When Do Decision-Makers Delegate Strategic Decisions to AI? The Effects of AI Accuracy and Ethical Decision Complexity

Lennart Samland

Dissertation written under the supervision of Peter V. Rajsingh PhD (Católica Lisbon) and Gwendoline Sajons (ESCP Business School)

Dissertation submitted in partial fulfilment of requirements for the MSc International Management with a Specialization in Strategy & Consulting, at Universidade Católica Portuguesa and for the MSc in Management with a Specialization in AI & Robotics in Business, at ESCP Business School, 02.06.2025.

Resumo

Esta tese avalia o impacto das diferenças nas taxas de exatidão da IA comunicadas e da presença de uma componente ética na decisão sobre a vontade dos decisores de delegar decisões estratégicas à IA. A primeira das três questões de investigação é: "Como é que uma menor precisão da IA comunicada afeta a vontade dos decisores de lhe delegarem decisões estratégicas?" Em segundo lugar, esta tese examina "Como é que a complexidade ética de uma decisão estratégica influencia a vontade dos decisores de delegar decisões na IA?" e, por fim, "Como é que a presença de uma componente ética altera o impacto de uma baixa precisão comunicada na vontade dos decisores de delegar decisões estratégicas à IA?".

A investigação atual sobre os efeitos da precisão comunicada e da complexidade ética é limitada e centrada em simulações ou setores como saúde e defesa. Contudo, há pouca consideração destes fatores no contexto da tomada de decisões empresariais, que em breve será uma área essencial para aplicação de IA (Gartner, 2023).

Esta tese conclui que o efeito da exatidão comunicada não é estatisticamente significativo, mas confirma que uma componente ética numa decisão estratégica torna os decisores menos propensos a delegar. Este efeito deixa de ser significativo com a introdução do termo de interação, que também é insignificante. Apesar de H1 e H3 não serem confirmadas e apenas H2 ser parcialmente apoiada, esta tese contribui para o debate sobre a aversão e apreciação de algoritmos e oferece recomendações práticas a gestores e desenvolvedores de IA.

Palavras-chave: IA, Tomada de decisão, Delegação estratégica, Aversão a algoritmos, Interação Humano-IA, Precisão da IA, Complexidade ética, Estratégia corporativa

Título: Quando os Decisores Delegam Decisões Estratégicas à IA? Os Efeitos da Precisão da IA e da Complexidade Ética da Decisão.

Autor: Lennart Samland

Abstract

This thesis assesses the impact of differences in communicated AI accuracy rates and the presence of an ethical decision component on decision-makers' willingness to delegate strategic decisions to AI. The first of three research questions is: “How does a lower communicated AI accuracy impact decision-makers’ willingness to delegate strategic decisions to it?” Secondly, this thesis examines “How does the ethical complexity of a strategic decision influence decision-makers’ willingness to delegate decisions to AI?” and the final research question of “How does the presence of an ethical decision component change the impact of a low communicated accuracy on decision-makers' willingness to delegate strategic decisions to AI?” will be answered.

Current research on the effects of communicated accuracy rate and ethical complexity is limited and confined to game simulations or sectors like healthcare and the military. However, there is little consideration of these factors in the context of corporate decision-making, which will soon pose an essential area of application for AI systems (Gartner, 2023).

This thesis finds the effect of communicated accuracy to be non-significant, while it confirms that an ethical component in a strategic decision makes decision-makers less likely to delegate the decision. This effect only holds until the introduction of the equally non-significant interaction term. Though H1 and H3 could not be supported, and H2 was only partially supported, this thesis contributes to the discussion of the causes of algorithm aversion. Additionally, we offer practical suggestions on factors that managers and developers must watch out for when implementing AI into strategic decision-making.

Keywords: AI, Decision-making, Strategy delegation, Algorithm aversion, Human-AI interaction, AI accuracy, Ethical complexity, Corporate strategy

Title: When Do Decision-Makers Delegate Strategic Decisions to AI? The Effects of AI Accuracy and Ethical Decision Complexity.

Author: Lennart Samland

List of Figures

List of Tables

List of abbreviations

AI	Artificial Intelligence
RQ	Research Question
ML	Machine Learning
AGI	Artificial General Intelligence
ASI	Artificial Super Intelligence
LLM	Large Language Model
TAM	Technology Acceptance Model
UTAUT	Unified Theory of Acceptance and Use of Technology
XAI	Explainable Artificial Intelligence
IV	Independent Variable
DV	Dependent Variable
VIF	Variance Inflation Factor

1. Introduction

“What steam was to the First Industrial Revolution is what AI will be to the fourth” (Bristol et al., 2023, p. 1). Just as steam power was essential to gain a competitive edge, the implementation of Artificial Intelligence (AI) into companies’ decision-making has become a cornerstone of nowadays competitive advantage (Halim et al., 2024). According to Gartner (2023), "79% of corporate strategists see AI and analytics as critical to their success over the next two years." (Gartner, 2023, p. 1). With AI being essential for a company's success in the following years, this thesis delves into different factors that influence decision-makers’ willingness to delegate strategic decisions to an AI.

Due to the importance of AI, many researchers have turned their attention to the topic of AI in decision-making. The implementation of AI in decision-making is especially incentivized as its superior computational power (Bigman & Gray, 2018) allows AI to outperform humans in many domains (Csaszar et al., 2024; Dietvorst et al., 2015; Mahmud et al., 2022). AI and algorithms are widely recognized to be beneficial for forecasting and data-driven insights (Csaszar et al., 2024; Dietvorst et al., 2015; Logg et al., 2019; Mahmud et al., 2022), which are essential components of strategic decisions. Despite the proven superiority of AI in many fields, humans tend to show algorithm aversion, which consists of the rejection of algorithmic help despite superior performance (Dietvorst et al., 2015). The thesis covers several potential causes of algorithm aversion, which can be rooted in the high accuracy expectations that individuals have in algorithmic decision aids (Dzindolet et al., 2002; Mahmud et al., 2022), as well as the disproportionally large loss of confidence after an observed error (Alarcon et al., 2024; Dietvorst et al., 2015; Jussupow et al., 2024). Additional literature dives into the trust dynamics that influence algorithm aversion by outlining learning perception gaps (Dietvorst et al., 2015; Mahmud et al., 2022; Sloan & Warner, 2018), error timing (Manzey et al., 2012), as well as error severity and context (Filiz et al., 2023; Parlangeli et al., 2024; Renier et al., 2021). Further authors also determine familiarity (Choung et al., 2024; Mahmud et al., 2022, 2024), AI quality (Pathak & Bansal, 2024), or demographics (Araujo et al., 2020; Brink et al., 2023; Ferraz et al., 2025; Mahmud et al., 2022) to be determinants of algorithm aversion.

Despite the extensive research in these fields, very little attention is paid to the impact of AI accuracy. During the literature research of this paper, only very few papers on the impact of AI accuracy on delegation behavior could be found so we had to resort to literature on general algorithm aversion to justify the hypotheses. None of the literature provided had directly studied the impact of communicated AI accuracy in a corporate decision context. Most of the research is focused on either hypothetical scenarios to measure the underlying aversion effects or is clustered in industries such as healthcare. A similar pattern can be seen in the literature on AI in ethical decision-making, which is mainly concentrated around the potential benefits and shortcomings of using AI for ethical decision-making. Scholars mainly outline the benefits to lie in the potential for explainability of otherwise intuitive decisions (Demaree-Cotton et al., 2022; Kvam et al., 2024). On the other hand, extensive research was done on shortcomings such as the reduction of complex ethical problems into statistical correlations (Benzinger et al., 2023) , as well as the problems of accountability (Parlangeli et al., 2024) and transparency (Cheong, 2024). Similar to the general research on causes of algorithm aversion, close to no attention is paid to the potential of AI for strategic decisions with ethical implications. The only business content that incorporates some degree of ethical considerations is composed of literature that analyzes the use of AI for hiring practices (Choung et al., 2024), however, this field is rarely subject to strategic decisions. The rest of the research on AI-based ethical decision-making is once again clustered in the domains of healthcare (Benzinger et al., 2023; Biller-Andorno et al., 2022; Demaree-Cotton et al., 2022; Kerstan et al., 2024) and the military (Bellaby, 2021; De Cremer & Narayanan, 2023; Kohn et al., 2024; Liao, 2020). This indicates a critical gap in the literature, as the previously outlined importance and potential benefits of the implementation of AI into strategic decision-making are not being leveraged. Therefore, the dynamics of algorithm aversion concerning AI accuracy and the presence of an ethical decision component need to be studied in a corporate context. This thesis will consider independent variables, which are currently not sufficiently represented, and apply them to the new context of strategic decision-making. Finally, the joint influence of communicated AI accuracy and ethical complexity in strategic decision contexts previously remained unexplored but will be addressed in the thesis.

The first two research questions of the thesis will address the variables of communicated AI accuracy and ethical complexity of a strategic decision. They will be framed as: “How does a lower communicated AI accuracy impact decision-makers’ willingness to delegate strategic decisions to it?” and “How does the ethical complexity of a strategic decision influence decision-makers’ willingness to delegate decisions to AI?”. The third research question of this thesis seeks to understand “How does the presence of an ethical decision component change the impact of a low communicated accuracy on decision-makers' willingness to delegate strategic decisions to AI?”.

Answering these three research questions allows us to fill the previously mentioned gap in the literature and contribute to the discussion of algorithm aversion. Additionally, the application of algorithm aversion to the context of AI-based strategic decision-making allows for new insights that were previously disregarded. The thesis will also deliver practical implications as it helps managers to understand how AI could best be introduced into the context of their strategic decision-making and which conditions facilitate or hinder the successful implementation.

To investigate these questions, we conducted a 2x2 experiment, which analyzed the impact of changes in the communicated accuracy of the AI system that participants had at their disposal and the presence of a strategic decision with ethical considerations on decision-makers' willingness to outsource strategic decisions to AI. The 2x2 experiment was then distributed to a target of 200 random participants using the platform Prolific (Prolific, 2024).

The analysis stage of this thesis concludes that the impact of a change in the communicated AI accuracy was not statistically significant. However, the presence of an ethical component in the strategic decision had a statistically significant impact throughout the models, which measured the main effects. This coefficient also turned insignificant as soon as the interaction term was introduced. Finally, the interaction of accuracy and ethics was also not statistically significant.

The following section of this thesis consists of the literature review. This part begins with a general assessment of AI in decision-making, before outlining research that pertains to algorithm aversion and the role of AI decision-making in ethically complex situations and industries. The literature review is concluded by a summary of key managerial theories, a chapter on the impact of AI decision-making on managers, and finally, the theoretical considerations. After covering the methods part of the thesis, we will explain the results that we got from the 2x2 experiments. These results are then used in the discussion, in which we discuss each of the hypotheses and control variables before clarifying the theoretical and practical relevance of the thesis. The thesis is concluded by a chapter on limitations and suggestions for future research.

2. Literature Review

2.1. Introduction

This literature review synthesizes literature regarding the topics of AI, algorithms, and machine learning (ML) with a particular focus on their role in automated decision-making. To understand the motivations behind automated decision-making, we will explain the underlying psychological effects, such as trust, aversion, and delegation.

2.1.1. Definitions of AI

AI came into being at the Dartmouth Conference of 1956, which is considered its founding event, and which coined the term of artificial intelligence (Moor, 2006). However, the idea of an intelligent machine was conceptualised earlier with the Turing test, which was originally called the imitation game. The Turing test was a test of a machine's ability to exhibit intelligent behavior to convince a human counterpart into believing a human being was responding (Turing, 1950). From these beginnings of AI, several definitions have branched off throughout the literature. Luger et al (1993), frame AI as “The branch of computer science concerned with the automation of intelligent behavior.” (Luger et al., 1993, p. 1). Meanwhile, others understand it as the “ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings” (Copeland, 2025, p. 1). AI does not just have to be limited to intelligent behavior, but can be specific to tasks that would usually be done by humans and which require adaptive decisions depending on the environment (Kritikos & Iphofen, 2023). Alternative definitions involve not only the fulfillment of human tasks but also the ability to interact and communicate in a human-like way (Gil de Zúñiga et al., 2024).

AI thus constitutes an intelligent agent (Russell & Norvig, 2021) who, according to some scholars, can be capable of using intent interference (Kvam et al., 2024). Intent interference refers to the ability of a system to make sense of usually non-quantifiable reasoning that humans use in their decision-making process. Kvam et al. (2024) state that an AI system using intent interference outperforms traditional models that only look at observed behavior, as they understand the deeper motivations of human decision-making to accurately predict future human behavior.

This literature review will also draw information from papers on Machine Learning (ML) and algorithms. Machine learning is a “subfield of artificial intelligence that gives computers the ability to learn without being explicitly programmed" (Brown, 2021, p. 1). While AI and ML are focused on a computer's ability to act independently, an algorithm is a computation process that strictly follows pre-defined rules and procedures (CSRC, n.d.). Therefore, the distinction between AI, algorithms, and ML lies in the fact that ML is a subsection of AI, while both are composed of algorithms. For consistency, throughout the thesis, we will use the term AI in all general discussions.

2.1.2. Benefits of AI Decision-Making

An increasing number of decisions are made by algorithms instead of humans (Bigman & Gray, 2018; Mahmud et al., 2024). A perceived key factor of why people use AI in decision-making traces back to the so-called algorithm appreciation. Originally coined by Logg et al. (2019), algorithm appreciation refers to a person's willingness to rely on algorithms when it comes to decision-making (Mahmud et al., 2024). Studies of AI decisions in a hiring context found that participants evaluated AI decisions as being more competent and trustworthy than human decisions (Choung et al., 2024). Trust thus plays an important role in the willingness of people to take advice from AI (Glikson & Woolley, 2020) or delegate decisions to algorithms (Dietvorst et al., 2015). Additional research on AI prediction systems also shows that the trustworthiness that some people perceive is heavily based on perceived fairness (Song et al., 2025). This theory is backed by other scholars who understand perceived fairness to be one of the essential determinants of trust, use, and acceptance of AI decisions (Shulner-Tal et al., 2023). Potential reasons for the tendency of many individuals to show algorithm appreciation lie in the benefits that such decision-making offers.

The first benefit of AI systems is that they can reduce human error and improve decision accuracy (Choi et al., 2025; Klingbeil et al., 2024). Bigman & Gray (2018) argue that due to superior computational power, machines with AI tend to make more accurate decisions in fields that involve calculations and logic. This advanced computational power is also seen as the reason why AI decision-making is considered superior to humans in contexts with great uncertainty and complexity, as AI can quickly recognize patterns and make sense of complex variable interrelations (Bennett et al., 2024). Other authors also highlight the high scalability and speed that AI automation of decision-making processes can provide. An implementation of AI promises real-time decision-making with instant data insights (Balbaa & Abdurashidova, 2024). These benefits could augment human capabilities across various industries (Balbaa & Abdurashidova, 2024).

Finally, algorithmic decisions are generally more consistent than those of humans and have lower variance. Meanwhile, humans are generally more likely to generate a perfect result, however, their decisions are, on average, more inaccurate and less consistent (Dietvorst & Bharti, 2020; Mahmud et al., 2022). Due to this pattern, algorithms excel at avoiding obvious mistakes, but their decisions are too consistent to achieve perfect scores (Dietvorst et al., 2015).

2.1.3. Shortcomings of AI Decision-Making

Issues that arise from the use of AI for decision-making include concerns about bias, transparency, and accountability (Benzinger et al., 2023; Nassar & Kamal, 2021). When it comes to transparency, AI-based systems are often perceived as a black-box, which creates scepticism about their use (Mahmud et al., 2024). Existing biases in training datasets are an additional, much-discussed issue (Benzinger et al., 2023). Furthermore, people tend to overrely on AI input after being told that the recommendation comes from an AI (Klingbeil et al., 2024). Agudo et al. (2024) describe the phenomenon of over-trusting an AI as automation bias. This even includes trusting an AI when it errs (Agudo et al., 2024). Logg et al. (2019) also found that people are likely to trust an algorithm more, except if the alternative is their own judgment.

2.1.4. General AI Use Cases

Artificial General Intelligence (AGI) describes machines matching human capacities across various tasks. This is distinguished from Artificial Super Intelligence (ASI), where machines surpass humans (Iqbal, 2024). Mahmud et al. (2022) and Csaszar et al. (2024) argue that is already the case in certain fields. Csaszar et al. (2024) point to the dominance of AI in quantifiable decisions, such as trading, where AI has taken over 78% of all trading decisions. The same can be said for evidence-based algorithms, which are almost always more accurate when it comes to predictions than human forecasters (Dietvorst et al., 2015). When it comes to decision-making, scholars state that people are most open towards algorithms in quantitative fields (Logg et al., 2019). However, this does not imply that companies are not also using AI in other applications. AI is being used increasingly in human resource management, where. It is taking over decisions and delivering data-based predictive analyses of employee turnover

(Rodgers et al., 2023). Moreover, AI in human resource management is leading to cost savings in recruitment and talent management (Rodgers et al., 2023).

2.2. Algorithm Aversion and AI Prediction Mistakes

This chapter explains factors that are compromising trust when it comes to AI in decision-making, based on the fundamental concept of algorithm aversion.

2.2.1. Definition of Algorithm Aversion

Algorithm aversion is understood as a general willingness of individuals to discard algorithmic advice even though it proves to be more accurate (Dietvorst et al., 2015; Mahmud et al., 2022). Algorithm aversion can generally be divided into two parts. Aversion that occurs before the interaction and aversion that results from an error by the AI (Jussupow et al., 2024).

2.2.2. Causes and Dynamics of Aversion

Expectation and regret – According to Dietvorst et al (2015), people are generally more likely to choose a human forecaster instead of an algorithm even when they see the algorithm outperform the human counterpart. This implies that people weigh human input more strongly. Additionally, they also judge professionals who use algorithmic input more strictly. People are also generally very aware of errors made by algorithms, as they expect a higher accuracy (Dzindolet et al., 2002; Mahmud et al., 2022). When humans observe an AI making a mistake of a diagnostic nature, trust in the particular AI erodes quicker than for a human who commits the same error (Jussupow et al., 2024). The reason lies in the fact that humans tend to excuse another human's errors and attribute them to external factors. This also applies to robots, which disproportionally lose more confidence relative to humans committing a similar competence error (Alarcon et al., 2024). Thus, people abandon algorithmic decision aids after perceiving decision-making imperfections (Dietvorst et al., 2015). This further negatively influences the likelihood of complying with an algorithm in the future (Mahmud et al., 2022). In order to restore trust, a longer error-free period is needed (Bigman & Gray, 2018).

Learning perception gap – People are also averse to inputs given by algorithms whenever they have previously encountered mistakes. This is based on the assumption that algorithms are unable to learn from previous errors (Dietvorst et al., 2015) or detect their own errors (Sloan & Warner, 2018). By proving an algorithm's ability to learn from mistakes, one could decrease people's aversion to said algorithms (Mahmud et al., 2022).

Error timing, severity, and context – Also relevant are the timing of the algorithmic error and the stage of the decision-making process. The primacy effect is more significant than the recency effect, which implies that an error during the early stages of a decision-making process more negatively impacts trust (Manzey et al., 2012). Similarly, AI errors can also be evaluated differently due to the severity of the consequences as well as the context of the error. People react differently depending on how severity of the mistake (Parlangeli et al., 2024; Renier et al., 2021). Thus, errors in high-stakes situations will sustainably compromise trust (Parlangeli et al., 2024). Moreover, studies found that people's tendency for algorithm aversion is higher in contexts with potentially more serious consequences of the decision (Filiz et al., 2023). The importance of context is supported by research in specific high-stakes industries. For the purpose of this thesis, the military (Bellaby, 2021; Liao, 2020) and healthcare (Rosenbacke et al., 2024) settings were selected to illustrate AI-based decision-making in industries that are composed of sensitive contexts, which leave little margin for error. This is supported by research in the medical field that identifies the absence of a 100% accurate AI to be a key risk in the respective high-stakes decisions (Lone et al., 2025).

Familiarity and trust calibration – Additional factors influencing algorithm aversion are familiarity with the algorithm, duration, and frequency of use (Mahmud et al., 2022). Familiarity refers to a person's knowledge in a field, which usually comes from practical experience or direct exposure (Mahmud et al., 2024). Increased familiarity reduces "black box effects" and positively influences trust, which in the end affects behavior (Mahmud et al., 2024). Choung et al. (2024) also found that increased familiarity and exposure to AI systems reduce distrust. Jussupow et al. (2024) argue that this can potentially go both ways. While some people shift from aversion to appreciation, others will shift in the reverse direction. Research also finds that perceived AI quality is another important component of cognitive trust, which is one of the core variables that help to explain people's intention to delegate decisions to AI (Pathak & Bansal, 2024).

User characteristics – Another factor influencing willingness to use AI is demographics (Mahmud et al., 2022). Older people are generally more averse to algorithms (Araujo et al., 2020) and therefore less likely to delegate decisions to an algorithm (Ferraz et al., 2025). Other authors also found age and education levels to be important drivers of managers' willingness to use AI for their decision-making (Brink et al., 2023). Finally, women tend to consider algorithms less useful in contrast to men (Araujo et al., 2020). This hypothesis is also elaborated on by research by Ferraz et al. (2025), who add that female participants generally react more negatively to observed algorithmic errors, showing a lower error tolerance than men.

2.3. AI in Ethical Decision-Making

This section not only discusses the implications of including AI in ethical decisions but also provides case studies from two industries. These examples should illustrate how AI is currently being applied to ethical decisions. The goal is to understand how humans integrate AI decision-making in sensitive domains. To enable machines to be capable of ethical decision-making, the creation of intelligent machines, which are able to autonomously make decisions that are considered to be "good" in our value system is needed (Etzioni & Etzioni, 2016).

2.3.1. Strengths of AI in Ethical Decision-Making

Said intelligent machines bring a set of benefits into the context of ethical decision-making. Firstly, AI, such as Large Language Models (LLMs), is capable of ethical reasoning and can challenge decisions by playing devil's advocate (Csaszar et al., 2024). Demaree-Cotton et al (2022) argue that, as ethical decision-making of humans cannot be reduced to exact criteria but instead is based on intuition without clear factors or weighting, employing AI for ethical decision-making allows for precise justification of the decisions that are taken through traceable reasoning. Therefore, AI can be used as a way to improve transparency and explainability of an ethical decision (Kvam et al., 2024). Moreover, AI could be used as a decision-making aid, and even if the AI's proposal is ultimately rejected, the steps it took to get to this decision can be traced and used by the human decision-maker to justify their otherwise intuitive reasoning (Demaree-Cotton et al., 2022). An additional benefit brought by AI decision-making is that it allows for making informed decisions in areas that are sensitive and might thus be subject to human biases or errors (Agbabiaka et al., 2025). Agbabiaka et al. (2025) thus argue that a correctly implemented AI can also protect the otherwise more vulnerable populations, as human biases can be excluded specifically.

2.3.2. Criticism against AI in Ethical Decision-Making

However, implementing the previously mentioned benefits can pose a challenge as several points of criticism need to be considered. Generally, AI decisions about human lives in autonomous driving, law, medicine, and military were studied, and a general algorithm aversion of participants was found (Bigman & Gray, 2018). Dietvorst & Bartels (2022) found that consumers are generally less willing to accept algorithmic decisions in decisions that include morally relevant tradeoffs. Instead, they only accept the algorithm's input in decisions without moral implications or with low stakes. A potential reason for this could be that the reduction of complex ethical problems to statistical correlations that lack empathy ignores human cultural and emotional inputs (Benzinger et al., 2023). A human mind that can think and feel is needed for difficult decisions (Bigman & Gray, 2018). In addition, Zhang et al. (2022) found that people consider AI's decision-making to be more utilitarian compared to humans. This implies that it aims at maximizing the outcome for the greatest number of people, but on the other hand, it accepts potential harm to reach this utilitarian goal. This way of prioritizing goals substantially differentiates AI from humans, who can act from deontological principles (Zhang et al., 2022).

Another problem of AI in ethical decision-making is scapegoat risk. This appears when humans seek to avoid blame by delegating everything to an AI. The introduction of AI in this context leaves responsibility gaps, which create accountability ambiguity and can erode trust in AI decision-making. It is easier for decision-makers to diminish punishment by delegating it to machines than to delegate it to humans (Parlangeli et al., 2024). Additionally, research finds that managers do not always employ algorithmic decision aids to improve the accuracy of their decisions, but to potentially avoid blame (Aschauer et al., 2024).

Additional issues arise from the fact that AI mainly acts as a mirror that reflects the behavior and biases of humans instead of thinking for itself (De Cremer & Narayanan, 2023). De Cremer & Narayanan (2023), therefore, argue that AI cannot be more ethical than its developers, which implies that the responsibility of ethical decision-making eventually remains with humans. Accordingly, they identify ethical AI as a myth, as even with an abundance of training data, AI will not acquire an ethical mind or moral intention (De Cremer & Narayanan, 2023). Cheong (2024) sees the solution in adopting explainable AI (XAI), which could enhance trust and transparency of AI systems. XAI, in combination with robust legal and ethical guidelines, could mitigate potential concerns about AI in ethical decision-making (Cheong, 2024).

2.3.3. Specific AI Use Cases in Ethically Tense Environments

Healthcare – AI applications in the healthcare field have seen a strong increase over the last years (Kerstan et al., 2024). Biller-Andorno et al. (2022) mention several examples where AI systems exhibit high levels of performance and can thus be used for difficult decisions such as patient resuscitation (Biller-Andorno et al., 2022). Additionally, when it comes to taking data-heavy decisions, AI can improve the accuracy in diagnostic fields such as radiology, dermatology, and cardiology (Benzinger et al., 2023). Mahmud et al (2024) also address the use of AI-based systems for disease diagnostics and treatment plans. AI-based decisions in hospitals also reduce stress on medical personnel that is associated with ethical decisions. The burden of regularly taking these challenging ethical decisions can lead to lower-quality decisions or even burnout (Benzinger et al., 2023). These negative effects could be reduced by the introduction of an AI system. Demaree-Cotton et al. (2022) argue that AI improves both accuracy and transparency of decision-making. Criticisms include the previously mentioned black-box effect (Mahmud et al., 2024) and lack of transparency into what values and principles are used for decisions and how these are weighted (Demaree-Cotton et al., 2022). Increased transparency is also necessary to prevent hospitals and healthcare institutions from implementing new biases, which would be financially favorable for them (Benzinger et al., 2023).

Military – AI weapons can reduce human casualties, but questions of morality and ethical responsibility remain (Bellaby, 2021). Differentiating fully autonomous AI from partial autonomy is essential, as the latter is still based on human input. Autonomous AI weapons also create an illusion of delegation of responsibility, since said weapons still fully rely on programming, which is ultimately done by humans (Bellaby, 2021). This confirms De Cremer & Narayanan's (2023) notion that AI can only be as ethical as its creator, which means that responsibility is not actually outsourced. That said, the implementation of autonomous weapons adds an entirely new layer of unpredictability to war (Liao, 2020). While Bellaby (2021) argues that humans remain responsible even for unpredictable AI decisions, which they can neither foresee nor control, Liao (2020) states that autonomous weapon systems create a vacuum of accountability. Choosing the right ethical framework would also be culturally and politically biased (Bellaby, 2021). A further consideration concerns human supervision for ethical decisions in combat situations within split seconds, for which humans lack the processing power (Kohn et al., 2024).

2.4. Managerial Relevance

This chapter outlines topics from managerial literature that bear upon attitudes towards a potential implementation of AI into decision-making. We then derive implications for managers on efficient AI implementation and use in decision-making.

2.4.1. Key Management Theories

Agency theory – The agency theory pertains to the potential problems that can arise from a conflict of interest between a firm's owners and its managers. These conflicts can often be traced back to diverging goals or information asymmetry (Panda & Leepsa, 2017; Trevino, 2024). A central problem that arises from this split of ownership and control is agency costs (Panda & Leepsa, 2017). However, the introduction of AI adds a new element to this mix. AI allows for immediate updates and sharing of intelligent information, which facilitates the reduction of information discrepancies between the parties (Moloi & Marwala, 2020).

Bounded rationality – Connected to the aforementioned information asymmetries, Herbert Simon’s theory of bounded rationality outlines the issues that managers face due to limited information or limited cognitive capacities. The incompleteness of information then leads managers to settle for satisficing rather than optimal solutions (Simon, 1972). Simon (1972) also emphasizes that people often make use of heuristics and biases to make sense of unknown information, which further impacts their decisions. Shick et al. (2023) outline that AI constitutes a potential solution to this conflict by allowing managers to move from bounded rationality to fully rational decisions. The way that AI can achieve this is by allowing immediate access to quantitative analyses and data sharing among decision-makers in companies.

Technology Acceptance Model (TAM) – An additional model, which significantly influences managers' perceptions of AI in decision-making, is the TAM. Originally coined by Davis (1989), the TAM outlines the connection of core constructs such as perceived usefulness and perceived ease of use to people's willingness to accept technologies. This implies that the key factors that shape technology acceptance are, firstly, the matter of how much this technology can improve current performance, and secondly, how little effort it requires to implement this technology.

Unified Theory of Acceptance and Use of Technology (UTAUT) – The UTAUT offers a more detailed perspective on the factors that influence the use of technology. The factors that are identified by Venkatesh (2003) to drive behavioral intentions are performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2003). Performance expectancy pertains to the expected benefit that an individual hopes to gain through the use of the respective technology, making it closely linked to Davis' (1989) concept of perceived usefulness. An additional item that is closely connected is Venkatesh's (2003) effort expectancy, which is comparable to the ease of use in the TAM. However, social influence adds a new layer to the model by considering the opinions of the social environment on the use of the technology. Finally, the facilitating conditions refer to the belief that the necessary infrastructure exists to ensure a smooth implementation of the technology (Venkatesh et al., 2003). Figure 1 shows that the UTAUT dimensions are also subject to moderation effects. As an example, performance expectancy is moderated by gender and age (Alghazi et al., 2021).

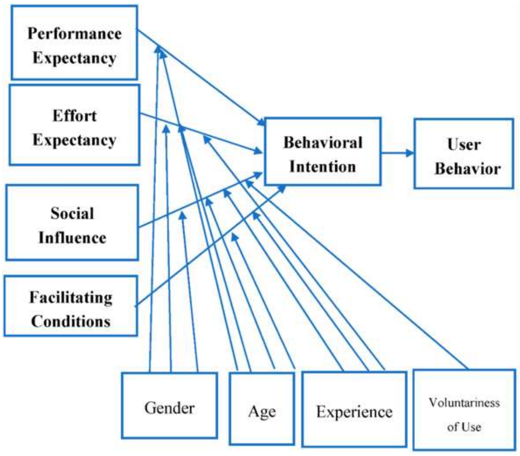


Figure 1: UTAUT model in its original form, coined by Venkatesh (2003).

Source: (Alghazi et al., 2021)

2.4.2. Implications for Managers

Generally, AI can be used to bridge many problems that emerge through issues such as agency theory (Moloi & Marwala, 2020) or bounded rationality (Shick et al., 2023). To capitalize on these benefits, AI needs to be well implemented into a company's decision-making procedures. To ensure a smooth implementation, components of the TAM, such as perceived usefulness and perceived ease of use, need to be communicated to key decision-makers. The UTAUT extends this perspective and also highlights the impacts of moderators on the dimensions that influence behavioral intention (Alghazi et al., 2021). Research by Pathak & Bansal (2024) backs the importance of perceived usefulness not only for implementation but also as an essential component that shapes trust in AI. Additional variables, which influence decision-makers' willingness to implement AI into their decision-making, are considered in the Discussion part of this thesis. In addition to these components, it is essential that AI decisions are transparent and explainable (Balbaa & Abdurashidova, 2024; Shao et al., 2024), and that human insight is essential as AI input is still often subject to errors (Shick et al., 2023). The importance of human oversight is also emphasized in light of the potential biases of the AI (Shick et al., 2023). Frameworks for efficient human-AI cooperation have been developed by different authors, such as De Cremer & Narayanan (2023), who distinguished between human-in-the-loop systems and AI-in-the-loop systems. Human-in-the-loop systems consist of AI-based decision-making, which is eventually accepted or rejected by a human. Meanwhile, AI-in-the-loop systems limit the role of the AI system to providing information about biases and flaws, which can then be taken into consideration by the human decision-maker to improve the decisions.

2.5. Theoretical Considerations

Algorithm appreciation, coined by Logg et al. (2019) and applied by scholars such as Choung et al (2024) to human resource management, is a key concept in the dynamics of trust towards algorithmic decision-making. Additionally, Mahmud et al. (2024) found an increase in algorithm appreciation through familiarity. However, despite these findings, algorithm aversion prevails despite proven superior performance (Dietvorst et al., 2015). People expect superior performance from algorithms (Dzindolet et al., 2002) and lose trust quickly whenever algorithms err (Dietvorst et al., 2015). Additionally, research found that people perceive that algorithms are unable to learn from their past errors, which intensifies aversion (Dietvorst et al., 2015; Sloan & Warner, 2018). This is further enhanced by strong emotional responses associated with such situations, such as post-decision regret when complying with an algorithm (Mahmud et al., 2022). The findings on algorithm aversion caused by errors motivate us to assert that lower algorithmic accuracy reduces trust and thus willingness to delegate. Additionally, studies also emphasized the effect of AI quality on the development of cognitive trust, which in turn influences delegation behavior (Pathak & Bansal, 2024). The first hypothesis, which addresses the first research question of this thesis, therefore, is:

> **H1:** A communicated lower AI accuracy rate (compared to a higher communicated accuracy) of an AI system negatively affects decision-makers' willingness to delegate the strategic decision to it instead of taking the decision themselves.

Even though some authors discuss the advantages of AI in ethical decision-making (Demaree-Cotton et al., 2022), the literature finds that people are rather willing to adopt AI for quantitative, low-stakes decisions (Bigman & Gray, 2018). This stems especially from the fact that people perceive that AI lacks required traits such as ethical sensitivity and empathy, essential for ethical decisions (Benzinger et al., 2023; Choung et al., 2024). An AI's utilitarian decision strategy is mentioned as an additional source of unwillingness of people to trust AI in morally sensitive contexts (Zhang et al., 2022). Many authors find that people consider AI to be unsuitable to handle moral decisions with social consequences (Dietvorst & Bartels, 2022; Bigman & Gray, 2018). In line with this research, we pose the following hypothesis as a potential answer to the second research question:

> **H2:** The presence of ethical complexity in a decision negatively affects decision-makers' willingness to delegate the strategic decision to an AI system rather than making the decision themselves, compared to a decision without ethical implications.

As addressed in the gap in the literature, the interaction of a change in communicated AI accuracy and a decision with versus without ethical complexity has not been discussed in a corporate context that deals with strategic decisions in companies. However, research in other industries, such as healthcare or the military, allowed us to understand the potential effects of such an interaction. Researchers in such fields identified potential issues that arise, especially when AI fails in morally concerning situations, such as the vacuum of accountability that is created by the reliance on AI for these decisions (Liao, 2020). However, ethical contexts increase people's demand for accountability (Bellaby, 2021; Liao, 2020). Additionally, individuals view AI errors in ethical fields not only as technical errors but as moral failures (Parlangeli et al., 2024). Rosenbacke et al. (2024) also find that the negative effects of AI inaccuracies were amplified in ethically charged situations. Similarly, Lone et al. (2025) discuss how the absence of a 100% accurate AI poses a key risk in high-stakes decisions. These observations, combined with other insights discussed, lead us to believe that people are more sensitive to AI errors in high-stakes situations. As we seek to prove that the interaction of ethical complexity and low accuracy amplifies distrust and reduces delegation willingness, the third hypothesis, and therefore the potential answer to the RQ3 is:

> **H3:** The negative effect of a communicated low AI accuracy on decision-makers' willingness to delegate strategic decisions to AI is stronger when the decision has ethical components.

The theoretical framework (Figure 2) consists of two individual constructs: AI accuracy and ethical complexity. These frameworks are operationalized by the measured decision-makers' willingness to delegate the decision. Additionally, according to hypothesis H3, this thesis examines whether the two variables interact in a manner that predicts the willingness of decision-makers to outsource strategic decisions to AI. The goal of this 2x2 experiment is to determine individual and interaction effects of these variables on a decision-maker's willingness to delegate a strategic decision to an algorithm.

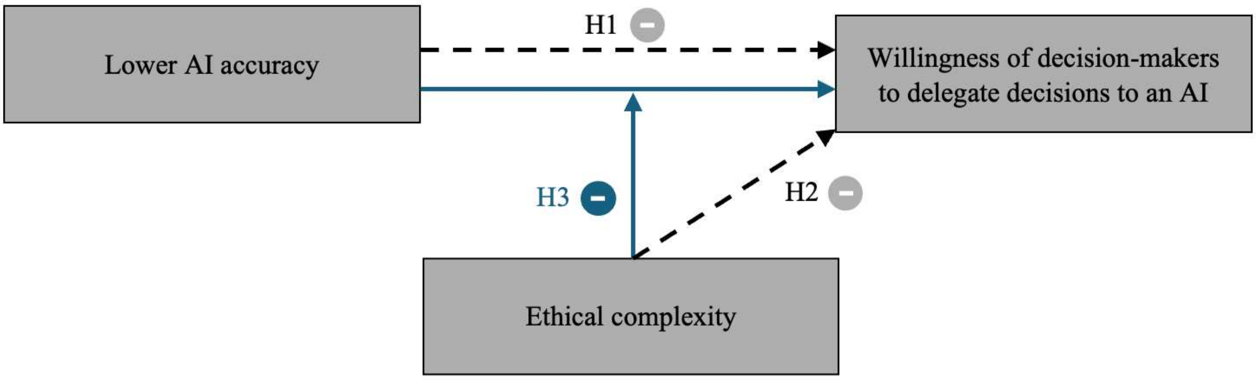


Figure 2: Conceptual framework of this thesis, including all three hypotheses

Note: Dotted black lines depict the main effects while the bold blue line depicts the interaction.

3. Methods

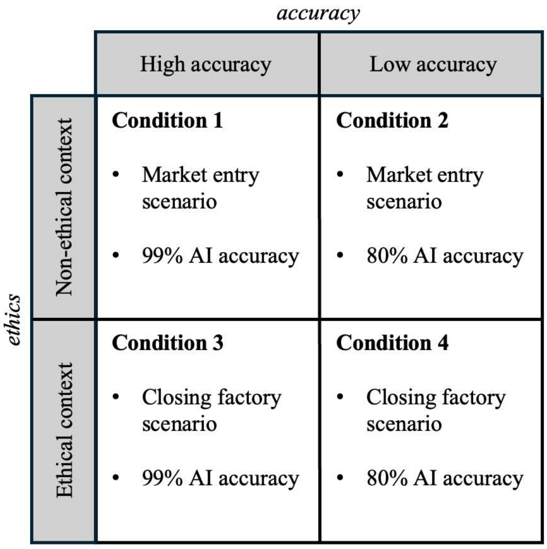
3.1. Experimental Design

Using a 2x2 experiment, this study aimed to test the influence of the two independent variables (IVs) on the dependent variable (DV). The utilization of a 2x2 experiment allowed us to capture the causal effect between those variables best, as we could manipulate single conditions of the experiment and capture the reaction of participants to said changes. A 2x2 design further enabled us to confirm that any changes in the dependent variable can be attributed to the manipulated condition. The conducted experiment not only allowed for testing the main effects but also the interaction effects of these IVs. Participants were randomly assigned to one of the four conditions, which ensured that all groups were balanced and equal in size.

The first IV (*accuracy*) entailed a manipulation of the accuracy of the AI, which the participants could choose to delegate their decision-making to. This accuracy rate was directly communicated to the participant in the assignment of each condition. Two of the conditions were confronted with a high communicated accuracy of the AI. The other two conditions had access to an AI whose accuracy was considerably lower.

The second IV (*ethics*) was developed by manipulating the decision itself. While two of the conditions were confronted with a purely financial decision, the other two were exposed to a decision that had ethical implications. To ensure that participants realize the ethical complexity of their decision, it was stated that the decision would impact the job security of people. On the other hand, the conditions without ethical complexity were instructed that their decision would have major financial implications for the company.

The resulting conditions of the 2x2 experiment are shown in the figure below. To see the exact wording of each condition, please consult Appendix 5.



ethics \ accuracy	High accuracy	Low accuracy
Non-ethical context	**Condition 1** • Market entry scenario • 99% AI accuracy	**Condition 2** • Market entry scenario • 80% AI accuracy
Ethical context	**Condition 3** • Closing factory scenario • 99% AI accuracy	**Condition 4** • Closing factory scenario • 80% AI accuracy

Figure 3: Manipulations per experimental condition

The DV (*delegate*) consisted of the participants' willingness to delegate the strategic decision to an AI. It captured the behavioral intention that participants would pursue in this hypothetical scenario. Even though this experiment was hypothetical, it was supposed to serve as a proxy for how decision-makers in companies would react in such situations. The DV was influenced by the two IVs (*accuracy* and *ethics*) but also by the control variables that we chose to include in the survey and some of the models. The DV (willingness to outsource strategic decisions to AI) was measured on a Likert scale from 1 to 5. One indicated the lowest likelihood to outsource the strategic decision (*Extremely unlikely),* while 5 implied the highest likelihood (*Extremely likely).* Even though a binary scale would have been useful to avoid having too many responses in the middle, a Likert scale was chosen as, in a real-life context, much more information would be needed to make an informed decision of this magnitude. Additionally, a scale allowed for more accurate tracking of the respondents' opinions as it captured the nuances of their decisions. This increased sensitivity would relieve the pressure for decision-makers to make a forced binary choice, which potentially does not fully reflect their opinion.

3.1. Stimuli and Manipulations

3.1.1. AI Accuracy Manipulation

The criteria of the AI accuracy manipulation consisted mainly of the fact that both communicated accuracy percentages needed to be realistically attainable for an AI system, but at the same time, different enough to capture a statistically significant difference between the conditions. High-performing diagnostic AI systems in the medical field inspired the hypothetical communicated accuracy of 99% in conditions 1 and 3 (Battineni et al., 2022). Although AI systems with comparable accuracies are currently not yet used in corporate decision-making, the fact that AI systems with this accuracy exist proves that it is theoretically possible to achieve this level of accuracy. When it came to the selection of the second accuracy rate, it was essential to choose a communicated accuracy still perceived as moderately high by the participants. If the accuracy was too low, none of the participants would likely choose to delegate the AI decision as they consider the AI system entirely unhelpful to their decision-making. These considerations led us to select 80% as the communicated accuracy for conditions 2 and 4.

Conclusions drawn from the literature review imply that this marginal change in communicated accuracy will already result in a different delegation pattern. This includes research by Dietvorst et al. (2015), who found that even smaller perceived flaws or errors lead to disproportionately higher distrust and rejection of algorithmic decision-making. Additionally, other authors also highlighted the importance of perceived AI quality in trust mechanics and thus delegation behavior (Pathak & Bansal, 2024) and in the adoption of technology (Davis, 1989; Venkatesh et al., 2003). The research on these topics led us to hypothesize that the manipulation of a 99% communicated accuracy versus an 80% communicated accuracy will result in significantly different delegation behavior.

3.1.2. Scenario Ethical Complexity Description

The market entry scenario was chosen for conditions 1 and 2 to ensure a neutral context for the decision. By choosing a decision that had purely financial implications with no direct human component attached, we sought to isolate the effect of the presence of ethical complexity on the participants' willingness to outsource the decision. On the other hand, the scenario with ethical considerations was represented by a potential shutdown of a manufacturing site, which would leave employees in the local community without jobs. The participants were only confronted with the situation as a text. This was done to ensure that they would not be influenced by pictures or additional information about the different contexts, which could impact their decision and therefore invalidate the results.

3.2. Design Decisions and Justifications

Further design considerations had to be decided before the launch of the survey. To avoid the previously mentioned vacuum of accountability, which researchers found in AI-based decision-making in ethical contexts (Liao, 2020), the question of responsibility needed to be formulated unambiguously. To achieve this, we explicitly included a sentence in the introduction text of the survey which stated that the participant will be held accountable whether he/she outsourced the decision or took it themselves.

To exclude the possibility of respondents thinking that the AI they were using has never been tested in a corporate setting, we included the fact that this AI has previously been in use in other companies. The detail of only having it in other companies was included to not introducing any kind of bias to the person's decision, which could result from previous experiences inside the hypothetical company.

Even though the scenario is purely hypothetical, the experiment design aimed at creating a scenario that realistically conveyed the fact that this is a high-stakes decision and that the consequences of this decision on others, but also themselves, are considerable. To ensure that respondents were aware of this, we included a part in the scenario description that highlights that the respondent was in the role of a high-level decision-maker in a multinational company. These details were aimed at conveying the magnitude of the decision that the respondent faces.

In addition to the content that focuses on the relation of the IV to the DV, an additional section of attention checks was included. The objective of this section was to improve the data quality by filtering out any respondents who did not pay attention throughout the survey (Abbey & Meloy, 2017). The attention checks asked respondents to recall what kind of decisions they had to take and what the accuracy of the AI was that was at their disposal.

3.3. Justification of Post-Survey Questions

In line with the literature that was presented in the literature review, we included some post-survey questions to measure age, gender, familiarity, perceived fairness, and perceived rationality.

Age was previously outlined as an important variable in algorithm aversion (Araujo et al., 2020; Mahmud et al., 2022), which significantly influences delegation behavior (Ferraz et al., 2025). Similarly, additional studies emphasized the impact of gender on algorithm aversion (Araujo et al., 2020; Ferraz et al., 2025). Finally, age was found to moderate the effect of performance expectancy in the UTAUT model (Alghazi et al., 2021).

As emphasized in the literature review, the impact of familiarity on algorithmic decision-making was studied by Mahmud et al. (2022), Mahmud et al. (2024), and Choung et al. (2024). According to their studies, increased familiarity should positively influence the participants' willingness to outsource the strategic decision. Participants of the survey were questioned about their use of AI to measure their familiarity. The exact content that was shown to respondents can be found in Appendix 5. Due to the impact of familiarity on delegation behavior, familiarity was used as a control variable to isolate further the effect of the IVs.

The same procedure was done with the construct of perceived fairness. Fairness was previously found to be an important driver of people's trust and acceptance of AI applications (Shulner-Tal et al., 2023; Song et al., 2025). While Song et al. (2025) did not directly link delegation behavior, research that was discussed in the literature review found that trust has a significant impact on users' willingness to delegate decisions to algorithms (Dietvorst et al., 2015). Shulner-Tal et al. (2023) also highlight the impact on trust but elaborate on the link between perceived fairness to AI use and acceptance of decisions.

The question on perceived rationality was included in the survey before the completion of the literature review. It was therefore an impulsive decision that was not grounded in a theoretical background. The concept of perceived rationality, therefore, lacked theoretical justification. In accordance with supervisor Peter Rajsingh, the variable was included to see whether it would turn out to be significant.

3.4. Survey and Sampling Procedure

The research was done in a quantitative format, using a 3-minute survey which was created on the platform “Qualtrics” (Qualtrics, 2024). The survey draft was assisted by peer feedback of 10 fellow students who received the survey in advance and communicated potential improvements that were implemented before the official launch.

The survey was then randomly distributed to participants on the platform “Prolific” (Prolific, 2024). This way of distribution ensured that the participants would be selected randomly. Additional benefits of this distribution method included the short time it took to reach the participants as well as the full anonymity that the platform ensures. The survey was published on the 27.03.2025 and was answered by a total of 219 participants, out of which 19 observations could not be counted due to missing answers or time-outs throughout the survey. The goal of 200 participants was reached within 30 minutes. The use of the random allocation function on the website “Prolific” (Prolific, 2024) aimed to ensure that there would be no bias in the population. The use of “Prolific” (Prolific, 2024) guarantees that there would be 50 participants per condition, so that the number of observations per condition was balanced and to ensure results could have statistical power. Out of the 200 participants, 53 results had to be disqualified from the dataset before the analysis due to one or more failed attention checks. This resulted in a slight imbalance of 37, 34, 33, and 43 participants for conditions 1-4, respectively. One additional observation in condition 4 could not be considered due to a missing observation in the age variable. The missing value would have led to an imbalance in the analysis stage, which is why the whole observation could not be considered. This brought the number of participants in condition 4 down to 42 and the overall number of observations to 146. The only pre-screening that was done in the data collection process was to require participants to be 18 years or older and to speak English.

3.5. Analysis and Results

The full analysis of the data set that was needed to get to the results was done in Stata18 (StataCorp, 2023). This analysis consisted of a general descriptive statistics part, assumption checks, tests of balance, a correlation matrix, and finally, the ordered logistic regression models. The regression analysis was performed using an ordered logistic regression to capture the effect of the IVs (communicated accuracy of the AI, prevalence of ethical complexity) on the DV (willingness to outsource the strategic decision). The ordered logistic regression was chosen to account for the fact that the intervals in the Likert scale of the DV are not comparable, but only in order. The regression analysis was composed of two steps. The first step consisted of an assessment of the main effects of the IVs (*accuracy, ethics*) on the dependent variable (*delegate*).

Step 1: The effect of a change in accuracy (*accuracy*) and the presence of ethical complexity of the decision (*ethics*) on the log odds of selecting a higher category in the delegation willingness.

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \hat{\beta}_0^{(j)} + \hat{\beta}_1\mathit{(accuracy)} + \hat{\beta}_2\mathit{(ethics)} + \varepsilon$$

Symbol	Interpretation	Condition comparison
$\hat{\beta}_0^{(j)}$	Cutpoints that are seperating the ordinal categories of the DV (*delegate*)	Depends on category j
$\hat{\beta}_1$	Change in log odds of being in a higher category for *delegate* when accuracy is low (1 vs 0)	Comparison of C1 & C3 to C2 & C4
$\hat{\beta}_2$	Change in log odds of being in a higher category for *delegate* when ethical complexity is present (0 vs 1)	Comparison of C1 & C2 to C3 & C4

Table 1: Interpretation of equation terms without interaction

Step 2: The effect of the interaction of a change in accuracy (*accuracy*) and the presence of ethical complexity of the decision (*ethics*) on the log odds of selecting a higher category in the delegation willingness.

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \hat{\beta}_0^{(j)} + \hat{\beta}_1\mathit{(accuracy} + \hat{\beta}_2\mathit{(ethics)} + \hat{\beta}_3\mathit{(accuracy{*}ethics)} + \varepsilon$$

Symbol	Interpretation	Condition comparison
$\hat{\beta}_0^{(j)}$	Baseline value for log odds when all variables are 0	Depends on category j
$\hat{\beta}_1$	Change in log odds of being in a higher delegation category when accuracy changes from high to low, when ethics = 0 (no ethical component)	Comparison of C1 & C2
$\hat{\beta}_2$	Change in log odds of being in a higher delegation category when the decision context changes from non-ethical to ethical, when accuracy = 0 (high accuracy)	Comparison of C1 & C3
$\hat{\beta}_3$	Interaction term: Additional change in log odds when both accuracy is low and the decision is ethical (anything beyond individual effects)	Comparison of C1 & C4 (vs others)

Table 2: Interpretation of equation terms with interaction

4. Results

4.1. Sample Description

Out of the 146 respondents who were finally included in the analysis, the majority were female (Appendix 2, Table 8). There were 76 women (mean = 0.521), 69 men (mean = 0.473), and one non-binary person who answered the survey and passed both attention checks (n = 1, mean = 0.007). As the potential dummy variable of "*nonbinary*" cannot be interpreted, the creation of the dummy was included in the code but not analyzed in a later stage of the thesis. However, the observation itself was still considered for all the other analyses to ensure that the sample is as representative as possible.

The age of participants ranged from 20 to 72, with a mean of 31.97 and a standard deviation of 9.902. This variable had a missing value, as we considered it to be polite not to make the age question mandatory to answer. It was the only question in the survey that was not mandatory for completion. The anonymity of the survey makes this precaution obsolete, which is why we would not repeat this procedure, as it only leads to unnecessary missing values.

The control variables, which captured respondents' perceptions of fairness and rationality, were measured by asking the participants about their agreement with the following two statements. In the fairness part, the statement was "I trust AI systems to make fair decisions." and for the rationality, it was "I consider AI systems to generally make more rational decisions than humans.". Answer options ranged on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The mean score of both variables was relatively high, with 3.350 for perceived fairness and 3.315 for perceived rationality. These variables were not converted into dummy variables because they are not categorical. However, since this scale is not continuous and we are using a logistic ordinal regression, the absolute value of the coefficients will not be interpretable in the Discussion part.

The variable of familiarity, which is emphasized by scholars and therefore potentially very relevant in the Discussion part of the thesis, was measured by asking the survey respondents the following question: "Do you have prior experience in AI-based decision making?". The answer options were "Yes", "Only with LLMs (large language models such as ChatGPT)", or "No". For the analysis, these three options were converted into *high_familiarity, only_LLMs,* and *no_familiarity,* respectively. As expected, most of the participants had used AI, but only in the form of LLMs (n = 90, mean = 0.616). The second biggest group in this category is composed of people who bring significant prior experience with AI, which extends over just the use of LLMs (n = 36, mean = 0.247). This variable had to be converted into dummy variables because it is a categorical variable, which cannot be properly numerically analyzed.

Another variable that also had to be converted into a set of dummy variables is the education variable. Five different dummies were created so that they could be properly implemented into the models. Most participants surveyed had completed a bachelor's degree as their highest level of education (n = 66, mean = 0.452), followed by high school (n = 40, mean = 0.274).

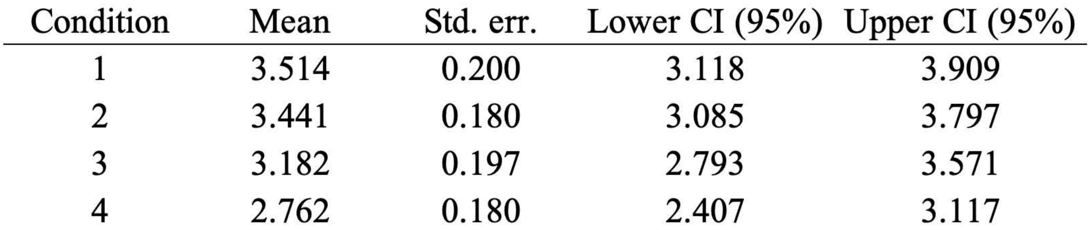
Variable	Obs	Mean	Std. dev.	Min	Max
delegate	146	3.205	1.174	1	5
ethics	146	0.514	0.502	0	1
accuracy	146	0.521	0.501	0	1
accuracy_ethics	146	0.288	0.454	0	1
female	146	0.521	0.501	0	1
nonbinary	146	0.007	0.083	0	1
age	146	31.973	9.902	20	72
fairness	146	3.349	1.074	1	5
rationality	146	3.315	1.131	1	5
high_familiarity	146	0.247	0.433	0	1
only_LLMs	146	0.616	0.488	0	1
no_familiarity	146	0.137	0.345	0	1
high_school	146	0.274	0.448	0	1
bachelor	146	0.452	0.499	0	1
master	146	0.199	0.400	0	1
doctoral_degree	146	0.034	0.182	0	1
other_degree	146	0.041	0.199	0	1

Table 3: Descriptive statistics of the entire sample

4.2. Dependent Variable and Condition Overview

On a scale of 1 (lowest willingness) to 5 (highest willingness), the average was 3.205 for the willingness of delegating the strategic decisions to an AI with a standard deviation of 1.174. Both 1 and the highest value of 5 appeared throughout the responses. As shown in the averages per condition in Table 4, the distribution of the DV was skewed towards the right, as 4 was the most common observation of the DV.

The willingness to outsource a decision to AI differed substantially between the different conditions (Table 4). Throughout the observations, the average to delegate the decision is highest in condition 1 (mean = 3.514) and lowest in condition 4 (mean = 2.762). A declining trend could be observed throughout the conditions, implying that the manipulations across conditions negatively influenced respondent behavior. The effects that could be observed in the table of delegation likelihood per condition were later confirmed through a correlation matrix and later specified in the regression.



Condition	Mean	Std. err.	Lower CI (95%)	Upper CI (95%)
1	3.514	0.200	3.118	3.909
2	3.441	0.180	3.085	3.797
3	3.182	0.197	2.793	3.571
4	2.762	0.180	2.407	3.117

Table 4: Distribution of DV values across conditions.

4.3. Correlation Matrix

To gain additional insights into how the different variables affected the DV, which is the willingness of respondents to outsource the strategic decision to an AI (*delegate*), we conducted a correlation matrix. Only a part of the matrix is included, as shown below in Table 5. This version includes the most important impacts of IVs and controls on the DV.

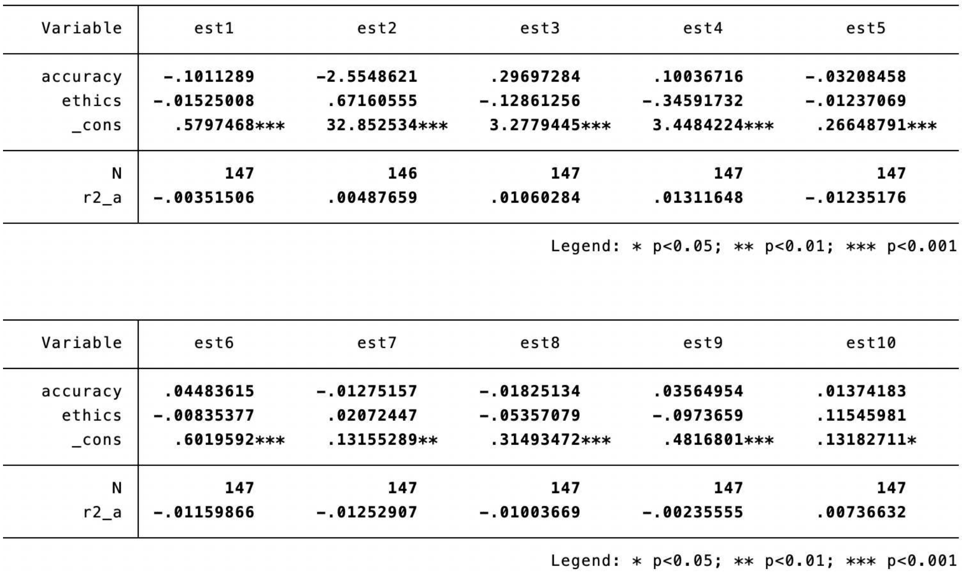
A look into the correlation matrix showed that neither the correlations between the DV (willingness to delegate the strategic decision to AI) and the IV of *accuracy*, nor any of the demographical control variables, were statistically significant. On the other hand, the IV *ethics* showed a weak, negative correlation, which was significant at the 10% level. The negative correlation implies that with an increase in this dummy variable (presence of an ethical complexity versus no ethical complexity), participants were less likely to outsource the strategic decision to AI. While this negative correlation could also be seen in the *accuracy* variable, this variable was statistically insignificant. An additional significant negative correlation can be observed in the interaction term. Table 5 shows the most important correlations with the DV (*delegate).* The full version of this table can be found in Appendix 4 as Figure 9.

Variable	delegate	ethics	accuracy	acc*ethics	female	non-binary	age	fairness	rationality
delegate	1								
ethics	-0.227*	1							
accuracy	-0.124	0.081	1						
accuracy*ethics	-0.241**	0.618***	0.6099***	1					
female	0.075	-0.001	0.094	0.034	1				
non-binary	0.056	-0.085	0.080	-0.053	-0.087	1			
age	-0.047	0.045	0.132	0.078	-0.071	-0.067	1		
fairness	0.566***	-0.067	-0.148*	-0.179**	0.044	-0.105	-0.072	1	
rationality	0.450***	-0.154*	-0.060	-0.191**	-0.072	0.051	-0.042	0.454***	1
high_familarity	0.253***	-0.016	0.040	-0.083	0.104	0.145*	-0.148*	0.274***	0.1643**
only_LLMs	-0.163*	-0.010	-0.052	0.035	-0.052	-0.105	-0.102	-0.137*	-0.067
no_familiarity	-0.087	0.029	0.024	0.055	-0.056	-0.033	0.330***	-0.149*	-0.111
high_school	0.142*	-0.048	0.006	-0.085	-0.148*	-0.051	-0.065	0.072	0.142*
bachelor	0.005	-0.108	-0.037	-0.060	0.183**	0.091	-0.049	0.025	0.027
master	-0.190**	0.1409*	-0.003	0.063	0.000	-0.041	0.081	-0.082	-0.185**
doctoral_degree	0.064	0.033	0.030	0.130	0.030	-0.016	0.027	0.115	-0.019
other_degree	-0.010	0.063	0.061	0.097	-0.147*	-0.017	0.081	-0.164**	0.003

Table 5: Correlation matrix, covering all correlations with the DV.

4.4. Test of Balance

To ensure that all conditions were balanced among themselves when it comes to the different control variables, we conducted a test of balance using a set of regressions for each control variable, which can be seen in Table 6. By regressing each control variable with both IVs, we ensured that none of the experiment conditions were imbalanced. Despite using a logistic ordered regression in this thesis, this test of balance was done using a linear regression as the coefficients themselves are not interpreted. In the resulting table, which summarizes the results for all control variables, none of the coefficients for accuracy or ethics were significant, which implied that the variables are indeed balanced.



```
    Variable |     est1          est2          est3          est4          est5
-------------+-----------------------------------------------------------------------
    accuracy |  -.1011289    -2.5548621     .29697284     .10036716    -.03208458
      ethics | -.01525008     .67160555    -.12861256    -.34591732    -.01237069
       _cons |   .5797468***  32.852534***  3.2779445***  3.4484224***  .26648791***
-------------+-----------------------------------------------------------------------
           N |        147           146           147           147           147
        r2_a | -.00351506     .00487659     .01060284     .01311648    -.01235176
-------------------------------------------------------------------------------------
                                            Legend: * p<0.05; ** p<0.01; *** p<0.001

    Variable |     est6          est7          est8          est9          est10
-------------+-----------------------------------------------------------------------
    accuracy |  .04483615    -.01275157    -.01825134     .03564954     .01374183
      ethics | -.00835377     .02072447    -.05357079    -.0973659      .11545981
       _cons |   .6019592***  .13155289**   .31493472***  .4816801***   .13182711*
-------------+-----------------------------------------------------------------------
           N |        147           147           147           147           147
        r2_a | -.01159866    -.01252907    -.01003669    -.00235555     .00736632
-------------------------------------------------------------------------------------
                                            Legend: * p<0.05; ** p<0.01; *** p<0.001
```

Table 6: Test of balance of all variables

4.5. Hypothesis Testing via Ordered Logistic Regression Models

We tested the hypotheses through 6 ordered logistic regression models. The main regression included Models 1-3, which only measured the main effects of the IVs without the interaction effect. Model 1 included only the IVs (*accuracy* and *ethics*), while Model 2 also considered imbalanced control variables to see what happens if we only controlled for the variables that might bias the results. As there were no imbalanced variables in this dataset, we chose to include control variables that were measured to be statistically significant throughout the ordered logit models: *fairness* and *rationality.* The rest of the control variables were then included in Model 3 to further refine the effect of *accuracy* and *ethics* and to best explore what motivated participants to delegate the decision.

The second step of the regression measured the influence of the interaction term on the model. Once again, this step consisted of 3 models (models 4-6). Model 4 included only the IVs and their interaction term, while models 5 and 6 mirrored the gradual addition of control variables that was previously done in models 2 and 3. The mirroring of the same regression process allowed for a more accurate comparison between the two steps, whose only difference is the interaction term. By only having this one difference, we could directly compare models 1-3 and 4-6 to see the effect of the interaction term.

4.6. Robustness Checks

Before presenting the coefficients of the analysis, it was crucial to conduct the robustness checks to see whether the OLS assumptions were fulfilled.

4.6.1. Robustness of the Dataset

Outliers – We tested specific variables of the dataset for outliers, such as the *age* variable. What can be seen in both the age histogram (Appendix 1, Figure 6) as well as the boxplot (Appendix 1, Figure 7) is that there are some outliers. Even though the furthest observation of 72 was far outside the outer whiskers of the boxplots, we chose to keep it in the dataset, as the value is still plausible.

4.6.2. Robustness of Models 1,2, and 5

Models 1,2, and 5 were identified as the core models. Model 1 measured the immediate effect of our main variables. Model 2 included selected control variables, which increased the explanatory power of the model, and Model 5 added the interaction term. The reason why these models were chosen lies in two factors. Firstly, the Brant test for proportional odds could only be conducted with a limited number of parameters and could therefore not effectively test the model for proportional odds if all control variables were added. To avoid any issues due to an undetected violation of this assumption, we chose to select models 2 and 5 instead of models 3 and 6. Additionally, the entire analysis was previously conducted using a linear regression for another thesis. The resulting adjusted R^2 showed that the inclusion of the rest of the control variables only adds very little explanatory power (R^2 increases by 0.0045) while adding a lot of additional complexity to the model.

Proportional odds – As we used an ordered logistic regression, we needed to check for the proportionality of odds by using a Brant test on all our core models. The Brant test for Model 1 (Appendix 3, Table 9) showed high p-values across all predictors and for the entire model ($p = 0.717$). Due to this high p-value, we accept the null hypothesis, which means that the proportional odds assumption holds for Model 1. The same test was conducted in Model 2 (Appendix 3, Table 10). This time, the p-value was significantly lower ($p = 0.154$), and for some of the variables, the p-value even came close to being significant at the 10% level ($p = 0.108$). However, none of the p-values were as low as 0.05, the null hypothesis of proportional odds could be accepted, meaning that the proportional odds assumption also held for Model 2. Finally, we used the Brant test on Model 5 (Appendix 3, Table 11). The p-value dropped once more to $p = 0.110$. Despite this drop, the null hypothesis could not be rejected. This implies that the assumption of proportionality of odds also held for Model 5. As the assumptions of proportionality of odds hold for all models, we could use the ordered logistic regression with the “ologit” function in Stata (StataCorp, 2023) instead of switching to a generalized ordered logit model (gologit2).

Multicollinearity – An additional step of the model-specific robustness checks consisted of the check for multicollinearity. To do this check, we used the variance inflation factor (VIF). All VIF scores for Model 1 came out at 1.01, which was low, meaning that the IVs among themselves are uncorrelated. The variables in Model 2 generally showed a higher degree of multicollinearity with rationality and fairness, with a VIF of 1.28 (Appendix 3, Table 13).

However, this degree of multicollinearity was still not concerning. Finally, Model 5 showed the highest VIF among the models, which could be attributed to the presence of an interaction term, which logically raised the degree of multicollinearity of a model. There was no need for corrective action as all VIF scores remained below 5 and the average VIF of the model was at only 2.02 (Appendix 3, Table 14).

Homoscedasticity – As ordered logit models, such as the ones we used in this thesis, do not assess the residuals of each observation, the assumption of heteroskedasticity was not needed for our models. Therefore, tests such as the White's test for heteroskedasticity were not conducted. Instead, robust standard errors were used across all models to account for potential violations of assumptions.

In conclusion, the robustness check part of the models, assumptions, proportionality of odds, and no perfect multicollinearity were fulfilled. Due to the nature of this study, potential violations such as heteroskedasticity or independence of errors could not be assessed. However, potential violations were accounted for by the inclusion of robust standard errors.

4.6.3. Regression Results and Hypothesis Evaluation

After ensuring that the models fulfill the required assumptions, the regression results can be evaluated. These results can be found in Table 7. The table also includes values for the adjusted R^2 of each of the models. These values were taken from a previously conducted thesis, which used a linear regression on the same dataset. Although the values for the adjusted R^2 cannot be taken at face value, they are meant to serve as a general indication of the models' explanatory power.

The ordered logistic regression model 1 estimated the direct main effects of the IVs (lower communicated AI accuracy, presence of an ethical complexity) on the DV (likelihood to delegate the decision). Both IVs negatively affected the willingness to delegate the strategic decision to AI (*delegate*). However, the effect of *accuracy* was not statistically significant. On the other hand, the effect of the variable *ethics* was significant at the 5% level. As for all the other models, the model had n = 146 observations. The main constraint of this model lies in the low explanatory power of $R^2 = 0.050$, which undermines the statistical power of the model.

The variable *ethics* remained statistically significant in Model 2; however, this time only at a 10% confidence level. Consistent with the findings of Model 1, *accuracy* remained statistically insignificant. The added control variables of *fairness* and *rationality* turned out to be significant at the 1% level and 5% level, respectively. Additionally, both variables showed a positive coefficient, which implied that participants who perceived AI to make fair or rational decisions, on average, were more likely to delegate the decision to the AI. This model saw a significant increase in the adjusted R^2 to 0.377, which means that it explains around 37.7% of the variance in the DV.

Model 3 then includes all control variables, which renders the *ethics* variable significant at the 10% level. The coefficient of accuracy and ethics remains negative, which means that participants who were exposed to a lower communicated accuracy rate or an ethical component in the decision were less likely to delegate the decision to AI. However, as *accuracy* remains insignificant, this coefficient can only be seen as a directional indication. Both the direction of the coefficients as well as the significances remained the same for the control variables of *fairness* and *rationality*. In addition, the control variable "*master*" turned out to be significant at the 10% level. The coefficient of this dummy was negative, which implies that participants with a completed master's degree, on average, were less likely to delegate the decision.

The second step of the regression analysis included the interaction term, which sought to answer RQ3. The inclusion of this interaction rendered both IVs and the interaction term statistically insignificant in Model 4. All IV effects on the DV were negative. If significant, the *accuracy* coefficient would imply that in conditions without ethical complexity, a communicated low AI accuracy slightly decreased the overall willingness of the delegation (*delegate*). If significant, the negative interaction term could suggest that the effect of low communicated accuracy was more negative for decisions with ethical complexity. This negative effect was, however, only present in model 4. The addition of controls such as *rationality* and *fairness* (models 5 and 6) resulted in the interaction term being positive. While the IVs were both insignificant, the added controls of *fairness* and *rationality* continued to show a strong, positive relation to the DV and statistical significance at the 1% and 5% level, respectively. The final ordered logit model, Model 6, aligned itself with models 2 and 3 regarding significances by finding *ethics*, *fairness,* and *rationality* to be significant. In line with the findings of Model 3, which also included all control variables, Model 6 supported the significance of the *master* variable.

Throughout all 6 models, the coefficient for *ethics* was consistently higher in absolute value than the coefficient for *accuracy*, which suggests that the presence of ethical decision complexity has a stronger (more negative) effect on the willingness of delegation (*delegate*) than a decrease in AI accuracy. The same can be said of the absolute value of fairness vs rationality, only this time, both values were positively influencing the DV.

Table 7: Ordered logistic regression results for models 1-6, measuring the effect of lower communicated AI accuracy (*accuracy*) and presence of ethical complexity (*ethics*) on the survey participants' willingness to delegate the strategic decision to AI (*delegate*).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
accuracy	-0.452	-0.127	-0.294	-0.222	-0.280	-0.538
ethics	-0.838**	-0.757*	-0.825*	-0.609	-0.916	-1.066*
fairness		1.024***	1.013***		1.031**	1.018***
rationality		0.452**	0.427**		0.459**	0.437**
female			0.472			0.492
age			0.008			0.009
high_familiarity			0.334			0.406
only_LLMs			-0.402			-0.377
no_familiarity			(omitted)			(omitted)
high_school			-1.084			-1.029
bachelor			-1.488			-1.464
master			-2.011*			-1.965*
doctoral_degree			-0.657			-0.737
other_degree			(omitted)			(omitted)
accuracy*ethics				-0.431	0.305	0.472
/cut1	-3.089***	1.388	-0.073	-2.983***	1.364	-0.031
/cut2	-1.497***	3.370***	1.996	-1.381***	3.339***	2.031
/cut3	-0.685*	4.458***	3.114*	-0.563	4.425***	3.145*
/cut4	1.508***	7.226***	6.127***	1.624***	7.206***	6.177***
N	146	146	146	146	146	146
r2_a	0.050	0.377	0.382	0.049	0.373	0.378

*Note: *p < 0.1, ** p < 0.05, *** p < 0.01*

Model 1: main effects; Model 2: main effects and significant control variables; Model 3: main effects and all control variables; Model 4: main effects and interaction; Model 5: main effects, interaction, and relevant control variables; Model 6: main effects, interaction, and all control variables.

After summarizing the results of the regression analysis, we will now comment on their implications for the three hypotheses that we sought to answer the initially outlined research questions. These hypotheses were:

> **H1:** A communicated lower AI accuracy rate (compared to a higher communicated accuracy) of an AI system negatively affects decision-makers' willingness to delegate the strategic decision to it instead of taking the decision themselves.
>
> **H2:** The presence of ethical complexity in a decision negatively affects decision-makers' willingness to delegate the strategic decision to an AI system rather than making the decision themselves, compared to a decision without ethical implications.
>
> **H3:** The negative effect of a communicated low AI accuracy on decision-makers' willingness to delegate strategic decisions to AI is stronger when the decision has ethical components.

When it comes to H1, the negative effect of a communicated lower accuracy rate (*accuracy* = 1) could be directly seen in all six ordered logistic regression models as the coefficient for accuracy was consistently negative. However, this coefficient never became statistically significant, which is why we concluded that hypothesis H1 could not be supported by the data. The second hypothesis, which focused on ethics, could partly be supported. The coefficient for *ethics* was consistently negative throughout all regressions. However, it was only statistically significant in the regression models 1-3 and in model 6. Additionally, the coefficients in models 2, 3, and 6 were only significant at the 10% level. Even though Model 1 suggested significance at the 5% level, it was also important to consider that this model only explained about 5% of the variation in the dependent variable. The final hypothesis, H3, was explained through models 4-6. Even though the interaction coefficient was strongly negative in Model 4, it soon turned out positive in models 5 and 6, which contradicted the hypothesis. Additionally, the interaction term was not statistically significant in all three models, which is why H3 could not be supported. Despite its statistical insignificance, we included Figure 4, which shows the graphical depiction of the interaction term. In line with previously described findings of the literature review, both the blue and the red lines follow a negative trend. Additionally, the red line is declining considerably steeper, which could indicate that the impact of low communicated AI accuracy (compared to communicated high accuracy rate) led to a substantially larger decrease in delegation likelihood than low communicated AI accuracy in a context without ethical implications. However, as previously discussed, this hypothesis is not supported, and the negative observed coefficient effect diminishes in models 5 and 6.

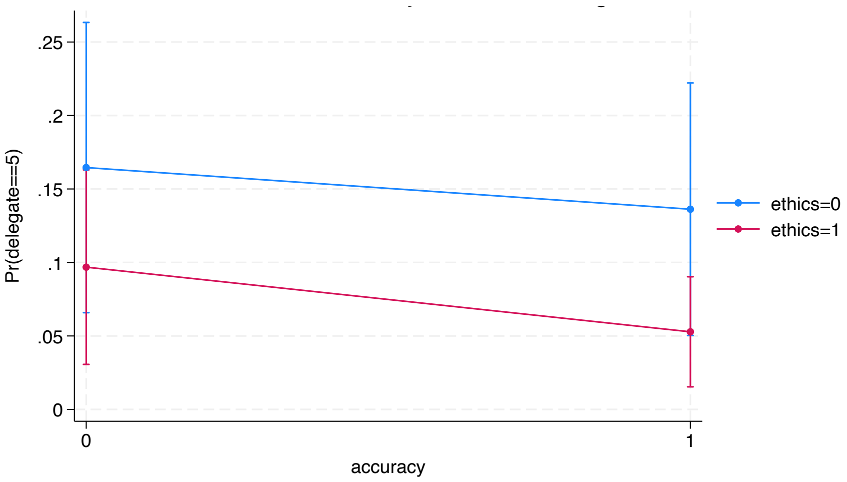


Figure 4: Visualization of the interaction of *accuracy* and *ethics*.

In conclusion, H1 and H3 were found to be unsupported, while H2 was partially backed empirically. However, the findings of H2 were not entirely robust, which meant they should be interpreted cautiously. Even though the coefficients of the IVs were mostly not significant, the consistently negative coefficients across models suggested that this might be a possible real-world trend, which leaves space for further investigation. As the sample size was small and the R^2 only moderately high, the low statistical significance did not imply complete absence of practical relevance.

Figure 5 below visualizes the effect of *accuracy* and *ethics* on *delegation* as well as the impact of the interaction term. This figure is based on the theoretical framework that was covered in the theoretical considerations, however, it includes findings from two different models: models 2 and 5. The reason why these models were chosen and not models 1 and 4, which measured the main and interaction effects of *accuracy* and *ethics* on *delegation,* was that the addition of control variables allows for a more realistic depiction of the coefficients. Both models 1 and 4 explained less than 5% of the variation in the DV. Models 2 and 5 were the better choices due to their adjusted R^2 of 0.377 and 0.373, respectively. The coefficients of model 2 are depicted in black, while the coefficients for *accuracy* and *ethics* of model 5 are shown in blue. The interaction is highlighted in orange. Even though the value of the coefficients cannot be directly interpreted, they can serve to indicate the direction of the relation and significance.

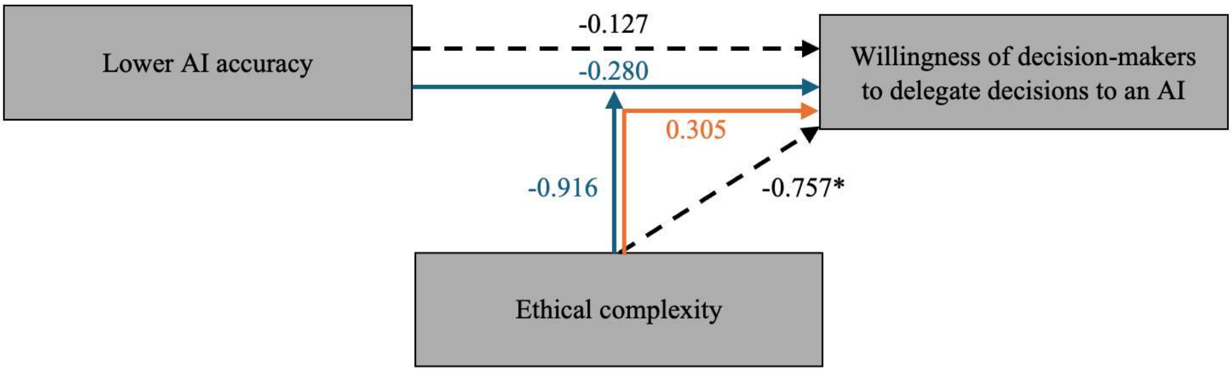


Figure 5: Visualization of the effects of *accuracy*, *ethics*, and the interaction term on the likelihood of delegation. *Note:* $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

5. Discussion

The following part discusses the potential reasons why certain hypotheses could be supported or could not be supported. Additionally, this part also includes the theoretical contributions that this thesis added to existing research, as well as the relevance of this thesis to certain stakeholders. Finally, key limitations and avenues for future research are covered.

5.1. Key Findings

5.1.1. Effect of Communicated AI Accuracy

Starting with the answer to the first research question, hypothesis H1, which sought to prove that participants are less willing to delegate strategic decisions to AI whenever the communicated AI accuracy is lower compared to scenarios in which a high communicated accuracy prevails. As outlined in the results section of this thesis, the observed effect was consistently negative across all ordered logistic regression models. Therefore, the general expected direction of this variable was confirmed. However, the accuracy variable never reached statistical significance. As the hypothesis is not supported, the results cannot back the findings of prior research in the field. The majority of the literature that was discussed in the literature review discussed the disproportionately larger loss of confidence in algorithms after observing errors (Dietvorst et al., 2015). A potential explanation for this dissonance between the literature and the results of the experiment could lie in the emphasis of the literature research on prior algorithmic failure, while the experiment confronted respondents merely with a lower communicated accuracy. Throughout the literature, little emphasis was put on the effect of lower AI accuracy on delegation behavior, which made it difficult to develop a strong theoretical foundation for this hypothesis. The insignificance of the *accuracy* variable could imply that the literature that was found on algorithm aversion after observed algorithmic error cannot be taken as a theoretical foundation of the hypothesis on communicated accuracy rate. This would imply that a simple communication of lower accuracy is not strong enough to trigger algorithm aversion, in contrast to perceived algorithmic failure. Despite this being the most probable explanation for the lack of support for the accuracy hypothesis, research by Pathak & Bansal (2024) previously confirmed that AI quality directly impacts the trust and finally the willingness of decision delegation. An additional explanation of the experiment results could be that, in line with Herbert Simon's theory of bounded rationality, participants were not able to make sense of all the different variables and just went for the satisfying solution due to cognitive or time constraints (Simon, 1972). Even though the accuracy rate was communicated clearly, this information is then filtered through biases, heuristics, and simplifications before a decision is taken. As a communicated accuracy rate does not trigger a large emotional reaction, participants might have struggled to make much sense of this number, which led them to take a hasty decision. A final potential explanation could lie in the research by Aschauer et al. (2024) on the objectives of managers who are using AI for their decision-making. According to their findings, managers primarily used AI to avoid blame for their decisions and not to improve decision accuracy. This could imply that the AI accuracy will not be the most important driver of delegation behavior, but instead the possibility to use the AI as a scapegoat (Aschauer et al., 2024).

Therefore, we conclude that the first research question of "How does a lower communicated AI accuracy impact decision-makers' willingness to delegate strategic decisions to it?" cannot be answered as our hypothesis was not supported due to the negative coefficients of the *accuracy* variable not being statistically significant.

5.1.2. Effect of Ethical Complexity

The second hypothesis revolved around the decrease in the willingness to delegate the strategic decision in situations where an ethical component was included in the decision. This hypothesis was backed in so far as the coefficient for *ethics* was consistently negative throughout all models. The coefficient was statistically significant at the 5% level in model 1 and the 10% level in models 2, 3, and 6. Therefore, H2 was partly supported. Ethical complexity seems to play a role but is not robust enough to sustain through all models to fully confirm the effect. As the statistical significance dropped from the 5% to 10% level once more, the perceptions on AI were added; we argue that the effect of said perceptions overshadows the effect of accuracy and ethics. The findings of Model 1 are additionally restructured due to the low explanatory power of the model ($R^2 = 4.99\%$). Despite these limitations, the partial support of this hypothesis is in line with research by Bigman & Gray (2018), who previously found that people are less likely to delegate decisions in a scenario with ethical implications. The reason for this effect was mainly found to be in the perception that AI lacks consciousness and human traits such as empathy or intent (Benzinger et al., 2023; Bigman & Gray, 2018). Additionally, the measured hesitancy of experiment participants could also underscore the previously found unwillingness of humans to delegate decisions to algorithms due to their consequentialist nature (Dietvorst &

Bartels, 2022). In a real-life context, the result of this experiment could also be traced to the concern of participants to be overly judged by others for using algorithmic decision aids, which could potentially lead to societal backlash or reputational damage (Dietvorst et al., 2015). Additionally, these results support research by Choung et al. (2024) on how AI is well accepted in low-stakes decisions. This could be seen from the experiment data, as respondents in conditions 1 and 2 were likely to delegate the decision. The answer to the second research question of: "How does the ethical complexity of a strategic decision influence decision-makers' willingness to delegate decisions to AI?" will be that in line with the second hypothesis, the presence of ethical complexity in a strategic decision lowers the willingness of decision-makers to delegate the decision to an AI. Due to the insignificance of this variable in models 3 and 4 and the low significance in the other models, this answer is to be taken with caution.

5.1.3. Effect of the Interaction of Accuracy and Ethics

This effect was discussed in the H3, whose coefficients turned out to be statistically insignificant in all models. The interaction effect even has a positive coefficient in models 5 and 6, which contradicts the hypothesis, but also the prior research in this field, which found imperfect accuracy rates to be a large risk in decisions with serious implications (Lone et al., 2025). A potential explanation of this effect could lie in the fact that participants might not be able to grasp all the different concepts that were part of condition 4 (e.g., low accuracy, strategic decision-making, ethical decision components, etc.). According to Simon's (1972) theory of bounded rationality, this problem could lead participants to settle for the satisficing option and considering that the accuracy of 80% is good enough to delegate the decision without investing further time into it. Another potential explanation could include that some participants of condition 3 (99% accuracy, ethical component) found the existence of an AI tool that scores an accuracy of 99% in ethically complex scenarios unrealistic and discounted the manipulation entirely, which resulted in a greater spread of results and therefore no statistical significance. Finally, the third research question of "How does the presence of an ethical decision component change the impact of a low communicated accuracy on decision-makers' willingness to delegate strategic decisions to AI?" cannot be explicitly answered, as the interaction coefficient was low and insignificant. The insignificance of the interaction effects makes it likely that the additional effect of an ethical decision in a context of low communicated accuracy will not be extraordinarily high. However, all significances and variable interpretations are to be taken with caution, as they could potentially be attributed to the low sample size and low adjusted R^2.

5.1.4. Role of Control Variables

Age – The impact of age on algorithm aversion was studied by authors such as Araujo et al. (2020) and Ferraz et al. (2025). Additionally, researchers such as Brink et al. (2023) consider the importance of age and the connected years of experience on delegation behavior. Due to these findings, *age* was included as a control variable. However, the effect of the *age* variable was very low, positive, and not statistically significant, which means that it cannot support the effects that were observed throughout the literature.

Gender – A similar pattern occurred for the gender control variable: *female*, which was, as research found, significant (Ferraz et al., 2025; Mahmud et al., 2022). Against the findings of Ferraz et al. (2025) and Mahmud et al. (2022), the effect of *female* was positive but could not be considered for the analysis as it was not statistically significant. This finding is in line with the findings of other researchers (Brink et al., 2023).

Education – While research on this topic is divided, with some scholars finding education to be important for technology (Alghazi et al., 2021) and AI (Brink et al., 2023) adoption behavior, all *education* dummies of the models were statistically insignificant. Against the findings of previously mentioned researchers, the education dummies with the most observations (*high_ school, bachelor, master*) showed that the willingness to delegate the decision generally decreased with rising education level.

Familiarity – Despite being outlined as an important variable throughout several pieces of literature (Brink et al., 2023; Choung et al., 2024; Mahmud et al., 2022), the dummy variables of familiarity were also statistically insignificant in the ordered logistic regression models.

Fairness – Fairness constitutes the first statistically significant control variable of the models. Throughout all 4 models in which it is included, fairness is significant at the 1% level and its impact is significantly positive. These findings align with the research in this field, which highlights the importance of fairness for trust (Song et al., 2025) and for the acceptance of AI decisions (Shulner-Tal et al., 2023). Implications that could be drawn from these significances include the importance of the development of an AI system that can make fair, data-driven decisions and communicate them to decision-makers accordingly.

Rationality – The final control variable of rationality could not be backed up with literature. Nonetheless, *rationality* had a positive impact on *delegate,* which was not as high as *fairness,* but still significant to the 5% level in models 2 and 5, and significant to the 10% level once the rest of the control variables are included. Due to the significance of this variable, we chose to keep it in the models. The significance of this variable once again highlights the importance of proving that an AI can make rational decisions to decision-makers.

5.2. Theoretical Contributions

This thesis contributes to the discussion of algorithm aversion and appreciation by partly backing prior research by understanding which psychological aspects motivate people to entrust decisions to an AI system. By doing this, the results of the 2x2 experiment confirmed some of the previously outlined factors of other scholars, such as the importance of ethical context (Benzinger et al., 2023; Cheong, 2024; Parlangeli et al., 2024) and perceived fairness (Shulner-Tal et al., 2023; Song et al., 2025) on whether people trust an AI system to make decisions. A new psychological aspect was also uncovered by the inclusion of rationality into the models, which had previously not been studied.

This thesis could also be considered as an extension of the TAM and UTAUT in the context of the use of AI. The TAM, which was previously only focused on perceived usefulness and ease of use, would therefore be extended through the introduction of ethical and specific performance considerations like the communicated accuracy rate. The same is true for the UTAUT, which also considers performance aspects but not in the context of communicated accuracy rate.

Additionally, the thesis tested a new interaction hypothesis, which has not previously been studied in a 2x2 environment. Although the findings of this effect did not align with the expected hypothesis, this thesis may open this subject for further research by other scholars. As mentioned in the existing literature gap, the application of these exact variables in a strategic decision-making context has not been previously explored. However, as the introduction considered AI in strategic decision-making to be an engine of the next industrial revolution (Bristol et al., 2023) and a driver of competitive advantage (Halim et al., 2024), we hope that the thesis lays a foundation for future research in this field.

5.3. Practical Implications

This thesis not only constitutes a contribution to the literature but should also result in practical implications. These practical implications combine literature and results of this thesis to draw recommendations for managers on how to best implement AI into corporate decision-making. Afterward, this part specifies to whom these recommendations will be most relevant.

5.3.1. Recommendations for Managers

The first recommendation, which was derived from literature, is to use AI only as a decision support but not a complete replacement (De Cremer & Narayanan, 2023; Demaree-Cotton et al., 2022). Such implementation would prevent a loss of control and ensure that AI is used to augment human decision-making. Potential systems that were proposed include Human-in-the-loop systems and AI-in-the-loop systems. These systems also preserve the accountability of decision-makers.

In addition to having AI as a decision aid, the characteristics that positively influence people's willingness to use AI for their decisions need to be emphasized. Such characteristics include the TAM factors of usefulness and ease of use (Davis, 1989), but also characteristics that were found in this thesis, such as fairness and rationality of the AI decisions. As the thesis confirmed that individuals who perceived AI to make fair and rational decisions were much more likely to delegate decisions to it, it might be a valid strategy to educate people on the fairness and rationality of said AI systems. To enable this, the AI cannot be a Blackbox (Mahmud et al., 2024) but instead, it needs to be explainable (XAI) (Cheong, 2024).

As this thesis also found people to be hesitant to use AI for decisions that involve higher stakes or ethical questions, the introduction of AI decision systems in companies should start with low stakes and quantitative decisions, which is coherent with the findings of other researchers (Bigman & Gray, 2018; Parlangeli et al., 2024).

5.3.2. Practical Relevance

This thesis will be most relevant for managers and decision-makers in companies. The results allow them to understand under which conditions AI in decision-making is accepted or rejected. These insights can be used to implement AI more efficiently and to prepare for potential obstacles to integration. The same insights will be valuable for AI system developers. The thesis analyzes user concerns about ethical sensitivity, accuracy concerns, and trust, which can be used by developers to account for them during the development of AI systems.

Policy makers could also benefit from the findings of this thesis. The results elaborate on situations in which human oversight is necessary, which pertains especially to ethically loaded decision contexts. The findings of this thesis could allow them to justify their decisions in regard of regulatory decisions in the field of AI decision-making.

Finally, as this thesis can be seen as an extension of existing models such as TAM and UTAUT in the field of AI, it becomes relevant to other researchers who are studying the interaction of humans and AI.

5.4. Limitations

Despite the previously outlined practical relevance of the thesis results, not all findings can be taken at face value, as there are several limitations throughout the thesis that need to be addressed and that would be improved in the next thesis.

Starting with the question of sample representativeness, while using the platform Prolific (Prolific, 2024) for the collection of survey responses ensured a full random sample, it also meant that we were not targeting the exact population, which is described in the thesis title. This thesis is designed to study the behavior of decision-makers in companies. However, it is safe to assume that the vast majority of the survey's respondents are not decision-makers in companies. It is also rather likely that many respondents are not familiar with or interested in the corporate context. The respondents who were recruited through this online platform, therefore, differ substantially from real-life decision-makers who are confronted with such a situation. This limits the validity and applicability to real-life corporate contexts.

At the same time, the participants knew about the highly hypothetical nature of this experiment. As their decisions were non-consequential, which could have reduced their emotional connection and risk sensitivity. The result deviations that are caused because of this phenomenon further limit the applicability to a real-life context.

An additional limitation presents itself in the framing of the conditions with ethical complexity. While conditions 1 and 2 revolve around a market entry, which could be perceived as a potential gain, conditions 3 and 4 present themselves as a loss frame and emphasize potential harm. Research by Tversky & Kahneman (1981) finds that such framing effects can have a significant impact on outcomes. This means that a part of the differences in conditions 1&2 versus 3&4, and therefore in the variable *ethics,* could be attributed to the unwanted framing effects.

Another limitation of the framing of conditions 3&4 consists in the fact that the social consequences that are occurring in the ethical complexity conditions are in addition to the still prevalent financial implications for the company. This makes it increasingly difficult to consider the isolated effect of the prevalence of ethical complexity. Both this limitation as well as the previous one were discussed with the thesis supervisor, Gwendolin Sajons, in advance of the surveying and writing of the thesis. Our discussion concluded that we are still to continue with the approach, but that these limitations need to be critically evaluated in this section of the thesis.

5.5. Suggestions for Future Research

The first avenue for future research refers to one of the central limitations of this thesis. This study used a general population sample, which does not entirely reflect a comparable experience or accountability pressure as a sample in a business environment. A future potential survey could aim to recruit specifically participants working in corporate environments and bringing real-world decision-making experience. This study could additionally include other independent variables. While this thesis developed the concepts of differences in communicated AI accuracy and the presence of an ethical context, another thesis could study decision-making behavior in the light of different personality traits (Ferraz et al., 2025) or compare AI perception across different cultural contexts (Park & Jung, 2025). The inclusion of cultural contexts could be especially interesting as particular cultures are hesitant to replace the human component despite high perceived usefulness and ease of use (Kelly et al., 2023).

The quantitative nature of this thesis implied that participants were not able to give a qualitative reasoning for their decisions. While decisions that potentially involve ethical components may involve subtle emotional or moral dilemmas for survey participants, none of these effects could be captured through this quantitative analysis. Therefore, future research could add a qualitative component to understand the persons' emotions and to potentially explain why some hypotheses could not be supported. A qualitative component could also provide further insights that could help to understand what would motivate people to be more willing to outsource strategic decisions with ethical components to algorithmic decision-making.

But even under the consideration of the same IVs, further studies could be done. For example, the analysis could be extended throughout a longer period. As Jussupow (2024) stated, a user's interaction with AI is not static but instead, trust and delegation behavior changes over time. Further studies could investigate how behavior changes throughout prolonged engagement or how a one-time error sustainably affects trust, especially under ethical complexity. This could potentially also involve more complex experiments with several decision stages.

The literature review of this thesis also mentioned the impact of XAI and the general impact of transparency (Cheong, 2024) in AI decision-making to improve trust in the respective systems. Further experiments could be conducted under the consideration of these additional factors to discover how this would impact respondents' willingness to delegate the decisions in scenarios with a lower communicated error rate and decisions with ethical considerations.

Finally, the vacuum of responsibility (Liao, 2020) could also be studied. While we had to clarify that the respective decision-maker will be fully responsible for the impact of their decision, another study could elaborate on the dynamics of accountability of AI decision-making by questioning participants about their perceived responsibility for the decision.

6. Conclusion

The main objective of this thesis was to conduct a 2x2 experiment to understand how communicated AI accuracy and ethical components of decisions influence decision-makers' willingness to delegate the strategic decision to an AI system.

The thesis could not confirm the findings of prior research that highlights the importance of perceived performance and accuracy in technology acceptance (Davis, 1989; Pathak & Bansal, 2024; Venkatesh et al., 2003), that a lower communicated AI accuracy rate significantly impacts people's willingness to delegate a strategic decision to an AI. On the other hand, studies on algorithm aversion in decisions with ethical complexity (Bigman & Gray, 2018) could be partially confirmed by the *ethics* variable, which is statistically significant at the 10% level until the introduction of the interaction term. Meanwhile, the interaction of accuracy and ethics was found to be statistically insignificant across all models. The thesis confirms the importance of perceived fairness (Shulner-Tal et al., 2023; Song et al., 2025) at the 1% level and additionally finds perceived rationality to be significant at the 10% level.

The findings of this thesis contribute to existing literature that studies factors that cause algorithm aversion (Dietvorst et al., 2015; Mahmud et al., 2022) and highlight that aversion even persists whenever high AI accuracy is communicated. Through these findings, the thesis adds new layers to the perception of the TAM (Davis, 1989) in the context of AI implementation. Moreover, the thesis provides practical relevance to managers and policymakers who seek to understand the dynamics of algorithm aversion to ensure an efficient implementation of AI systems into decision-making.

The limitations of this study mainly consisted of the lack of real-world decision-makers in the survey population, as well as difficulties to isolate the effect of *ethics* due to unwanted framing effects and the consideration of ethical costs as an additional cost in addition to the financial cost. Future studies could therefore be conducted to understand the reasoning of respondents through a qualitative component.

In conclusion, this thesis finds that while rationality and fairness can encourage delegation, people are hesitant to delegate decisions, especially when ethical decision components are involved. Therefore, the thesis finds that while AI in strategic decisions is the future (Gartner, 2023), the role of AI lies in augmenting human decision-making, especially in high-stakes decisions.

7. References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53–56*, 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Agbabiaka, O., Ojo, A., & Connolly, N. (2025). Requirements for trustworthy AI-enabled automated decision-making in the public sector: A systematic review. *Technological Forecasting and Social Change*, *215*, 124076. https://doi.org/10.1016/j.techfore.2025.124076

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, *9*(1), 1. https://doi.org/10.1186/s41235-023-00529-3

Alarcon, G. M., Lyons, J. B., Hamdan, I. aldin, & Jessup, S. A. (2024). Affective Responses to Trust Violations in a Human-Autonomy Teaming Context: Humans Versus Robots. *International Journal of Social Robotics*, *16*(1), 23–35. https://doi.org/10.1007/s12369-023-01017-w

Alghazi, S. S., Kamsin, A., Almaiah, M. A., Wong, S. Y., & Shuib, L. (2021). For Sustainable Application of Mobile Learning: An Extended UTAUT Model to Examine the Effect of Technical Factors on the Usage of Mobile Devices as a Learning Tool. *Sustainability*, *13*(4), Article 4. https://doi.org/10.3390/su13041856

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

Aschauer, F., Sohn, M., & Hirsch, B. (2024). *Managerial advice-takingSharing responsibility with (non)human advisors trumps decision accuracy.*

Balbaa, M. E., & Abdurashidova, M. S. (2024). (PDF) The Impact of Artificial Intelligence in Decision Making: A Comprehensive Review. *ResearchGate*. https://doi.org/10.36713/epra15747

Battineni, G., Chintalapudi, N., Hossain, M. A., Losco, G., Ruocco, C., Sagaro, G. G., Traini, E., Nittari, G., & Amenta, F. (2022). Artificial Intelligence Models in the Diagnosis of Adult-Onset Dementia Disorders: A Review. *Bioengineering*, *9*(8), Article 8. https://doi.org/10.3390/bioengineering9080370

Bellaby, R. W. (2021). Can AI Weapons Make Ethical Decisions? *Criminal Justice Ethics*, *40*(2), 86–107. https://doi.org/10.1080/0731129X.2021.1951459

Bennett, N., Lemoine, G. J., & Molnar, P. (2024). *USING ARTIFICIAL INTELLIGENCE to Navigate the Risks of a VUCA Environment: Boards must be savvy about the benefits of AI, but also be aware of the limitations of current solutions*. p28-32. edsinc.

Benzinger, L., Ursin, F., Balke, W.-T., Kacprowski, T., & Salloch, S. (2023). Should Artificial Intelligence be used to support clinical ethical decision-making? A systematic review of reasons. *BMC Medical Ethics*, *24*(1), 48. https://doi.org/10.1186/s12910-023-00929-6

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Biller-Andorno, N., Ferrario, A., Joebges, S., Krones, T., Massini, F., Barth, P., Arampatzis, G., & Krauthammer, M. (2022). AI support for ethical decision-making around resuscitation: Proceed with care. *Journal of Medical Ethics*, *48*(3), 175–183. https://doi.org/10.1136/medethics-2020-106786

Brink, A., Benyayer, L.-D., & Kupp, M. (2023). Decision-making in organizations: Should managers use AI? *Journal of Business Strategy*, *45*(4), 267–274. https://doi.org/10.1108/JBS-04-2023-0068

Bristol, H., De Boer, E., De Kroon, D., Shahani, R., & Torti, F. (2023). *Adopting AI in manufacturing at speed and scale | McKinsey*. https://www.mckinsey.com/capabilities/operations/our-insights/adopting-ai-at-speed-and-scale-the-4ir-push-to-stay-competitive?utm_source=chatgpt.com

Brown, S. (2021, April 21). *Machine learning, explained | MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*. https://doi.org/10.3389/fhumd.2024.1421273

Choi, S., Kang, H., Kim, N., & Kim, J. (2025). How does artificial intelligence improve human decision-making? Evidence from the AI-powered Go program. *Strategic Management Journal*, *n/a*(n/a). https://doi.org/10.1002/smj.3694

Choung, H., Seberger, J. S., & David, P. (2024). When AI is Perceived to Be Fairer than a Human: Understanding Perceptions of Algorithmic Decisions in a Job Application Context. *International Journal of Human–Computer Interaction*, *40*(22), 7451–7468. https://doi.org/10.1080/10447318.2023.2266244

Copeland, B. J. (2025, April 30). *Artificial intelligence (AI) | Definition, Examples, Types, Applications, Companies, & Facts | Britannica*. https://www.britannica.com/technology/artificial-intelligence

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors. *Strategy Science*, *9*(4), 322–345. https://doi.org/10.1287/stsc.2024.0190

CSRC. (n.d.). *Algorithm—Glossary | CSRC*. Retrieved May 3, 2025, from https://csrc.nist.gov/glossary/term/algorithm

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

De Cremer, D., & Narayanan, D. (2023). How AI tools can—And cannot—Help organizations become more ethical. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1093712

Demaree-Cotton, J., Earp ,Brian D., & and Savulescu, J. (2022). How to Use AI Ethically for Ethical Decision-Making. *The American Journal of Bioethics*, *22*(7), 1–3. https://doi.org/10.1080/15265161.2022.2075968

Dietvorst, B. J., & Bartels, D. M. (2022). Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies. *Journal of Consumer Psychology*, *32*(3), 406–424. https://doi.org/10.1002/jcpy.1266

Dietvorst, B. J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

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

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. *Ethics and Information Technology*, *18*(2), 149–156. https://doi.org/10.1007/s10676-016-9400-6

Ferraz, V., Houf, L., Pitz, T., Schwieren, C., & Sickmann, J. (2025). Trust in the machine: How contextual factors and personality traits shape algorithm aversion and collaboration. *Computers in Human Behavior Reports*, *17*, 100578. https://doi.org/10.1016/j.chbr.2024.100578

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

Gil de Zúñiga, H., Goyanes ,Manuel, & and Durotoye, T. (2024). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research. *Political Communication*, *41*(2), 317–334. https://doi.org/10.1080/10584609.2023.2290497

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Halim, H. A., Waqas, A., Hanifah, H., & Ahmad, N. H. (2024). *Strategic foresight and big data analytics as antecedents of SMEs' sustainable competitive advantage: Role of AI utilization*. *16*(4).

Iqbal, S. (2024). The Intelligence Spectrum: Unraveling the Path from ANI to ASI. *Journal of Computing & Biomedical Informatics*, *7*(02), Article 02. https://jcbi.org/index.php/Main/article/view/779

Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making. *MIS Quarterly*, *48*(4), 1575–1590. https://doi.org/10.25300/MISQ/2024/18512

Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925. https://doi.org/10.1016/j.tele.2022.101925

Kerstan, S., Bienefeld, N., & Grote, G. (2024). Choosing human over AI doctors? How comparative trust associations and knowledge relate to risk and benefit perceptions of AI in healthcare. *Risk Analysis*, *44*(4), 939–957. https://doi.org/10.1111/risa.14216

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, *160*, 108352. https://doi.org/10.1016/j.chb.2024.108352

Kohn, S., Cohen ,Marvin, Johnson ,Athena, Terman ,Mikhail, Weltman ,Gershon, & and Lyons, J. (2024). Supporting Ethical Decision-Making for Lethal Autonomous Weapons. *Journal of Military Ethics*, *23*(1), 12–31. https://doi.org/10.1080/15027570.2024.2366094

Kritikos, M., & Iphofen, R. (2023). Adopt a Broad Definition of AI and Algorithmic Management. In *Ethical AI Surveillance in the Workplace* (world; Vol. 10, pp. 79–79). Emerald Publishing Limited. https://doi.org/10.1108/S2398-601820230000010034

Kvam, P. D., Sokratous, K., Fitch, A., & Hintze, A. (2024). Using artificial intelligence to fit, compare, evaluate, and discover computational models of decision behavior. *Decision*, *11*(4), 599–618. https://doi.org/10.1037/dec0000237

Liao, S. M. (2020). *Ethics of Artificial Intelligence*. Oxford University Press.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lone, M. R., Naruka, Y. S., Viswhanath, S., & Pandey, A. (2025). Balancing AI accuracy and medical guidance: Evaluating ChatGPT's role in UTI-related queries. *Infectious Diseases Now*, *55*(3), 105051. https://doi.org/10.1016/j.idnow.2025.105051

Luger, G. F., LUGER, Stubblefield, & Stubblefield, W. A. (1993). *Artificial intelligence: Structures and strategies for complex problem solving* (2nd ed.). Benjamin/Cummings Pub. Co. https://www.uoitc.edu.iq/images/documents/informatics-institute/exam_materials/artificial%20intelligence%20structures%20and%20strategies%20for%20%20complex%20problem%20solving.pdf

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 and Social Change*, *175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mahmud, H., Islam, A. K. M. N., Luo, X. (Robert), & Mikalef, P. (2024). Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decision Support Systems*, *179*, 114168. https://doi.org/10.1016/j.dss.2024.114168

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 57–87. https://doi.org/10.1177/1555343411433844

Moloi, T., & Marwala, T. (2020). The Agency Theory. In T. Moloi & T. Marwala (Eds.), *Artificial Intelligence in Economics and Finance Theories* (pp. 95–102). Springer International Publishing. https://doi.org/10.1007/978-3-030-42962-1_11

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4), Article 4. https://doi.org/10.1609/aimag.v27i4.1911

Panda, B., & Leepsa, N. M. (2017). Agency theory: Review of Theory and Evidence on Problems and Perspectives. *Indian Journal of Corporate Governance*, *10*(1), 74–95. https://doi.org/10.1177/0974686217701467

Park, J., & Jung, Y. (2025). Exploring Cultural Differences in AI-Based Interviews: Innovativeness and Justice Perceptions Among Job Applicants in the United States and South Korea. *Human Resource Management*, *n/a*(n/a). https://doi.org/10.1002/hrm.22303

Parlangeli, O., Curro', F., Palmitesta, P., & Guidi, S. (2024). Moral judgements of errors by AI systems and humans in civil and criminal law. *Behaviour & Information Technology*, *43*(9), 1718–1728. https://doi.org/10.1080/0144929X.2023.2283622

Pathak, A., & Bansal, V. (2024). AI as decision aid or delegated agent: The effects of trust dimensions on the adoption of AI digital agents. *Computers in Human Behavior: Artificial Humans*, *2*(2), 100094. https://doi.org/10.1016/j.chbah.2024.100094

Prolific. (2024). *Participant recruitment platform*. https://www.prolific.com

Qualtrics. (2024). *Qualtrics Experience Management Platform* [Computer software]. Qualtrics. https://www.qualtrics.com

Renier, L. A., Schmid Mast, M., & Bekbergenova, A. (2021). To err is human, not algorithmic – Robust reactions to erring algorithms. *Computers in Human Behavior*, *124*, 106879. https://doi.org/10.1016/j.chb.2021.106879

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, *33*(1), 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Rosenbacke, R., Melhus, Å., & Stuckler, D. (2024). False conflict and false confirmation errors are crucial components of AI accuracy in medical decision making. *Nature Communications*, *15*(1), 6896. https://doi.org/10.1038/s41467-024-50952-3

Russell, S. J., & Norvig, P. (with Chang, M., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I., Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J.). (2021). *Artificial intelligence: A modern approach* (Fourth Edition). Pearson.

Shao, C., Nah, S., Makady, H., & McNealy, J. (2024). Understanding User Attitudes Towards AI-Enabled Technologies: An Integrated Model of Self-Efficacy, TAM, and AI Ethics. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2024.2331858

Shick, M., Johnson, N., & Fan, Y. (2023). Artificial intelligence and the end of bounded rationality: A new era in organizational decision making. *Development and Learning in Organizations: An International Journal*, *38*(4), 1–3. https://doi.org/10.1108/DLO-02-2023-0048

Shulner-Tal, A., Kuflik, T., & Kliger, D. (2023). Enhancing Fairness Perception – Towards Human-Centred AI and Personalized Explanations Understanding the Factors Influencing Laypeople's Fairness Perceptions of Algorithmic Decisions. *International Journal of Human–Computer Interaction*, *39*(7), 1455–1482. https://doi.org/10.1080/10447318.2022.2095705

Simon, H. A. (1972). *Theories of bounded rationality*. https://iiif.library.cmu.edu/file/Simon_box00065_fld05032_bdl0001_doc0001/Simon_box00065_fld05032_bdl0001_doc0001.pdf

Sloan, R. H., & Warner, R. (2018). *When Is an Algorithm Fair? Errors, Proxies, and Predictions in Algorithmic Decision Making* (SSRN Scholarly Paper No. 3163664). Social Science Research Network. https://doi.org/10.2139/ssrn.3163664

Song, Y., Li, C., Xing, W., Lyu, B., & Zhu, W. (2025). Investigating perceived fairness of AI prediction system for math learning: A mixed-methods study with college students. *The Internet and Higher Education*, *65*, 101000. https://doi.org/10.1016/j.iheduc.2025.101000

StataCorp. (2023). *Stata Statistical Software: Release 18* [Computer software]. StataCorp LLC.

Trevino, G. A. (2024). *The Relationship Between Agency Theory and Lack of Control Discounts*.

Turing, Alan. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Zhang, Z., Chen, Z., & Xu, L. (2022). Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. *Journal of Experimental Social Psychology*, *101*, 104327. https://doi.org/10.1016/j.jesp.2022.104327

8. Appendix

Appendix 1: Relevant figures and tables for robustness checks

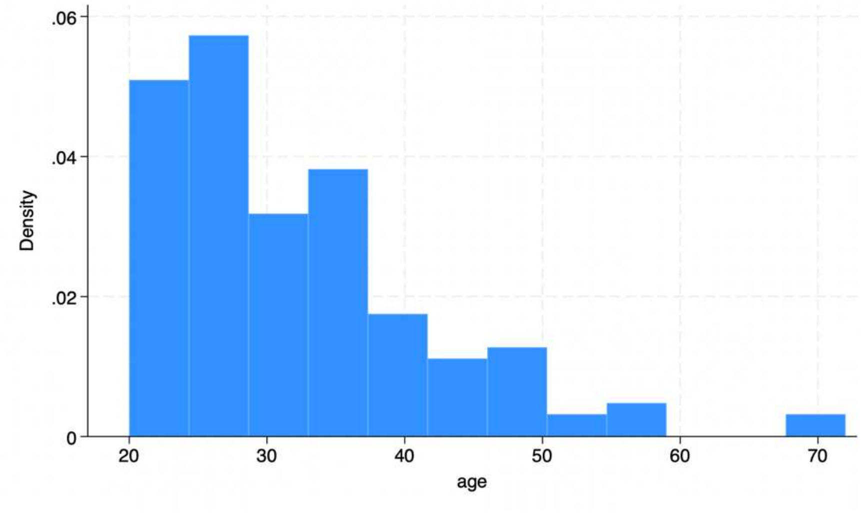


Figure 6: Histogram of the distribution of the *age* variable.

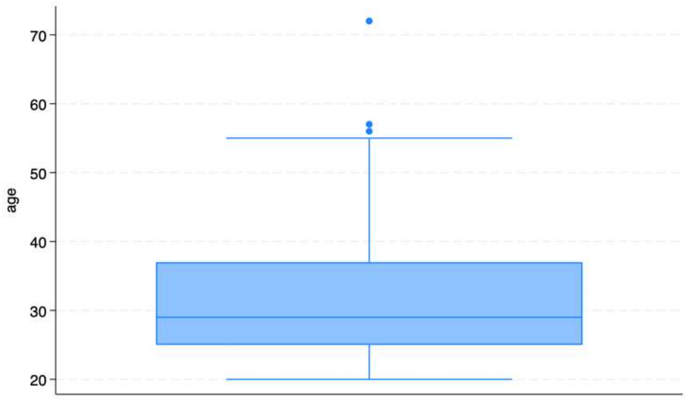
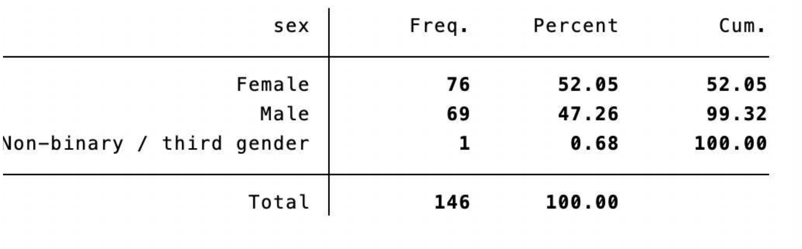


Figure 7: Boxplot of the distribution of the *age* variable.

Appendix 2: Relevant figures and tables for descriptive statistics



```
                     sex |      Freq.     Percent        Cum.
-------------------------+-----------------------------------
                  Female |         76       52.05       52.05
                    Male |         69       47.26       99.32
Non-binary / third gender |          1        0.68      100.00
-------------------------+-----------------------------------
                   Total |        146      100.00
```

Table 8: Gender composition of the sample

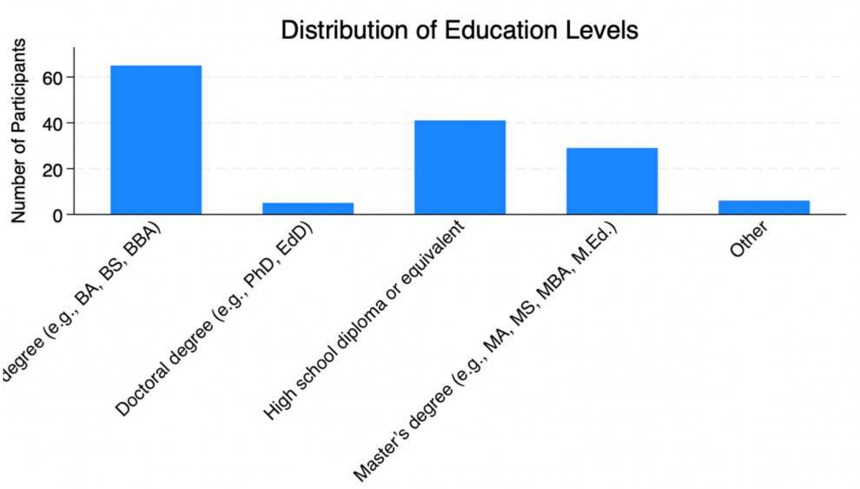
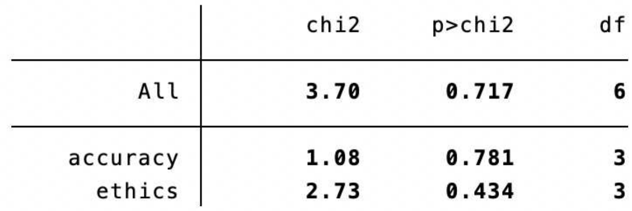


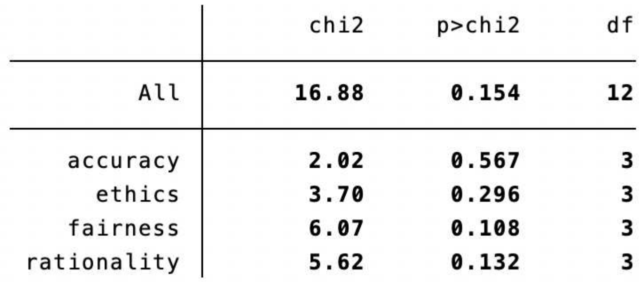
Figure 8: Education level composition of the sample

Appendix 3: Ordered regression assumption checks



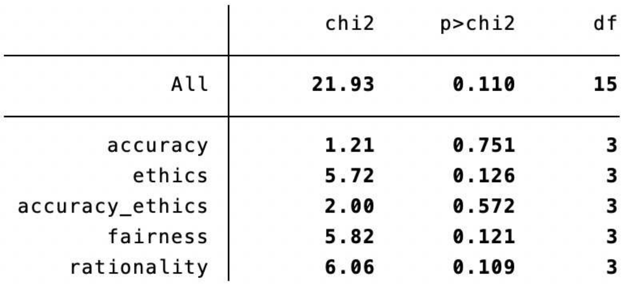
	chi2	p>chi2	df
All	**3.70**	**0.717**	**6**
accuracy	**1.08**	**0.781**	**3**
ethics	**2.73**	**0.434**	**3**

Table 9: Brant test for Model 1



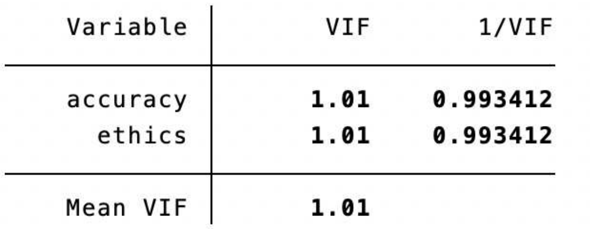
	chi2	p>chi2	df
All	**16.88**	**0.154**	**12**
accuracy	**2.02**	**0.567**	**3**
ethics	**3.70**	**0.296**	**3**
fairness	**6.07**	**0.108**	**3**
rationality	**5.62**	**0.132**	**3**

Table 10: Brant test for Model 2



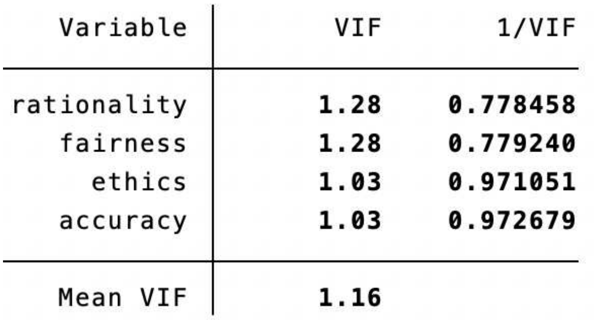
	chi2	p>chi2	df
All	**21.93**	**0.110**	**15**
accuracy	**1.21**	**0.751**	**3**
ethics	**5.72**	**0.126**	**3**
accuracy_ethics	**2.00**	**0.572**	**3**
fairness	**5.82**	**0.121**	**3**
rationality	**6.06**	**0.109**	**3**

Table 11: Brant test for Model 5



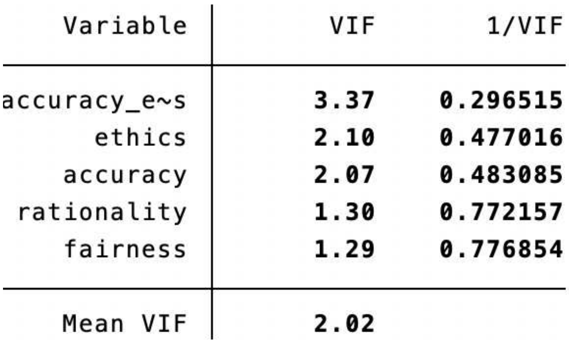
Variable	VIF	1/VIF
accuracy	**1.01**	**0.993412**
ethics	**1.01**	**0.993412**
Mean VIF	**1.01**	

Table 12: VIF test for multicollinearity for Model 1



Variable	VIF	1/VIF
rationality	1.28	0.778458
fairness	1.28	0.779240
ethics	1.03	0.971051
accuracy	1.03	0.972679
Mean VIF	1.16	

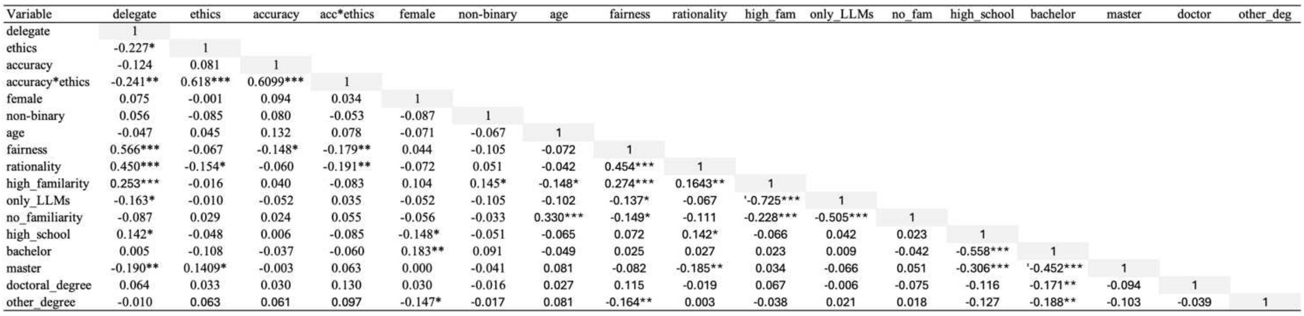
Table 13: VIF test for multicollinearity for Model 2



Variable	VIF	1/VIF
accuracy_e~s	3.37	0.296515
ethics	2.10	0.477016
accuracy	2.07	0.483085
rationality	1.30	0.772157
fairness	1.29	0.776854
Mean VIF	2.02	

Table 14: VIF test for multicollinearity for Model 5

Appendix 4: Correlation matrix



Variable	delegate	ethics	accuracy	acc*ethics	female	non-binary	age	fairness	rationality	high_fam	only_LLMs	no_fam	high_school	bachelor	master	doctor	other_deg
delegate	1																
ethics	-0.227*	1															
accuracy	-0.124	0.081	1														
accuracy*ethics	-0.241**	0.618***	0.6099***	1													
female	0.075	-0.001	0.094	0.034	1												
non-binary	0.056	-0.085	0.080	-0.053	-0.087	1											
age	-0.047	0.045	0.132	0.078	-0.071	-0.067	1										
fairness	0.566***	-0.067	-0.148*	-0.179**	0.044	-0.105	-0.072	1									
rationality	0.450***	-0.154*	-0.060	-0.191**	-0.072	0.051	-0.042	0.454***	1								
high_familarity	0.253***	-0.016	0.040	-0.083	0.104	0.145*	-0.148*	0.274***	0.1643**	1							
only_LLMs	-0.163*	-0.010	-0.052	0.035	-0.052	-0.105	-0.102	-0.137*	-0.067	'-0.725***	1						
no_familiarity	-0.087	0.029	0.024	0.055	-0.056	-0.033	0.330***	-0.149*	-0.111	-0.228***	-0.505***	1					
high_school	0.142*	-0.048	0.006	-0.085	-0.148*	-0.051	-0.065	0.072	0.142*	-0.066	0.042	0.023	1				
bachelor	0.005	-0.108	-0.037	-0.060	0.183**	0.091	-0.049	0.025	0.027	0.023	0.009	-0.042	-0.558***	1			
master	-0.190**	0.1409*	-0.003	0.063	0.000	-0.041	0.081	-0.082	-0.185**	0.034	-0.066	0.051	-0.306***	'-0.452***	1		
doctoral_degree	0.064	0.033	0.030	0.130	0.030	-0.016	0.027	0.115	-0.019	0.067	-0.006	-0.075	-0.116	-0.171**	-0.094	1	
other_degree	-0.010	0.063	0.061	0.097	-0.147*	-0.017	0.081	-0.164**	0.003	-0.038	0.021	0.018	-0.127	-0.188**	-0.103	-0.039	1

Figure 9: Full correlation matrix

Appendix 5: Survey

Welcome!

My name is Lennart Samland and I am a Master's student at ESCP Business School. I greatly appreciate your willingness to participate in this research survey, which will take less than 5 minutes to complete. The aim of this survey is to better understand individual decision-making. Please read the instructions carefully.
There is no risk to you in participating and you are free to quit at any time by closing your browser. There is absolutely no lying or deception in this survey. Your responses will remain anonymous, and it will not be possible to trace them back to an individual.

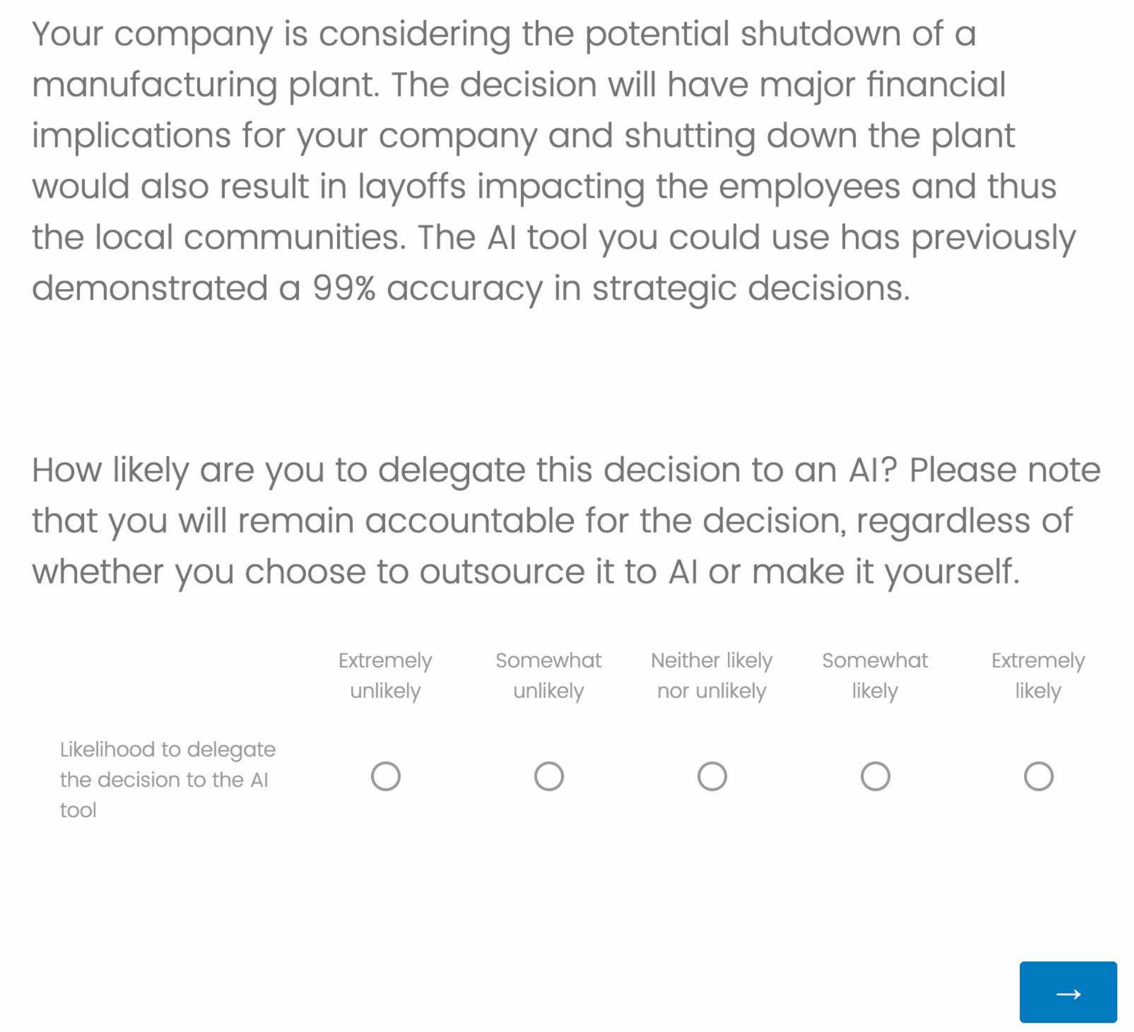
By submitting this form (clicking the "next" button) you are confirming that you are over the age of 18 and agree to participate in this study.

Feel free to contact me under lennart.samland@escp.edu.eu in case you have any questions, concerns or feedback.

Page 1: Standard thesis introductory text

Imagine you are a high-level executive at a multinational company making a critical strategic decision. Your company has recently gained access to an AI-based decision-making system that has been used for similar cases in other companies.

Page 2: Scenario description

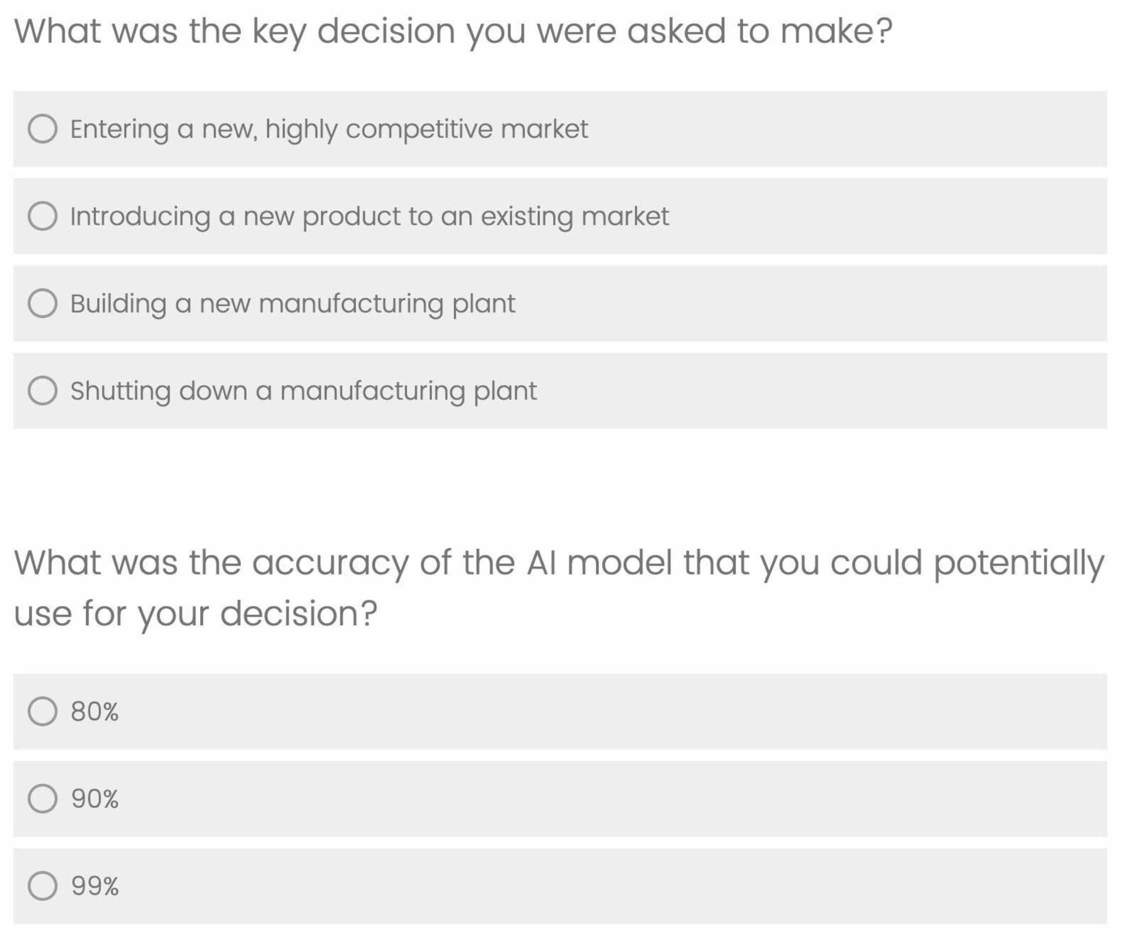


Your company is considering the potential shutdown of a manufacturing plant. The decision will have major financial implications for your company and shutting down the plant would also result in layoffs impacting the employees and thus the local communities. The AI tool you could use has previously demonstrated a 99% accuracy in strategic decisions.

How likely are you to delegate this decision to an AI? Please note that you will remain accountable for the decision, regardless of whether you choose to outsource it to AI or make it yourself.

	Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
Likelihood to delegate the decision to the AI tool	○	○	○	○	○

Page 3: Condition and assignment (e.g.: Condition 3)



What was the key decision you were asked to make?

○ Entering a new, highly competitive market

○ Introducing a new product to an existing market

○ Building a new manufacturing plant

○ Shutting down a manufacturing plant

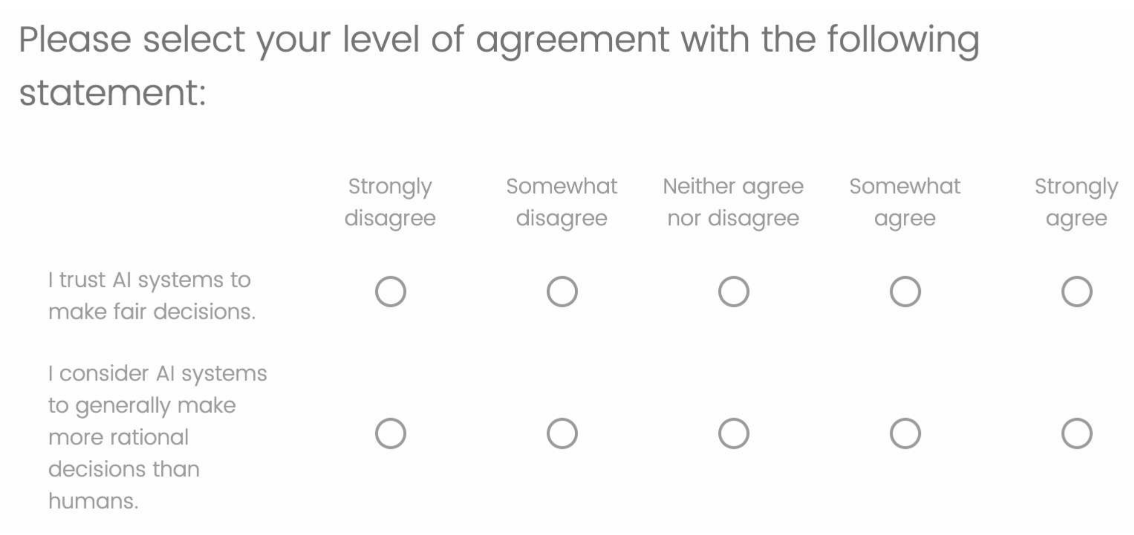
What was the accuracy of the AI model that you could potentially use for your decision?

○ 80%

○ 90%

○ 99%

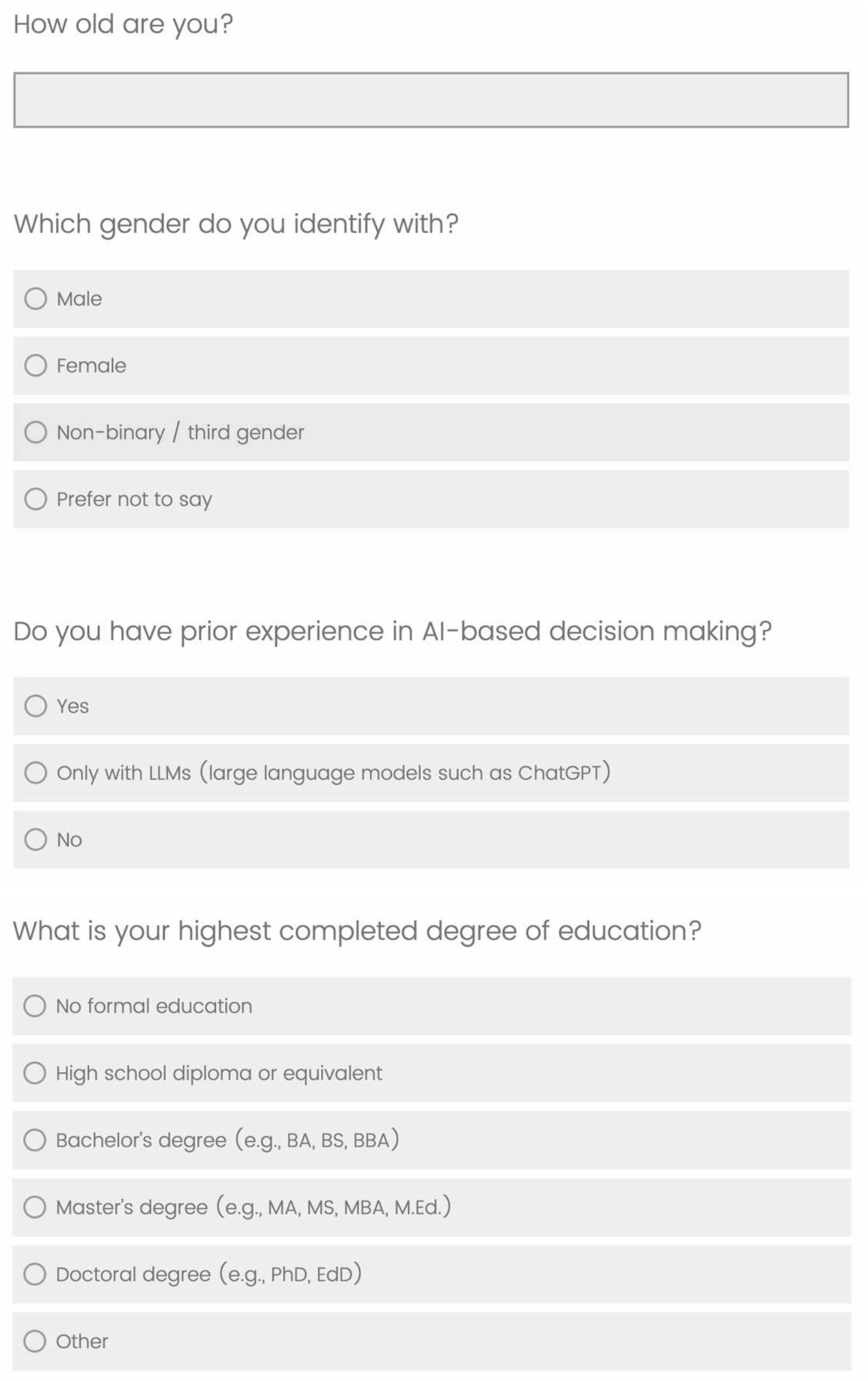
Page 4: Attention checks



Please select your level of agreement with the following statement:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I trust AI systems to make fair decisions.	○	○	○	○	○
I consider AI systems to generally make more rational decisions than humans.	○	○	○	○	○

Page 5: Post survey questions – measurement of perceived fairness and perceived rationality



How old are you?

Which gender do you identify with?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer not to say

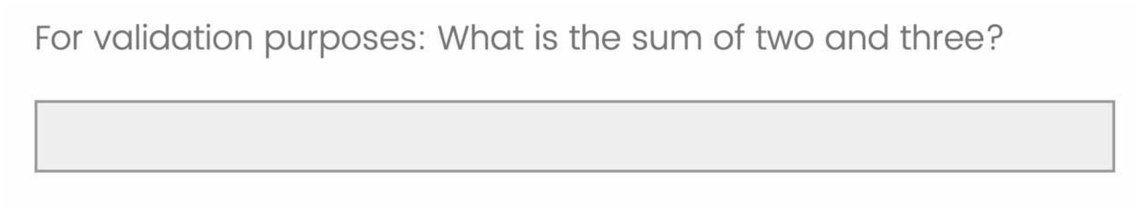
Do you have prior experience in AI-based decision making?

- ○ Yes
- ○ Only with LLMs (large language models such as ChatGPT)
- ○ No

What is your highest completed degree of education?

- ○ No formal education
- ○ High school diploma or equivalent
- ○ Bachelor's degree (e.g., BA, BS, BBA)
- ○ Master's degree (e.g., MA, MS, MBA, M.Ed.)
- ○ Doctoral degree (e.g., PhD, EdD)
- ○ Other

Page 6: Post survey questions – demographics and measurement of familiarity



For validation purposes: What is the sum of two and three?

Page 7: Validation check