 |
| 65-74 | 0.667 | 3.333 |
| 75-84 | 0.333 | 1.667 |

| $(O-E)^2/E$ | | |
|---|---|---|
| | AI | Human |
| Under 18 | 0.167 | 0.033 |
| 18-24 | 0.298 | 0.060 |
| 25-34 | 0.167 | 0.033 |
| 35-44 | 0.083 | 0.017 |
| 45-54 | 0.742 | 0.148 |
| 55-64 | 0.833 | 0.167 |
| 65-74 | 0.667 | 0.133 |
| 75-84 | 0.333 | 0.067 |

| | |
|---|---|
| $X^2$ | 3.948 |
| *df* | 7 |
| *p*-value | 0.786 |

As presented in Table 4.3, the p-value of this chi-square test is greater than .05 ($p > .05$), as $X^2$ $(7, N = 96) = .786$, $p > .05$. Consequently, the null hypothesis ($H_0$) is accepted, and the alternative hypothesis (H1) is rejected, indicating no significant association between the age of respondents and the Dutch citizens' preferences for AI versus human judicial decision-makers.

#### 4.2.2 The association between gender and judicator

As emphasised by Mahmud et al. (2022), the perception of AI decision-makers varies among different genders. Research, notably in the field of justice, has shown that women are more averse to AI applications compared to men. This underscores the necessity of determining whether there is a significant association between Dutch citizens' preferences for AI or human judicial decision-makers and the gender of respondents.

The null hypothesis and alternative hypothesis:

*$H_0$: 'The gender of the respondents is not associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*$H_1$: 'The gender of the respondents is associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*Table 4.4 – Chi-square of the association between gender and preferred judicial decision-maker*

| Observed (O) | AI | Human | |
|---|---|---|---|
| Male | 10 | 32 | 42 |
| Female | 5 | 48 | 53 |
| Other | 1 | 0 | 1 |
| | 16 | 80 | 96 |

| Expected (E) | AI | Human |
|---|---|---|
| Male | 7 | 35 |
| Female | 8.833 | 44.167 |
| Other | 0.167 | 0.833 |

| $(O-E)^2/E$ | AI | Human |
|---|---|---|
| Male | 1.286 | 0.257 |
| Female | 1.664 | 0.333 |
| Other | 4.167 | 0.833 |

| | |
|---|---|
| $X^2$ | 8.539 |
| *df* | 2 |
| *p*-value | 0.014 |

As stated in Table 4.4, the p-value of this chi-square test is smaller than .05 ( * $p < .05$), as $X^2$ (2, $N = 96$) = .014, $p < .05$. Consequently, the null hypothesis ($H_0$) is rejected, and the alternative hypothesis ($H_1$) is accepted, indicating a significant association between the gender of respondents and the Dutch citizens' preferences for AI versus human judicial decision-makers.

#### 4.2.3 Association between education and judicator

As emphasised by Mahmud et al. (2022) and Gillespie et al. (2023), the perception of AI decision-makers varies among citizens with different levels of education. Citizens with lower levels of education are less likely to trust and accept AI compared to citizens with a university degree. This underscores the necessity of determining whether there is a significant association between Dutch citizens' preferences for AI or human judicial decision-makers and the educational level of the respondents.

The null hypothesis and alternative hypothesis:

*$H_0$: 'The education of the respondents is not associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*$H_1$: 'The education of the respondents is associated with Dutch citizen's preference for AI or human judicial decision-makers.'*

*Table 4.5 - Chi-square of the association between education and preferred judicial decision-maker*

| Observed (O) | | | |
|---|---|---|---|
| Educatior | AI | Human | |
| Primary | 0 | 5 | 5 |
| Secondar | 7 | 23 | 30 |
| Bachelor | 8 | 43 | 51 |
| Master | 1 | 7 | 8 |
| No say | 2 | 0 | 2 |
| | 18 | 78 | 96 |

| Expected (E) | | |
|---|---|---|
| | AI | Human |
| Under 18 | 0.938 | 4.063 |
| 18-24 | 5.625 | 24.375 |
| 25-34 | 9.563 | 41.438 |
| 35-44 | 1.500 | 6.500 |
| 45-54 | 0.375 | 1.625 |

| (O-E)²/E | | |
|---|---|---|
| | AI | Human |
| Under 18 | 0.938 | 0.216 |
| 18-24 | 0.336 | 0.078 |
| 25-34 | 0.255 | 0.059 |
| 35-44 | 0.167 | 0.038 |
| 45-54 | 7.042 | 1.625 |

| | |
|---|---|
| $X^2$ | 10.754 |
| *df* | 4 |
| *p*-value | 0.029 |

As emphasised in Table 4.5, the p-value of this chi-square test is smaller than .05 ( * $p < .05$), as $X^2$ (4, $N$ = 96) = .029, $p < .05$. Consequently, the null hypothesis ($H_0$) is rejected, and the alternative hypothesis ($H_1$) is accepted, indicating a significant association between the educational level of respondents and the Dutch citizens' preferences for AI versus human judicial decision-makers.

# 5. Discussion

This study examined Dutch citizens' preferences for AI or human judges (the dependent variable) in a hypothetical vignette scenario with four possible combinations of two independent variables (punishment severity and accident gravity, together stakes at risk). The objective was to find out if Dutch people prefer AI decision-making over human judgment in cases with low stakes and, on the other hand if they prefer human decision-making over AI judgement in cases with high stakes.

Three chi-square tests were performed to determine whether demographics affected the sample of Dutch citizens used to determine whether they preferred an AI or a human decision-maker. Chi-square tests are utilised to assess associations between categorical variables (Franke et al., 2011). External validity concerns the degree to which results can be generalised to other populations (Fitzner, 2007). According to Mahmud et al. (2022), the degree of algorithm aversion is influenced by demographic variables, including age, gender, and obtained educational level. In contrast to earlier studies by Gillespie et al. (2023) and Mahmud et al. (2022), which reported a significant age gap in AI acceptance, the first chi-square test found no significant association between the age of respondents and the preferred judicial decision-maker. However, since the age of the respondents has no significant impact on the results, this increases the external validity of the sample that was utilised and, as a result, the reliability of the statistical results. Additionally, a significant association was found in the second chi-square test assessing the relationship between gender and preferred decision-making. This finding is consistent with the findings of Gillespie et al. (2023), which indicated that women, particularly in the judiciary, are less inclined than men to make decisions utilising algorithms. But since respondents' gender has a significant impact on the results, this lowers the external validity of the sample that was utilised and, in turn, the reliability of the statistical results. Finally, the third chi-square test examining the relationship between education and preferred decision-maker also revealed a significant association, in line with studies by Mahmud et al. (2022) and Gillespie et al. (2023), which indicate that individuals with a university degree accept AI more than people without one. However, since respondents' obtained educational level significantly affects the outcomes, this lowers the external validity of the sample utilised and, consequently, the reliability of the statistical results.

Overall, there is a definite preference for human judicial decision-makers when comparing the survey output data. However, the combination of independent variables with the lowest stakes at risk shows a slight, relatively small increase. Given that Dutch citizens consistently oppose the use of AI in the court, it appears that their preferences are not greatly influenced by the independent variables. This lack of significant influence is confirmed by the binomial logistic regression model's results, which show that these independent variables have no discernible impact on the dependent variable of a judge's decision. Consequently, both research hypotheses, $H_1$ and $H_2$, need to be rejected, as Dutch citizens do not base their preferences on the level of stakes at risk. This study builds upon previous research on whether citizens prefer AI or humans in the judiciary. Prior research suggests that individuals exhibit algorithm aversion in AI sentencing, with a preference for human decision-making as stakes at risk increase (Araujo et al., 2020; Bambauer & Risch, 2021). In high-stakes cases like traffic disputes, individuals prefer human decision-makers due to uncertainty about outcomes and concerns over potential consequences, leading to decreased trust in algorithmic decisions. (Mahmud et al., 2022).

Consequently, in a hypothetical scenario involving a minor traffic accident with no injuries (low-gravity accident) and a traffic fine (low-severity punishment) (collectively a low-stakes situation), individuals would be more inclined to choose an AI decision-maker. Nevertheless, upon examining the significance numbers of the binomial logistic regression model, it becomes evident that there is no significant association between the preferred judicial decision-maker and the stakes at risk. This lack of significance could potentially be attributed to a type-II error. A type-II error indicates a false negative, it would mean in this context, that failing to find an association between the preferred judicial decision-maker and the stakes at risk, even though there might be one in the real population, following Chapter 2 (Banerjee et al., 2009). Nonetheless, these research findings close a knowledge gap about the perception of the Dutch population regarding the potential utilisation of AI in the Dutch judiciary. While future research on this topic is needed, the results of this study suggest that the general AI aversion theory may not apply to the sample population, as there is no significant difference in preferences between high and low stakes at risk.

# 6. Limitations and further research

The stakes at risk is an important factor in the assessment of whether Dutch citizens prefer an AI or human judicial decision-maker; however, it is far from the only factor that may influence the acceptance or rejection of AI in the judiciary. First, of all, the application of AI in judicial decision-making affects the resources needed, in terms of efficiency and effort, to retrieve data, search through documents, and apply the law to a legal case (Aini, 2020; Barysė & Sarel, 2022; Kulk & Deursen, 2020). Citizen preference is relatively sensitive to merits as speed, and consequently costs, often prioritised over the accuracy of the decision-maker (Bambauer & Risch, 2021). Other factors, such as level of transparency and equality of treatment can also be at stake. However, these factors are often questioned by scholars as to which type of decision-maker they belong (Bambauer & Risch, 2021).

Erfanian et al. (2020) and O'Dell et al. (2012) highlight that respondents may interpret the vignettes utilised in this study differently or incorrectly, and these vignettes can never fully capture the real-world elements being examined. Participants in vignette surveys are often presented with hypothetical scenarios, which may not accurately reflect actual situations. The method of obtaining information through vignettes, and the assumptions participants make during interpretation, can influence their responses, and thus the results of the vignettes (Erfanian et al., 2020). Additionally, as Erfanian et al. (2020) emphasised, vignettes typically focus on isolated scenarios, thereby oversimplifying reality. Consequently, respondents' reactions may not fully reflect the nuances and complexities of their actual decision-making processes. Further research might consider to utilise video or real-life events instead of written vignettes, as these could provide deeper insights into individuals' behaviours and the situations they encounter.

Consequently, as indicated by Emerson (2015), because convenience and snowball sampling techniques introduce uncontrolled random variables like individuals from similar geographic areas, socioeconomic statuses, and ethnic backgrounds, the results may not be generalisable to larger populations or contexts. This is because the sample utilised is not formally representative of the Dutch population of policy interest, resulting in the research producing biased outputs for the targeted population (Emerson, 2015; Olsen & Orr, 2016). The sample utilised for this research consisted largely of female, educated, and young individuals, potentially influencing the research outcomes. Previous research by Mahmud et al. (2022) suggests that demographic variables such as age, gender, and educational level impact the degree of algorithm aversion. As emphasised by Mahmud et al. (2022) and Gillespie et al. (2023), these findings partially support the results of the chi-square tests conducted in this research, which identified gender and educational level as factors influencing preferences for AI and human judicial decision-making. Consistent with the results of this study's chi-square test, individuals with a university education are more willing to accept AI applications. Moreover, younger generations are also more receptive to AI applications, although this trend was not evident in the sample of this research, as the chi-square test found no significant association between age and preferred judicial decision-making (Gillespie et al., 2023). This necessitates further research into the utilisation of an adequate representation of the population of the policy of interest before public administrators set policy (Olsen & Orr, 2016).

Moreover, the research has been conducted in the Netherlands, thus limiting the scope of the study. As stated by Gillespie et al. (2023), there are significant differences in the acceptance of AI systems across countries. For instance, countries with the highest levels of AI acceptance include India (67%), China (66%), Brazil (54%), and South Africa (48%). In contrast, the Netherlands has one of the lowest acceptance scores for AI systems at only 18%, which aligns with the generally low acceptance scores observed in Western countries. This confines the research to the Netherlands and other Western countries with similar scores, highlighting the need for additional research to explore public perception regarding the application of AI in the judiciary.

**Policy implications**

This research provides the incentive for further investigation into the perception of Dutch citizens towards the utilisation of AI in judicial decision-making. The following policy implications should be considered:

- *Analyse the perception of citizens:* it is required to conduct further research on the Dutch population because the methods utilised in this study produced a sample that is not representative of the entire Dutch population and is therefore giving a biased representation of the complete population (Emerson, 2015; Olsen & Orr, 2016).

- *The need of public consultation:* it is of essence to comprehend citizen perceptions to gauge public attitudes towards potential policy implementation. Neglecting this aspect could ultimately erode the legitimacy of legal reforms, like the implementation of AI judicial decision-making (Bambauer & Risch, 2021; Yalcin et al., 2022).
- *Investigate other types of legal technology:* when further research confirms that the Dutch population is averse to AI judicial decision-making, alternative legal technologies should be considered, such as online dispute resolution and online legal guidance (Susskind, 2023).
- *Prioritise citizens:* as mentioned by Susskind (2023), legal professionals may be disrupted by new technologies like AI sentencing. However, it's crucial to consider the citizens' perspective, as new techniques can lead to lower costs and greater convenience.

## 7. Conclusion

Scholars debate on AI's (Artificial Intelligence) role in judicial decision-making, but generally agree that the implementation of AI significantly affects the current perception of the legal process (Susskind, 2013; Završník, 2020). As stated by Bambauer & Risch (2021), critics advocate for significant reforms to tackle the problems related to AI implementation in judicial decision-making, creating the potential development of more sophisticated AI sentencing within the next decade is likely (Sourdin, 2018). AI implementation could alleviate congestion in civil courts, preventing less-complex cases from being overlooked due to capacity limitations (Yalcin et al., 2022). Consequently, failure to report incidents undermines the legal system's ability to fulfil its duties, impacting the affected (Bosick et al., 2012). However, courts view AI as a potential solution to this issue (Yalcin et al., 2022). Public perceptions of AI in judicial decision-making and trust in the justice system are crucial for effective governance (Yalcin et al., 2022). Governments and international organisations discuss policies regarding the integration of AI judges, pivotal for upholding the rule of law (Yalcin et al., 2022). Public trust influences citizen's willingness to report crimes, underscoring the importance of considering public opinion in social policy decisions (Bambauer & Risch, 2021; Yalcin et al., 2022). Despite this, legal scholarship often overlooks individuals' reactions to decision systems involving algorithms, potentially undermining the legitimacy of legal reforms (Bambauer & Risch, 2021; Yalcin et al., 2022). Understanding citizen preferences offers insights into AI and human decision-making errors, guiding policymakers in understanding how citizens perceive decision-making processes (Yalcin et al., 2022). Previous research suggests that attitudes toward decision-makers vary based on the stakes at risk, emphasising the need for differentiated treatment in high-stakes decision-making processes (Bambauer & Risch, 2021; Yalcin et al., 2022). This inquiry resulted in the following research question *'Do Dutch citizens preferences for utilising AI versus human decision-making in the judiciary change when the stakes at risk vary in terms of incident gravity and severity of punishment?'* Prior research suggests that individuals exhibit algorithm aversion in AI sentencing, preferring human decision-making as stakes increase (Araujo et al., 2020; Bambauer & Risch, 2021). In high-stakes situations, such as a traffic accident with severe injuries, individuals opt for human decision-makers due to uncertainty (Mahmud et al., 2022). However, in low-stakes scenarios, individuals may prefer

AI decision-makers. Surprisingly, the binomial logistic regression analysis revealed no significant association between preferred judicial decision-maker and stakes at risk. Therefore, both research hypotheses, $H_1$ and $H_2$, need to be rejected, indicating that Dutch citizen's preferences are not influenced by stakes at risk. However, the lack of significance in the analysis could be attributed to type-II error. The sample utilised in the research may not be representative to the entire Dutch population, potentially skewing the results (Banerjee et al., 2009). For instance, the sample may overrepresent individuals with higher education, females, and younger age groups, leading to inaccuracies in the findings. This partially aligns with the chi-square test conducted, which indicates that participant's education and gender significantly influence their preference between AI and human judicial decision-makers. The answer to the main question is that there is no significant relationship between the stakes at risk (case gravity and punishment severity) and the preference of Dutch citizens for either an AI or human judicial decision-maker. In summary, while scholars and public administrators debate the role of AI in criminal sentencing and critics advocate for reforms, this study concludes that there is no significant relationship between the stakes at risk of a legal case and Dutch citizen's preference for AI or human judicial decision-makers.

## Bibliography

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351–371. https://doi.org/10.1177/1094428114547952

Aini, G. (2020). A summary of the research on the judicial application of artificial intelligence. *Chinese Studies*, *09*(01), 14–28. https://doi.org/10.4236/chnstd.2020.91002

Alexander, C., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly*, *42*(1), 93. https://doi.org/10.1086/268432

Alon-Barkat, S., & Busuioc, M. (2022). Human–AI interactions in public sector decision making: "Automation bias" and "Selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, *33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Atzmüller, C., & Steiner, P. (2010). Experimental vignette studies in survey research. *Methodology*, *6*(3), 128–138. https://doi.org/10.1027/1614-2241/a000014

Bambauer, D. E., & Risch, M. (2021, July 31). *Worse than human?* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3897126#paper-references-widget

Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal/Industrial Psychiatry Journal*, *18*(2), 127. https://doi.org/10.4103/0972-6748.62274

Barendrecht, M. (2019). De best mogelijke rechtspraak. *Justitiële Verkenningen*, *45*(1), 97–114. https://doi.org/10.5553/jv/016758502019045001008

Barysė, D., & Sarel, R. (2022). Algorithms in the Court: Does it Matter Which Part of the Judicial Decision-Making is Automated? *Social Science Research Network*. https://doi.org/10.2139/ssrn.4208203

Bosick, S. J., Rennison, C. M., Gover, A. R., & Dodge, M. (2012). Reporting violence to the police: Predictors through the life course. *Journal of Criminal Justice*, *40*(6), 441–451. https://doi.org/10.1016/j.jcrimjus.2012.05.001

Buocz, T. J. (2018). Artificial intelligence in court. Legitimacy problems of AI assistance in the judiciary. *Retskraft – Copenhagen Journal of Legal Studies*, 41–59. https://static1.squarespace.com/static/59db92336f4ca35190c650a5/t/5ad9da5f70a6adf9d3ee842c/1524226655876/Artificial+Intelligence+in+Court.pdf

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing*, *25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Centraal Bureau voor de Statistiek. (n.d.). *Bevolkingspiramide*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingspiramide

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Chereni, S., Sliuzas, R., & Flacke, J. (2020). An extended briefing and debriefing technique to enhance data quality in cross-national/language mixed-method research. *International Journal of Social Research Methodology, 23*(6), 661–675. https://doi.org/10.1080/13645579.2020.1731992

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology. General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Duan, Y., Edwards, J., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of Visual Impairment & Blindness*, *109*(2), 164–168. https://doi.org/10.1177/0145482x1510900215

Erfanian, F., Roudsari, R. L., Heydari, A., & Bahmani, M. N. D. (2020). A Narrative on using vignettes: Its advantages and drawbacks. *Journal of Midwifery and Reproductive Health*, *8*(2), 2134–2145. https://doi.org/10.22038/jmrh.2020.41650.1472

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2023). The extent of algorithm aversion in decision-making situations with varying gravity. *PloS One*, *18*(2), e0278751. https://doi.org/10.1371/journal.pone.0278751

Fitzner, K. (2007). Reliability and validity A quick review. ˜ *the œDiabetes Educator/˜ the œDiabetes Educator*, *33*(5), 775–780. https://doi.org/10.1177/0145721707308172

Franke, T., Ho, T. S., & Christie, C. A. (2011). The Chi-Square Test. *American Journal of Evaluation/˜ the œAmerican Journal of Evaluation*, *33*(3), 448–458. https://doi.org/10.1177/1098214011426594

Galloway, A. (2005). Non-Probability sampling. In *Elsevier eBooks* (pp. 859–864). https://doi.org/10.1016/b0-12-369398-5/00382-0

Gillespie, N., Lockey, S., Curtis, C., Pool, J. K., & Akbari, A. (2023). *Trust in Artificial Intelligence: A global study*. https://doi.org/10.14264/00d3c94

Hardy, I., Kloetzer, B., Moeller, K., & Sodian, B. (2010). The Analysis of Classroom Discourse: Elementary School Science Curricula Advancing Reasoning with evidence. *Educational Assessment/Educational Assessment :*, *15*(3–4), 197–221. https://doi.org/10.1080/10627197.2010.530556

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing/Evidence-based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Helberger, N., Araujo, T., & De Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law and Security Report/Computer Law & Security Report*, *39*, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and Updating Demographic Questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*, *21*(3), 138–151. https://doi.org/10.24839/2164-8204.jn21.3.138

Hupe, P. (2019). Chapter 25 Using vignettes in street-level bureaucracy research. In *Research Handbook on Street-Level Bureaucracy* (pp. 392–408). Edward Elgar Publishing. https://doi.org/10.4337/9781786437631

Jenkins, N., Ritchie, L., & Quinn, S. (2020). From reflection to diffraction: exploring the use of vignettes within post-humanist and multi-species research. *Qualitative Research*, *21*(6), 975–989. https://doi.org/10.1177/1468794120920258

Jussupow, E., Benbasat, I., & Heinzl, A. (n.d.). *WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION*. AIS Electronic Library (AISeL). https://aisel.aisnet.org/ecis2020_rp/168?utm_source=aisel.aisnet.org%2Fecis2020_rp%2F168&utm_medium=PDF&utm_campaign=PDFCoverPages

Kulk, S., & Deursen, S. (2020). Juridische aspecten van algoritmen die besluiten nemen Een verkennend onderzoek. In *Montaigne Centrum Voor Rechtsstaat En Rechtspleging*. Montaigne Centrum voor Rechtsstaat en Rechtspleging. https://repository.wodc.nl/bitstream/handle/20.500.12832/2417/2947_volledige_tekst_tcm28-452340.pdf

Lelkes, Y., Krosnick, J. A., Marx, D. M., Judd, C. M., & Park, B. (2012). Complete anonymity compromises the accuracy of self-reports. *Journal of Experimental Social Psychology, 48*(6), 1291–1299. https://doi.org/10.1016/j.jesp.2012.07.002

Levy, J. S. (1992). An introduction to prospect theory. *Political Psychology*, *13*(2), 171–186. http://en.journals.sid.ir/ViewPaper.aspx?ID=145044

Liu, B. (2021). In AI we Trust? Effects of Agency locus and transparency on uncertainty Reduction in Human–AI Interaction. *Journal of Computer-mediated Communication*, *26*(6), 384–402. https://doi.org/10.1093/jcmc/zmab013

Lotfian, M., Ingensand, J., & Brovelli, M. A. (2021). The partnership of citizen science and machine learning: benefits, risks, and future challenges for engagement, data collection, and data quality. *Sustainability*, *13*(14), 8087. https://doi.org/10.3390/su13148087

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting & Social Change/Technological Forecasting and Social Change*, *175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Ministerie van Justitie en Veiligheid. (2023, January 2). *Richtlijn voor strafvordering verkeersongevallen, gevaarlijk verkeersgedrag en verlaten plaats ongeval (2022R004)*. Beleid En Straffen | Openbaar Ministerie. https://www.om.nl/onderwerpen/beleidsregels/richtlijnen-voor-strafvordering-resultaten/richtlijn-voor-strafvordering-verkeersongevallen-2022r004

Ministerie van Onderwijs, Cultuur en Wetenschap. (2024, March 1). *Hoogst behaald onderwijsniveau*. Onderwijs Algemene Feiten En Cijfers | OCW in Cijfers. https://www.ocwincijfers.nl/sectoren/onderwijs-algemeen/hoogst-behaald-onderwijsniveau

Nowotko, P. M. (2021). AI in judicial application of law and the right to a court. *Procedia Computer Science*, *192*, 2220–2228. https://doi.org/10.1016/j.procs.2021.08.235

O'Dell, L., Crafter, S., De Abreu, G., & Cline, T. (2012). The problem of interpretation in vignette methodology in research with young people. *Qualitative Research*, *12*(6), 702–714. https://doi.org/10.1177/1468794112439003

O'Donnell, G. (2004). Why the Rule of Law Matters. *Journal of Democracy*, *15*(4), 32–46. https://doi.org/10.1353/jod.2004.0076

Olsen, R. B., & Orr, L. L. (2016). On the "Where" of Social Experiments: Selecting More Representative Samples to Inform Policy. *New Directions for Evaluation*, *2016*(152), 61–71. https://doi.org/10.1002/ev.20207

Osborne, J. (2010). Arguing to Learn in Science: The role of collaborative, critical Discourse. *Science*, *328*(5977), 463–466. https://doi.org/10.1126/science.1183944

Plutzer, E. (2019). Privacy, sensitive questions, and informed consent. *Public Opinion Quarterly*, *83*(S1), 169–184. https://doi.org/10.1093/poq/nfz017

Roberts, J. V., & Plesničar, M. M. (2014). Sentencing, legitimacy, and public opinion. In *Springer eBooks* (pp. 33–51). https://doi.org/10.1007/978-3-319-09813-5_2

Ryberg, J., & Roberts, J. V. (2022). Sentencing and artificial intelligence. In *Oxford University Press eBooks*. https://doi.org/10.1093/oso/9780197539538.001.0001

Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, *18*(2), 179–183. https://doi.org/10.1002/nur.4770180211

Sela, A. (2018). Can computers be fair? How automated and Human-Powered online dispute resolution affect procedural justice in mediation and arbitration. *Social Science Research Network*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3203190_code604376.pdf?abstractid=3074311&mirid=1&type=2

Sourdin, T. (2018). Judge v Robot? Artificial Intelligence and Judicial Decision-Making. *University of New South Wales Law Journal*, *41*(4). https://doi.org/10.53637/zgux2213

Steiner, P., Atzmüller, C., & Su, D. (2017). Designing valid and reliable vignette experiments for survey Research: A case study on the fair gender income gap. *Journal of Methods and Measurement in the Social Sciences, 7*(2). https://doi.org/10.2458/v7i2.20321

Stipak, B. (1980). Local governments' use of citizen surveys. *PAR. Public Administration Review/Public Administration Review*, *40*(5), 521. https://doi.org/10.2307/3110206

Susskind, R. (2013). *Tomorrow's lawyers: An Introduction to Your Future*. OUP Oxford.

Susskind, R. (2023). *Tomorrow's lawyers: An Introduction to Your Future*. Oxford University Press.

Törrönen, J. (2018). Using vignettes in qualitative interviews as clues, microcosms or provokers. *Qualitative Research Journal*, *18*(3), 276–286. https://doi.org/10.1108/qrj-d-17-00055

UCLA. (n.d.). *Logistic Regression | SPSS annotated output*. Statistical Methods and Data Analytics. Retrieved May 10, 2024, from https://stats.oarc.ucla.edu/spss/output/logistic-regression/

Van Gelder, E. (2018). Online Dispute Resolution: een veelbelovend initiatief voor toegang tot het recht? *Maandblad Voor Vermogensrecht*, *16*(7–8), 262–267. https://doi.org/10.5553/mvv/157457672018016708007

Van Wingerden, S. (2020). Een computer stapt nooit met het verkeerde been uit bed, maar kan hij ook straffen opleggen? *Proces*, *99*(2), 81–84. https://doi.org/10.5553/proces/016500762020099002001

VZinfo. (n.d.). *Bevolking | Bevolkingsomvang | Leeftijd en geslacht*. Volksgezondheid En Zorg. https://www.vzinfo.nl/bevolking-bevolkingsomvang-leeftijd-en-geslacht#:~:text=De%20verhouding%20tussen%20mannen%20en,worden%20(51%2C3%25).

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2022). Perceptions of justice by algorithms. *Artificial Intelligence and Law*, *31*(2), 269–292. https://doi.org/10.1007/s10506-022-09312-z

Završník, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA-Forum*, *20*(4), 567–583. https://doi.org/10.1007/s12027-020-00602-0

Universiteit Leiden

# Appendix

## Appendix 1 – Survey (English)

### Survey introduction

Introduction Thank you for considering participation in this Master's thesis research. Before proceeding, it's essential to understand the purpose and methodology of this study. Please review the following information carefully.

This research constitutes a Master's thesis project and involves an online survey aimed at investigating the preferred methods of judicial decision-making among Dutch citizens.

It will present a hypothetical scenario involving a traffic accident. The gathered data will contribute to scientific research publications, with all results anonymised to ensure confidentiality. Your participation is voluntary, and you are free to withdraw from the survey at any point without explanation. Incomplete responses will not be included in the final dataset. For further details regarding privacy regulations at Leiden University, please refer to the following link: https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection#:~:text=We%20are%20all%20responsible%20for,in%20a%20privacy%2Dproof%20way.

The survey is expected to take approximately 3-5 minutes of your time.

Should you have any inquiries about the survey or the research itself, please don't hesitate to contact Wessel S. van Alphen, a Master's student in Public Administration (specialsing in Public Management and Organisational Sciences), who is conducting this research at w.s.van.alphen@umail.leidenuniv.nl.

Are you willing to participate in this research?

○ Yes, I want to participate (1)

○ No, I do not want to participate (3)

### Scenario 1, High-gravity accident – High-severity punishment

HiLe Ac, HiLe Pu Please review the following information carefully:

You were on your way home from work during rush hour. Due to the traffic jam, you wanted to inform your family about your late arrival. You began to grab your phone and write a message to them. Unfortunately, the text message distracted you from paying attention to the road, resulting in your car colliding with someone else's vehicle and causing severe injuries to the driver of that car. This leads to the conclusion that the accident occurred due to your fault, resulting in a relatively high-level punishment: a potential civil traffic fine of €420 (for using your phone while driving), along with a two-month prison sentence and a one-year suspension from driving (for causing a serious traffic accident due to negligence). Following the court's decision, you have been convicted of "Causing a serious traffic accident with negligence." After your conviction, you are informed that a new program has been implemented offering you two options for selecting your sentence: human decision-making by a traffic court judge or algorithmic decision-making by an AI system incorporating a machine learning model based on past decisions in similar cases.

What option do you choose?

- ◯ Human decision-maker (1)
- ◯ Algortihm decision-maker (3)

### Scenario 2, High-gravity accident – Low-severity punishment

HiLe Ac, LoLe Pu Please review the following information carefully:

You were on your way home from work during rush hour. Due to the traffic jam, you wanted to inform your family about your late arrival. You began to grab your phone and write a message to them. Unfortunately, the text message distracted you from paying attention to the road, resulting in your car colliding with someone else's vehicle and causing severe injuries to the driver of that car. This leads to the conclusion that the accident occurred due to your fault, resulting in a relatively low-level punishment: a potential civil traffic fine of €420 (for using your phone while driving), along with an additional €1,000 and a three-month suspension from driving (for causing a serious traffic accident due to negligence). Following the court's decision, you have been convicted of "Causing a serious traffic accident with negligence." After your conviction, you are informed that a new program has been implemented offering you two options for selecting your sentence: human decision-making by a traffic court judge or algorithmic decision-making by an AI system incorporating a machine learning model based on past decisions in similar cases.

What option do you choose?

- ◯ Human decision-maker (1)
- ◯ Algortihm decision-maker (2)

**Scenario 3, Low-gravity accident – High-severity punishment**

LoLe Ac, HiLe Pu Please review the following information carefully:

You were on your way home from work during rush hour. Due to the traffic jam, you wanted to inform your family about your late arrival. You began to grab your phone and write a message to them. Unfortunately, the text message distracted you from paying attention to the road, resulting in your car colliding with someone else's vehicle and causing minor damage to the other car. This leads to the conclusion that the accident occurred due to your fault, resulting in a relatively high-level punishment: a potential civil traffic fine of €420 (for using your phone while driving), along with an additional €100 and two-month suspension for driving (for causing a traffic accident with only material damage). Following the court's decision, you have been convicted of "Causing a traffic accident with only material damage." After your conviction, you are informed that a new program has been implemented offering you two options for selecting your sentence: human decision-making by a traffic court judge or algorithmic decision-making by an AI system incorporating a machine learning model based on past decisions in similar cases.

Which option do you choose?

- ◯ Human decision-maker (1)
- ◯ Algortihm decision-maker (2)

**Scenario 4, Low-gravity accident – Low-severity punishment**

LoLe Ac, LoLe Pu Please review the following information carefully:

You were on your way home from work during rush hour. Due to the traffic jam, you wanted to inform your family about your late arrival. You began to grab your phone and write a message to them. Unfortunately, the text message distracted you from paying attention to the road, resulting in your car colliding with someone else's vehicle and causing minor damage to the other car. This leads to the conclusion that the accident occurred due to your fault, resulting in a relatively low-level punishment: a potential civil traffic fine of €420 (for using your phone while driving) and an additional €100 (for causing a traffic accident with only material damage). Following the court's decision, you have been convicted of "Causing a traffic accident with only material damage." After your conviction, you are informed that a new program has been implemented offering you two options for selecting your sentence: human decision-making by a traffic court judge or algorithmic decision-making by an AI system incorporating a machine learning model based on past decisions in similar cases.

Which option do you choose?

- ◯ Human decision-maker (1)
- ◯ Algorithm decision-maker (3)

**Are you a Dutch citizen?**

Nationality Are you a Dutch citizen?

- Yes (1)
- No (2)

**Demographic questions**

Age What is your current age?

- Under 18 years (1)
- 18-24 years (2)
- 25-34 years (3)
- 35-44 years (4)
- 45-54 years (5)
- 55-64 years (6)
- 65-74 years (7)
- 75-84 years (8)
- Over 84 years (9)

Gender What best describes your gender?

- Male (1)
- Female (2)
- Prefer not say (3)
- Other (4)

Education What is the highest educational level you have optained?

- None (1)
- Primary school (2)
- Secondary school (3)
- Bachelor's degree (4)
- Master's degree (5)
- Ph.D. (6)
- Doctorate (7)
- Prefer not to say (8)

**Control question**

Control What was your preferred judicial decision-maker?

- Human decision-maker (1)
- Algortihm decision-maker (2)

## Appendix 2 – Survey (Dutch)

### Survey introduction

Introduction Bedankt voor uw overweging om deel te nemen aan dit masterthesisonderzoek. Alvorens verder te gaan, is het essentieel om het doel en de methodologie van deze studie te begrijpen. Gelieve de volgende informatie zorgvuldig door te nemen.

Dit onderzoek vormt een master scriptie project en omvat een online enquête gericht op het onderzoeken van de voorkeursmethoden van gerechtelijke besluitvorming onder Nederlandse burgers.

Er zal een hypothetisch scenario worden gepresenteerd met betrekking tot een verkeersongeval. De verzamelde gegevens zullen bijdragen aan wetenschappelijke onderzoekpublicaties, waarbij alle resultaten geanonimiseerd worden om vertrouwelijkheid te waarborgen. Uw deelname is vrijwillig en u bent vrij om op elk moment zonder opgaaf van redenen uit de enquête terug te trekken. Onvolledige reacties worden niet opgenomen in de uiteindelijke dataset. Voor verdere details over privacyregels van de Universiteit Leiden, verwijzen wij u graag naar de volgende link: https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection#:~:text=We%20are%20all%20responsible%20for,in%20a%20privacy%2Dproof%20way.

De enquête zal naar verwachting ongeveer 3-5 minuten van uw tijd in beslag nemen.

Mocht u vragen hebben over de enquête of het onderzoek zelf, aarzel dan niet om contact op te nemen met Wessel S. van Alphen, een masterstudent in Public Administration (gespecialiseerd in Public Management & Organisational Sciences), die dit onderzoek uitvoert via w.s.van.alphen@umail.leidenuniv.nl.

Bent u bereid om deel te nemen aan dit onderzoek?

- ○ Ja, ik wil graag deelnemen (1)
- ○ Nee, ik wil graag niet deelnemen (3)

**Universiteit Leiden**

**Scenario 1, High-gravity accident – High-severity punishment**

HiLe Ac, HiLe Pu Gelieve de volgende informatie zorgvuldig door te nemen:

Je was op weg naar huis vanaf je werk tijdens de spits. Vanwege de file wilde je je familie informeren over je late aankomst. Je pakte je telefoon en stuurde een bericht naar hen. Helaas leidde het tekstbericht je af van het opletten op de weg, waardoor je auto in botsing kwam met een ander voertuig en ernstige verwondingen veroorzaakte bij de bestuurder van dat voertuig. Dit leidt tot de conclusie dat het ongeval is veroorzaakt door jouw schuld, wat resulteert in een relatief zware straf: een mogelijke civiele verkeersboete van €420 (voor het gebruik van je telefoon tijdens het rijden), samen met een gevangenisstraf van twee maanden en een rijontzegging van één jaar (voor het veroorzaken van een ernstig verkeersongeval door nalatigheid). Na de beslissing van de rechtbank ben je veroordeeld voor "Het veroorzaken van een ernstig verkeersongeval door nalatigheid." Na je veroordeling word je geïnformeerd dat er een nieuw programma is geïmplementeerd dat je twee opties biedt voor het kiezen van je straf: menselijke besluitvorming door een verkeersrechter of algoritmische besluitvorming door een AI-systeem dat een machine learning-model omvat op basis van eerdere beslissingen in soortgelijke gevallen.

Welke optie kies je?

- ○ Menselijke besluitnemer (1)
- ○ Algoritmische besluitnemer (3)

**Scenario 2, High-gravity accident – Low-severity punishment**

HiLe Ac, LoLe Pu Gelieve de volgende informatie zorgvuldig door te nemen:

Je was onderweg naar huis vanaf je werk tijdens de spits. Vanwege de file wilde je je familie informeren over je late aankomst. Je pakte je telefoon en stuurde een bericht naar hen. Helaas leidde het tekstbericht je af van het opletten op de weg, waardoor je auto in botsing kwam met een ander voertuig en ernstige verwondingen veroorzaakte bij de bestuurder van dat voertuig. Dit leidt tot de conclusie dat het ongeval is veroorzaakt door jouw schuld, wat resulteert in een relatief lage straf: een mogelijke civiele verkeersboete van €420 (voor het gebruik van je telefoon tijdens het rijden), samen met een extra €1.000 en een rijontzegging van drie maanden (voor het veroorzaken van een ernstig verkeersongeval door nalatigheid). Na de beslissing van de rechtbank ben je veroordeeld voor "Het veroorzaken van een ernstig verkeersongeval door nalatigheid." Na je veroordeling word je geïnformeerd dat er een nieuw programma is geïmplementeerd dat je twee opties biedt voor het kiezen van je straf: menselijke besluitvorming door een verkeersrechter of algoritmische besluitvorming door een AI-systeem dat een machine learning-model omvat op basis van eerdere beslissingen in soortgelijke gevallen.

Welke optie kies je?

- ○ Menselijke besluitnemer (1)
- ○ Algoritmische besluitnemer (2)

**Scenario 3, Low-gravity accident – High-severity punishment**

LoLe Ac, HiLe Pu Gelieve de volgende informatie zorgvuldig door te nemen:

Je was onderweg naar huis vanaf je werk tijdens de spits. Vanwege de file wilde je je familie informeren over je late aankomst. Je pakte je telefoon en stuurde een bericht naar hen. Helaas leidde het tekstbericht je af van het opletten op de weg, waardoor je auto in botsing kwam met een ander voertuig en lichte schade veroorzaakte aan de andere auto. Dit leidt tot de conclusie dat het ongeval is veroorzaakt door jouw schuld, wat resulteert in een relatief zware straf: een mogelijke civiele verkeersboete van €420 (voor het gebruik van je telefoon tijdens het rijden), samen met een extra €100 en een rijontzegging van twee maanden (voor het veroorzaken van een verkeersongeval met alleen materiële schade). Na de beslissing van de rechtbank ben je veroordeeld voor "Het veroorzaken van een verkeersongeval met alleen materiële schade." Na je veroordeling word je geïnformeerd dat er een nieuw programma is geïmplementeerd dat je twee opties biedt voor het kiezen van je straf: menselijke besluitvorming door een verkeersrechter of algoritmische besluitvorming door een AI-systeem dat een machine learning-model omvat op basis van eerdere beslissingen in soortgelijke gevallen.  Welke optie kies je?

○ Menselijke besluitnemer (1)

○ Algoritmische besluitnemer (2)

**Scenario 4, Low-gravity accident – Low-severity punishment**

LoLe Ac, LoLe Pu Gelieve de volgende informatie zorgvuldig door te nemen:

Je was onderweg naar huis vanaf je werk tijdens de spits. Vanwege de file wilde je je familie informeren over je late aankomst. Je pakte je telefoon en stuurde een bericht naar hen. Helaas leidde het tekstbericht je af van het opletten op de weg, waardoor je auto in botsing kwam met een ander voertuig en lichte schade veroorzaakte aan de andere auto. Dit leidt tot de conclusie dat het ongeval is veroorzaakt door jouw schuld, wat resulteert in een relatief lichte straf: een mogelijke civiele verkeersboete van €420 (voor het gebruik van je telefoon tijdens het rijden) en een extra €100 (voor het veroorzaken van een verkeersongeval met alleen materiële schade). Na de beslissing van de rechtbank ben je veroordeeld voor "Het veroorzaken van een verkeersongeval met alleen materiële schade." Na je veroordeling word je geïnformeerd dat er een nieuw programma is geïmplementeerd dat je twee opties biedt voor het kiezen van je straf: menselijke besluitvorming door een verkeersrechter of algoritmische besluitvorming door een AI-systeem dat een machine learning-model omvat op basis van eerdere beslissingen in soortgelijke gevallen.

Welke optie kies je?

○ Menselijke besluitnemer (1)

○ Algoritmische besluitnemer (3)

**Are you a Dutch citizen?**

Nationality Bent u een Nederlands staatsburger?

- Ja (1)
- Nee (2)

**Demographic questions**

Age Wat is uw huidige leeftijd?

- Onder de 18 jaar (1)
- 18-24 jaar (2)
- 25-34 jaar (3)
- 35-44 jaar (4)
- 45-54 jaar (5)
- 55-64 jaar (6)
- 65-74 jaar (7)
- 75-84 jaar (8)
- Boven de 84 jaar (9)

Gender Wat beschrijft jouw geslacht het best?

- Man (1)
- Vrouw (2)
- Wil ik liever niet zeggen (3)
- Anders (4)

Education Wat is het hoogste opleidingsniveau wat je ooit hebt behaald?

- Geen (1)
- Basisonderwijs (2)
- Middelbaaronderwijs (3)
- Bachelor studie (4)
- Master studie (5)
- Ph.D. (6)
- Doctoraat (7)
- Wil ik liever niet zeggen (8)

**Control question**

Control Wat was je voorkeurs besluitvormer bij de casus?

- Menselijke besluitnemer (1)
- Algoritmische besluitnemer (2)

## Appendix 3 – SPSS Output tables, Binomial logistic regression

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 83,22 | ,03 | ,06 |

**Categorical Variables' Codings**

| | | Frequency | Parameter coding | | |
|---|---|---|---|---|---|
| | | | (1) | (2) | (3) |
| Ac_Pu | HiLeAc_HiLePu | 24 | 1 | 0 | 0 |
| | HiLeAc_LoLePu | 23 | 0 | 1 | 0 |
| | LoLeAc_HiLePu | 25 | 0 | 0 | 1 |
| | LoLeAc_LoLePu | 24 | 0 | 0 | 0 |

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% CI for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1 | Ac_Pu | | | 3,42 | 3 | ,332 | | | |
| | Ac_Pu(1) | -1,06 | ,76 | 1,92 | 1 | ,165 | ,35 | ,08 | 1,55 |
| | Ac_Pu(2) | -1,01 | ,76 | 1,74 | 1 | ,187 | ,36 | ,08 | 1,63 |
| | Ac_Pu(3) | -1,11 | ,76 | 2,10 | 1 | ,147 | ,33 | ,07 | 1,47 |
| | Constant | -,89 | ,45 | 3,90 | 1 | ,048 | ,41 | | |