In AI we trust? Perceptions about automated decision-making by artificial intelligence

Araujo, T.B.; Helberger, N.; Kruikemeier, S.; de Vreese, C.H.

Published in:
AI & Society

DOI:
10.1007/s00146-019-00931-w

In AI we trust? Perceptions about Automated Decision-Making by Artificial Intelligence

Theo Araujo[A]

Natali Helberger[B]

Sanne Kruikemeier[A]

Claes H. de Vreese[A]

[A] Amsterdam School of Communication Research (ASCoR), University of Amsterdam

[B] Institute for Information Law (IViR), University of Amsterdam

Corresponding author:

Theo Araujo, t.b.araujo@uva.nl Nieuwe Achtergracht 166, 1018 WV Amsterdam

Acknowledgements: This study was funded by the Research Priority Area Communication and its Digital Communication Methods Lab at the University of Amsterdam.

This is a pre-print ("before peer-review") of an article published in **AI & Society**. The final authenticated version is available online at https://doi.org/10.1007/s00146-019-00931-w

In AI we trust? Perceptions about Automated Decision-Making by Artificial Intelligence

Abstract

Fueled by ever growing amounts of (digital) data and advances in artificial intelligence, decision-making in contemporary societies is increasingly delegated to automated processes. Drawing from social science theories and from the emerging body of research about algorithmic perceptions and appreciation, the current study explores the extent to which personal characteristics can be linked to perceptions of automated decision-making by AI, and the boundary conditions of these perceptions, that is the extent to which such perceptions differ across media, (public) health, and judicial contexts. Data from a scenario-based survey experiment with a national sample (N = 958) show that people are by and large concerned about risks, and have mixed opinions about fairness and usefulness of automated decision-making by AI at a societal level, with general attitudes influenced by individual characteristics. Interestingly, however, decisions taken automatically by AI were often evaluated *on par* or even *better* than human experts for specific decisions. Theoretical and societal implications about these findings are discussed.

Keywords: Automated Decision-Making, Artificial Intelligence, Algorithmic Fairness, Algorithmic Appreciation, User perceptions.

This is a pre-print ("before peer-review") of an article published in **AI & Society**. The final authenticated version is available online at https://doi.org/10.1007/s00146-019-00931-w

In AI we trust? Perceptions about Automated Decision-Making by Artificial Intelligence

Fueled by ever growing amounts of (digital) data and advances in artificial intelligence (AI), decision-making is increasingly delegated to automated processes. These automated decision-making processes take place for example in communication, with algorithms making (personalized) news recommendations (e.g., Thurman and Schifferes 2012; Diakopoulos and Koliska 2017; Carlson 2018), personalizing advertising based on online behavior (Boerman et al. 2017), regulating user activity on social media platforms (van Dijck et al. 2018), automatically identifying suspicious profiles (Chu et al. 2012; Ferrara et al. 2016; Siddiqui et al. 2017), or even automatically generating news stories (Graefe et al. 2018). They also make their way into (public) health, with virtual health coaches recommending activities to individual users (Grolleman et al. 2006; Hudlicka 2013; Bickmore et al. 2016), or ongoing discussions on how to integrate AI to the decision-making process within healthcare (e.g., Agarwal et al. 2010; Dilsizian and Siegel 2013; Jha and Topol 2016; Yu and Kohane 2018). Their relevance is also growing in the judicial and law enforcement sector (for an overview of the possibilities, see Nissan 2017). Examples, primarily in the United States, show algorithms already being used to evaluate who might be eligible to early release from jail (Dressel and Farid 2018) and for criminal sentencing (Angwin et al. 2016). A wider discussion also takes place on how algorithms may help with pre-emptive policing tasks (Kennedy et al. 2011; Stanford University 2016) as well as the potential for "algorithmic states of exception" (McQuillan 2015).

Automated decision-making (ADM) can be defined in several ways. Narrowly, it can be described as "decisions by technological means without human involvement" (European Commission 2018, p. 7). More broadly, however, it may be seen as the process through which the ever-growing amount – and variety – of personal data collected "are subsequently processed by algorithms, which are then used to make (data-driven) decisions" (Newell and Marabelli 2015, p. 5). This leads to completely automated decision-making processes or can serve as aid for human decision-makers across a wide variety of contexts and varying levels of autonomy.

A growing body of research begins to investigate the consequences of ADM and AI-driven decisions in contemporary societies, as well as its limitations and risks, with increasing discussions in academic and technical circles highlighting that automated decisions can be biased (Hajian et al. 2016; Zarsky 2016; Angwin et al. 2016). Surveys also indicate that general concerns about fairness and usefulness of algorithms in decision-making, with a majority of Americans considering the usage of algorithms for decisions that may impact humans as unacceptable

(Smith 2018). However, much is yet to be explored about what influences people's perceptions of ADM by AI especially when it comes to its perceived usefulness and expectations regarding fairness (Lee and Baykal 2017; Lee 2018) or risk. Exploring what influences people's perceptions about ADM is especially important because ADM, and algorithms in general, can be seen as socio-technical concepts (Kitchin 2017; Elish and boyd 2018) that do not function in isolation but are embedded in the context of particular societal, institutional or organizational structures, with their own mechanisms, incentives, (power) relationships and roles in society. The importance of understanding the drivers for ADM perceptions becomes even more pressing considering that people's perceptions of what algorithms are capable of - instead of merely their performance - play a critical part in how they evaluate and ultimately accept ADM (Lee 2018).

Drawing from social science theories (Sundar and Nass 2000, 2001; Sundar 2008) and the emerging body of research about algorithmic perceptions (Logg et al. 2018), the current study makes several contributions to our understanding about attitudes towards ADM by AI. First, it extends the notion of algorithmic appreciation (Logg et al. 2018; Thurman et al. 2018) by explicitly investigating perceptions not only of *usefulness*, but also of *risk* and of *fairness* towards this technology. Second, using a survey with a representative sample from the Netherlands (N = 958), this study examines how individual characteristics influence general attitudes towards algorithms, AI, and ADM. As such, it extends earlier findings about algorithmic perceptions in other countries such as the US (Smith 2018) and provides a deeper level of insight of what *influences* such perceptions. Third, this study investigates not only automated *recommendations,* but also scenarios with *decisions* by AI and algorithms, thus going one step further in the level of autonomy provided to the algorithm vis-à-vis the subject of the decision. We also distinguish between different levels of impact of the decision and explore a range of scenarios in which the impact may be as low as creating a list of recommended news articles or fitness advice, to as high as sentencing in the judicial sector or treatment decisions in health. While doing so, we explicitly compare perceptions of the same decisions being taken by a human expert or by AI, thus providing a critical level of nuance of overall perceptions of algorithmic decision-making by contrasting it to baseline perceptions of decisions taken by humans. Finally, by taking a holistic approach and exploring ADM by AI in general, as well as perceptions about specific scenarios, the results of this study help paint a more complete picture about the perceptions and concerns about ADM by AI across societal sectors. While there are many potential sectors that could be explored, media, (public) health, and justice were chosen. In these three sectors, ADM can have a significant impact on individual rights, well-being and functioning in a society (as citizens and voters in the case of the media, as members of a society in the case of justice and as humans in the case of health). Taken together, the following research question will be answered: Which personal characteristics can be linked to perceptions of Automated Decision Making by Artificial Intelligence and what are the boundary conditions of these perceptions (i.e., do perceptions differ in different domains)?

Automated decision-making

The notion of ADM can be conceptualized as instances in which algorithms or AI collects, processes, models, and uses data to make automated decisions. In turn, the feedback from these decisions is then used by the system to improve itself for future use. This mechanism, however, should be seen as a largely socio-technical concept that goes beyond the technical aspect of what an algorithm or AI might be (Kitchin 2017; Elish and boyd 2018). While an algorithm may be considered as a set of “encoded procedures for transforming input data into a desired output, based on specified calculations” (Gillespie 2014, p. 1), how we conceptualize it further evolves through time, societal, and institutional contexts and human-machine interaction. Algorithms, therefore, “can be thought about in a number of ways: technically, computationally, mathematically, politically, culturally, economically, contextually, materially, philosophically, ethically and so on” (Kitchin 2017, p. 17). This study will focus on automation in the sense of an "the ongoing production of a process without the mediation of a person" (Dodge and Kitchin 2007, p. 270), that takes the role of an automated decision-maker.

Automated decision-makers can be conceptualized as an algorithm, a recommender system or, more in general, as being simply “artificial intelligence”, depending on how it is framed and presented to the user of the system or subject of the decision. ADM can take multiple forms. It ranges from decision-support systems that make recommendations to human decision-makers and/or nudge users of these systems in a certain direction, to fully automated decision-making processes that take decisions on behalf of institutions or organizations without human involvement. In this sense, human decision-makers rely to varying degrees on automated decision support systems when taking decisions that either relate to themselves (e.g., health app that coaches healthy behavior) or to third parties (e.g., judge using an ADM to determine a fine). The level of involvement of humans into these automated decisions varies. On the one hand, recommender systems, when deciding what to recommend to a user (e.g., a health advice or a newspaper article), still might leave a substantial level of autonomy to this user in terms of choosing whether to accept, or not, these recommendations.

What is more, such data-driven decision support systems bring also the implicit expectation that future actions by the system will be influenced by the behavior of the user of the system (or of the decision-maker) by means of an algorithmic feedback loop. Fully automated decision-making processes, when incorporated in “managerial and governance processes” (Lee 2018, p. 2), on the other hand, often only communicate the results of a decision without any room for human involvement in the making of the decision itself. In the worst case, such a system will leave the recipient of the decision not only in the dark about to what data were used to take the decision, or even how to contest such an outcome, but also even in terms of choice as to whether to participate in the process in the first place.

These technologies and methods, however, go through “fear and hype cycles” - as what has been arguably taking place with the growing importance of big data, machine learning, and AI (Elish and boyd 2018). This makes them particularly interesting for social scientific research. What influences attitudes and perceptions about ADM is not just the technological solution they offer, or their actual performance, but also the way in which these ADM processes are framed or communicated to the user or subject of the decision (Lee 2018). Most importantly is the assumption of neutrality and objectivity of these systems. They are “created for purposes that are often far from neutral: to create value and capital; to nudge behavior and structure preferences in a certain way; and to identify, sort and classify people” (Kitchin 2017, p. 19). However, these systems are often carefully and strategically articulated as impartial and objective socio-technical actors in the discourse surrounding their implementation and usage in different aspects of daily life, as happens for example with search engines or other recommender systems (Gillespie 2014). This may lead to expectations of objectivity and fairness of a machine that may very well exceed the level of objectivity and fairness that human decision makers are capable of.

Attitudes towards ADM

Earlier studies show a general tendency of considering an “expert system to be more objective and rational than a human adviser” (Dijkstra et al. 1998, p. 160). This tendency, often based on the assumption of statistical methods outperform human judgement (Dawes et al. 1989), gives rise to the idea of *algorithmic appreciation*, showing in several instances that people prefer judgement or recommendations by algorithms when compared to human recommendations (Logg et al. 2018). This tendency may be partially attributed to the notion of the *machine heuristic* (Sundar 2008), which suggests that the less a user anthropomorphizes an interface, the more she or he will consider its decisions and selections to be objective and free of (ideological) biases. In addition, research based on the Computer Are Social Actors (CASA) paradigm (Nass et al. 1994), and in particular on source orientation suggests that that computers - or, in the case of this study, ADM by AI - are treated as autonomous sources (e.g., Sundar and Nass 2000), with the assumptions or rules decided by the programmer when creating or managing the system disappear from view, or not being as prominent in the user perceptions.

Several studies, however, contextualize this general tendency, suggesting it might be dependent on several factors. For example, compared to human decision-makers, ADM or recommendation systems can be seen as inscrutable, which might impact the user's willingness to accept the system, or its recommendations (Yeomans et al. 2016). People seem also to be far less forgiving towards ADM than to humans: Recognizing that an algorithm makes a mistake - even when its overall performance is better than of a human - was seen to be sufficient to make people choose the human decision-maker, and thus leading to *algorithmic aversion* (Dietvorst et al. 2015). Moreover, the type of (human) decision maker serving as the reference is also important: machines were trusted more than human non-experts, but less trusted than human experts (Madhavan and Wiegmann 2007), for objective or subjective decisions (Logg 2017) or for management decisions requiring human or mechanical skills (Lee 2018). Interestingly, participants in the latter study associated authority as a source of fairness in human decision-makers, whereas, also in line with earlier research, algorithms were seen as “unbiased and efficient decision-maker” (Lee 2018, p. 8). The current study contributes to this line of research by investigating how perceptions of fairness, usefulness, and risk associated with ADM by AI are influenced both by contextual and individual factors, as outlined in the following sections.

Personal characteristics and attitudes towards ADM

When delving deeper into how users or subjects of the decision perceive fairness, usefulness or risk associated with ADM by AI, the first question that arises is to what extent do individual characteristics play a role in the process? We briefly review some of the key findings from earlier research below.

Knowledge

Earlier research highlights the role of knowledge in how people perceive ADM and algorithmic recommendations. The evidence, however, is mixed. On the one hand, comfort with mathematics (Logg 2017), or level of education

(Thurman et al. 2018) were shown to have, in general, positive associations with better attitudes towards algorithmic recommendations. More specifically, when it comes to news personalization, lower levels of education showed a stronger negative effect for attitudes towards automated personalization based on user behavior compared to other forms of news selection such as journalistic curation (Thurman et al. 2018). However, when it comes to fairness in algorithmic decisions involving groups, the higher the level of computer programming knowledge, the less that the algorithmic-mediated decision was perceived to be fair (Lee and Baykal 2017). Given these mixed findings, we propose the following research question, making a distinction between general (i.e., educational attainment) and domain-specific (i.e., knowledge about AI, algorithms and computer programming) knowledge:

RQ1. To what extent does general and domain-specific knowledge influence perceptions about usefulness, fairness and risk of ADM by AI?

Online privacy concerns and self-efficacy

ADM is usually associated with data-driven decisions (Newell and Marabelli 2015), based, for the most part, on automated and large-scale collection of digital trace data about individual behavior. We expect that general concerns about how personal data are collected and use (privacy concerns) - as well as one's assumptions about their own level of ability to protect her or his data (online self-efficacy) - will also influence general perceptions about ADM by AI. This is supported, for example, by earlier research indicating higher levels of privacy concern are associated with worse attitudes towards automated personalization of news based on user behavior (Thurman et al. 2018). We, therefore, assume that users with higher levels of privacy concerns will have a more critical evaluation of ADM by AI, whereas the more certain a user is of her or his own online self-efficacy in protecting their own privacy, the more confident - and therefore more positive - this user will be about these systems. This leads to the following hypotheses:

H1. Online privacy concern levels are negatively associated with perceptions of fairness and usefulness of ADM by AI, and positively associated with perceived risk.

H2. Online self-efficacy is positively associated with perceptions of fairness and usefulness of ADM by AI, and negatively associated with perceived risk.

Demographics

Beyond general and domain-specific knowledge, and online privacy concerns and self-efficacy, the role of demographics - especially age and gender - may also be relevant when it comes to perceptions about ADM by AI although the evidence is mixed. Age, for example, has been shown to partially influence perceptions about algorithms: Older people tended to show higher levels of agreement that human editors are a better way to receive news than through algorithms (Thurman et al. 2018) and be less supportive than younger Americans to the notion that algorithmic decisions might be free from biases (Smith 2018). The same results, however, were not seen in other contexts (Logg 2017). Gender, however, was not seen as relevant for algorithmic appreciation (Logg 2017; Thurman et al. 2018), yet it was seen to partially influence perception or attitudes towards algorithmic fairness (Dineen et al. 2004; Pierson 2017). Moreover, considering differences in scope - from recommendation to automated decision-making systems - and in types of attitudes - from appreciation to perceptions of usefulness, fairness and risk - between this study and earlier research, we believe it is important to update our knowledge about the influence of demographic characteristics on ADM by AI perceptions, and therefore propose the following research question:

RQ2. To what extent do demographic characteristics - i.e., age and gender - influence perceptions about usefulness, fairness, and risk of ADM by AI?

Belief in Equality

As per the *machine heuristic* (Sundar 2008), automated recommendation or decision-making systems might be perceived as being unbiased and objective, applying always the same set of (objective) rules especially if its users consider that it is a machine that takes the decisions. These perceptions may be consistent with the socio-technical discourse (by technology companies, among others) that tries to bring forward an image of lack of bias, objectivity, neutrality, consistency and predictability of outcomes of algorithmic processes (Gillespie 2014) despite increasing evidence of bias and risks brought by these technologies (e.g., O'Neil 2017). It stands to reason, we argue, that personal beliefs about the need for equality might also influence attitudes towards ADM by AI. We therefore propose the following research question:

RQ3. To what extent does the belief in the need for equality influence perceptions about usefulness, fairness and risk of ADM by AI?

Attitudes towards AI decision-makers: differences across contexts

As indicated earlier, research shows that the type of task, or decision, has an influence on attitudes towards or reliance on algorithmic recommendations (Logg 2017; Lee 2018). Given the pervasiveness of ADM across an increasing set of societal settings, we investigate how perceptions of usefulness, fairness and risk associated with AI, when compared to expert human decision-makers, vary across contexts. To provide a more complete picture, we explore this taking in consideration both the impact of the decision - comparing scenarios with low and high-stake consequences - and the level of *personalization* of the consequence - i.e., whether the subject of the decision is the person evaluating the ADM system, or rather someone else. We therefore propose the following research question:

RQ4. To what extent does ADM by AI usefulness, fairness and risk perceptions vary across the Media, Justice and Health sectors?

Methods

Sample

Participants to this study were recruited among members of a database from a public opinion research company Kantar Public, as part of a larger project investigating public perceptions of automated decision-making by AI. This database contains more than 115,000 participants, being representative of the Netherlands. A random sample of 3,072 panel participants of this database were invited. This sample reflects the Dutch population when it comes to gender, age, region and educational levels, with random sampling applied to each stratum. From the 1,069 who accepted the invitation, 958 provided informed consent and completed the questionnaire. The final sample (N = 958) was composed of 49% females, had an average age of 50.9 years (*SD* = 16.7), and with 34% of the respondents having completed up until the secondary level of education, 27% having a bachelor's degree (or equivalent) and 15% having a master's degree (or higher).

Procedure

Participants in the study first provided their informed consent, and answered a set of initial questions, including their level of knowledge about algorithms, AI, and computer programming. They then read a definition for ADM by AI which stated that "automated decision-making by artificial intelligence or computers can be defined as computer programs that can take decisions that were previously taken by humans. These decisions are taken automatically by computers based on data". Participants were then requested to answer a set of questions about their perceptions about ADM by AI. These answers are used to explore the influence of individual characteristics.

After answering these initial questions, participants were randomly assigned to scenarios that described decision-making across the Media, Health, and Justice sectors, and provided evaluations to the potential usefulness, fairness and risk of these decisions. These scenarios were part of a vignette-experiment with 2 (decision-maker: Human vs. AI) x 2 (subject of the decision: self vs. others) x 2 (impact of the decision: high vs. low) between-subjects design with the context - Media, Health, and Justice - being the within-subjects design. The contexts were presented in a randomized order to participants. Table 1 provides an overview of the scenarios. For all scenarios, respondents were told that the decisions were based on the data about (the behavior) of the subject of the decision. The subjects of the decision were either the participant her or himself ("you") or "someone else", depending on the condition. These results were used to answer RQ4.

<<< TABLE 1 ABOUT HERE >>>

Measures

Independent variables

Online self-efficacy (M = 3.3, SD = 1.21, α = .70) was measured with three questions from previous research (Boerman et al. 2018). *Privacy concerns* (M = 5.07, SD = 1.15, α = .83) was measured with five questions (based on (Baek and Morimoto 2012; Bol et al. 2018). *Technical knowledge* (M = 2.41, SD = 1.33, α = .88), partially in line with earlier studies on algorithmic fairness (Lee and Baykal 2017) was measured with three questions about self-reported knowledge about AI, algorithms, and computer programming. The notion of *belief in equality* (M = 3.59, SD = 1.70) was operationalized by a measure of belief in need for income equality with the same wording as in the World Values Survey (Inglehart et al. 2005). All variables were measured in seven-point scales. Gender, age, and education levels (1 - no or basic education, 7 - master/doctorate) were also included.

Dependent variables

The notions of usefulness, fairness, and risk were included as dependent variables in the analyses. For RQ1, participants indicated the extent to which they agreed that automated decision-making by AI could be considered useful, fair and risk for society as a whole. *Usefulness* was measured with three items adapted from earlier research on technology adoption (Davis 1989; Nysveen 2005). Given the intermediate levels of reliability, analyses were executed with (*M* = 4.21, *SD* =.92, α = .54) and without (*M* = 4.01, *SD* = 1.13, α = .75) the one item that was reversed for this study[1]. *Risk* (*M* = 4.51, *SD* = 1.01, α = .81) was measured with five items adapted from Cox and Cox (2001). For *fairness* (*M* = 3.87, *SD* = 1.34), we have used a single item in line with earlier research (e.g.,(Lee 2018).

For RQ4, given the number of scenarios that respondents had to evaluate, single-items were used for *fairness* (*M* = 3.86, *SD* = 1.59), *usefulness* (*M* = 3.72, *SD* = 1.62) and *risk* (*M* = 3.94, *SD* = 1.53). All responses were measured using 7-point scales. Participants answered a manipulation check question, which asked how large the impact of the decision would be. The results of a multilevel linear model with the respondent placed as the contextual level indicate that the manipulations worked as intended. Contrasts with Bonferroni adjustments showed that the high impact scenario was perceived as having a larger impact than the low impact scenario for justice (Contrast = 1.27, SE = 0.11, $p < 0.001$), health (Contrast = 1.62, SE = 0.11, $p < 0.001$) and media (Contrast = 0.69, SE = 0.11, $p < 0.001$). Participants also answered an attention check as to who was the decision-maker in the scenario (human or AI/computer). While the human and AI scenarios differed significantly in the answers in the correct direction, we adopted a strict approach and removed the responses not aligned with the decision-maker communicated in the scenario from the subsequent analyses[2].

Results

General perceptions about automated decision making by AI

Respondents first evaluated, in general, how they perceived ADM by AI to be potentially *useful*, *fair,* and *risky* when considering society as a whole. For the perceived usefulness scale, a slightly optimistic picture emerged, with approximately 40% of the respondents scored above the midpoint of the scales in which the items were measured (4) while approximately 35% scored below the midpoint. For perceived fairness, respondents were more split, with 29% scoring above the midpoint, and 32% below. For perceived risk, however, a large majority of the respondents (66%)

[1] Results are reported for the measure with higher reliability (without the reversed item), but differences are communicated in the notes (Table 2).

[2] Approximately 28% (818) of the responses for all the scenarios combined (N = 2874) were removed because of the manipulation check. When running the analyses with these responses included, the results stay largely the same with regards to direction and significance levels. Exceptions are indicated in the notes.

were negative about ADM by AI, scoring above the midpoint when evaluating its potential risk, with only 22% being somewhat optimistic, scoring below the midpoint for perceived risk.

The influence of individual characteristics on general perceptions of usefulness, fairness and risk associated with ADM by AI was tested with OLS regression in three separate models (Table 2). When it comes to *general and domain-specific knowledge* (RQ1), the results show that general knowledge (education) has a positive association with beliefs in usefulness of ADM by AI, while domain-specific knowledge (knowledge about computer programming, AI and algorithms) has a positive association with perceptions of usefulness and of fairness of ADM by AI. There was no significant association between (general or domain-specific) knowledge and perceptions of risk.

The results also provide support to H1 and H2. More specifically, *online privacy concerns* were negatively associated with perceptions of usefulness and of fairness of ADM by AI, and positively associated with perceptions of risk, supporting H1: The higher the levels of privacy concern about online activity, the more negative the attitudes about ADM by AI. *Online self-efficacy* had an opposite effect, supporting H2: The stronger the person's belief in their ability to protect their privacy online, the more positive their perceptions about usefulness and fairness of ADM by AI, and the lower the perceived risk.

In response to RQ2, *age* was shown to have a negative association with the perceived usefulness of ADM by AI, and a positive association with risk. *Gender* was only relevant for perceptions of usefulness, with females seeing ADM by AI as significantly less useful than males, and marginally associated with perceptions of risk. Finally, in response to RQ3: higher levels of belief in need of (economic) equality were associated with higher levels of usefulness and fairness perceptions about ADM for AI, whereas there was no significant association with perceptions of risk.

<<< TABLE 2 ABOUT HERE >>>

Attitudes towards ADM by AI across contexts

To investigate differences in fairness, usefulness, and risk perceptions across contexts depending on the level of impact and the subject of the decision (RQ4), a series of models were executed with the evaluation of the decision (fairness, usefulness, risk) as the dependent variable. First, we evaluated the differences across the context using multilevel models with the participant set as the contextual level. Multilevel linear analyses are used because the data are hierarchical, meaning that the evaluations of the decision (i.e., the dependent variable) are nested within individuals (i.e., because the respondents evaluated three scenarios). In other words, the evaluations thus depends on the person (see (Field 2013) for a detailed explanation). As a result, evaluations are not independent observations. Thus, to control for this dependency in the data, we estimated multilevel models when comparing responses across contexts. Second, when investigating the boundary conditions *within* a context, we used regression models – as each participant only evaluated one scenario per context. In both cases, the decision-maker (human expert or AI), impact (high vs. low), subject of the decision (self vs. other) and the context (health vs. media vs. justice) were set as the independent variables of the models. After running the models, we evaluated the differences between human and AI decision-makers by estimating the marginal effects of AI vs. human expert decision makers across high and low impact scenarios, and reporting contrasts with Bonferroni adjustments.

<<< TABLES 3 AND 4 ABOUT HERE >>>

Fairness

When analyzing each context with all scenarios combined (Table 4), there are no significant differences in perceived fairness between AI and human experts across Justice, Health and Media. When investigating the boundary conditions of fairness perceptions, however, ADM by AI was perceived as fairer than human experts with significantly higher levels for Justice and for Health in high-impact decisions, as revealed by the contrasts with Bonferroni adjustments (Table 3). No significant differences were seen for low-impact scenarios and for Media.

Risk

The analysis of the combined scenarios (Table 4) revealed no significant differences in perceived risk in decisions taken by AI when compared to human experts across the three contexts. However, in the case of Media, the differences in perceived risk between AI and human experts were marginally significant, with the contrasts with Bonferroni adjustments showing that AI decisions are perceived as less risky than those by human experts. When analyzing the boundary conditions within each context (Table 3), the contrasts revealed significant differences in the perceptions of risk of decisions taken by ADM by AI versus human experts: in all high-impact scenarios, AI decisions were perceived to be less risky. No differences were seen for low-impact scenarios.

Usefulness

Finally, when it comes to usefulness, the combined scenario analysis (Table 4) showed that in the Health context, decisions by ADM by AI were seen to be more useful than those taken by humans, whereas the same differences were not seen in the contrasts with Bonferroni adjustments for Justice or for Media. When investigating the boundary conditions within each context (Table 3), decisions by ADM by AI was perceived as being more useful than those by human experts in high-impact scenarios for Justice and Health. ADM by AI was also seen as more useful in low-impact decisions in Health, although the differences with human decision-makers were just marginally significant. The contrasts did not reveal differences for Media.

Discussion

The present study, drawing from social science research and the emerging body of research on algorithmic perceptions, explored the extent to which individual and contextual characteristics influence attitudes towards ADM by AI. The results of a survey and a scenario-based experiment with a representative sample of the Dutch population show that, when thinking in general about ADM by AI as a societal development, respondents are split when it comes to the potential usefulness or fairness of automated decision-making processes but are concerned about the potential risks. However, when contrasting respondents' perceptions of fairness, usefulness, and risk for the specific decisions within media, (public) health and justice, ADM by AI was for the most part seen as *on par*, and at times *better* evaluated than human experts. These results make several important contributions to our understanding of how users perceive automated decision-making processes, validating and extending earlier research on algorithmic appreciation in general (e.g., Logg et al. 2018) and with regards to algorithmic recommendations within the media sector (e.g., Thurman et al. 2018).

When considering ADM by AI in general, individual characteristics partially explain these views. Firstly, the results show that knowledge - both domain-specific (operationalized as knowledge about AI, algorithms and computer programming) and general (operationalized as levels of education) - is associated with increased expectations about *usefulness* of automated decision-making processes. This is in line with earlier research on automated news recommendations (Thurman et al. 2018). When it comes to fairness or risk perceptions, however, knowledge did not show any significant associations with increased risk (general and domain-specific knowledge), or lower fairness perceptions among people with higher levels of domain-specific knowledge unlike earlier research on algorithmic fairness (e.g., Lee and Baykal 2017). This suggests that people with more knowledge are actually *more* optimistic about ADM by AI when it comes to its usefulness, whereas knowledge seems to impact less perceptions of fairness or of risk.

Secondly, *online self-efficacy* was associated with higher expectations of fairness and of usefulness of ADM by AI, and with lower levels of perceived risk. This suggests that the more a person believes that she can protect her own privacy online -, i.e., higher online self-efficacy -, the more confident she is about the usefulness and fairness of ADM. This is striking, as particularly in situations in which ADM operates autonomously and without the possibility for users to exercise agency, however, this confidence can easily turn into a sense of false security as people may feel more confident but in fact do lack the means to exercise meaningful control. Along the same lines, *online privacy concerns* were associated with negative attitudes towards ADM. This supports earlier findings regarding attitudes towards ADM in news recommendations (Thurman et al. 2018), also extending these findings when it comes to perceptions of *fairness* and of *risk*. It is important to note that this influence of online self-efficacy and online privacy concerns were seen for *general* attitudes towards ADM by AI at a societal level. As more data could potentially translate in the assumption of more accurate decisions, future research should extend these findings and investigate the extent to which these characteristics influence the *willingness* to use an ADM – or *acceptability* to become a subject of its decision.

Thirdly, the results show not only age and gender differences when it comes to perceptions about usefulness and risk of ADM by AI, but also that actually people with higher levels of belief in equality - in our study operationalized as belief in economic equality - are actually more optimistic about ADM by AI when it comes to perceptions of fairness and of usefulness. These results further reinforce the notion that, despite warnings in the media and in the academic debate, a somewhat optimistic view emerges about ADM by AI precisely among those that would be the most critical or concerned about biases in decision-making.

While these findings provide an overview of how individual characteristics influence general attitudes towards ADM by AI, this study went one step further and compared *specific perceptions* to common automated decision-making scenarios across important societal sectors in which these decisions are increasingly prevalent: media, (public) health and justice. It is interesting that when comparing perceptions about ADM by AI and human experts as decision makers in specific scenarios, a different somewhat more positive view of ADM by AI emerged than when exploring *general* attitudes towards ADM by AI for society as a whole. For higher impact decisions, decisions by human experts were, in general, associated with more negative attitudes than when the same decisions were taken by ADM by AI. For justice and health human experts scored as *less fair* than ADM by AI - while for media there were no significant differences. With regards to usefulness, in all higher impact scenarios - with the exception of media - ADM by AI emerged with higher scores than human experts. And for risk decisions by ADM by AI were perceived as having lower levels of risk than when the same decisions were taken by human experts. For lower impact decisions, there were almost no significant differences in how human experts and ADM by AI as decision-makers were evaluated on fairness, usefulness or risk. The only exception to this trend was for justice in which a human expert decision-maker received higher scores for usefulness than AI, although that difference was only marginally significant.

The findings that emerged for the specific scenarios of decision-making were partially aligned with the notion of the *machine heuristic* (Sundar and Nass 2000, 2001; Sundar 2008) and of *algorithmic appreciation* (Logg et al. 2018), and provide a new level of nuance to the theory of algorithmic decision-making and the notion of the machine heuristic in several ways. Firstly, they show that in all scenarios ADM by AI was evaluated *on par* or *better* than decisions taken by human experts, thus reinforcing that the machine heuristic does play a role in attitudes towards ADM by AI. Secondly, we complement earlier research that has shown differences in attitude towards algorithmic decisions depending on whether the decisions were considered objective or subjective (Logg 2017), or requiring human or mechanical skills (Lee 2018) by revealing differences in attitudes depending on the level of impact of the decision - and in some sense on the level of autonomy given to the subject of the decision vis-à-vis the (human expert or AI) decision maker. Thirdly, unlike earlier research that has shown that human experts were more trusted than AI when it comes to recommendations (Madhavan and Wiegmann 2007), our findings either do not show differences, or show even better attitudes towards ADM by AI. This may be due to our emphasis on *automated decisions* instead of mere *recommendations* and with the evaluation being done from the perspective of the subject of the decision, instead of a person that will receive the input a recommendation to take the decision by her or himself. This may also be due to shifts in societal attitudes about ADM by AI and should be investigated by future research. Another reason could be a lack of trust in human decision making more generally, with ADM being perceived more as an alternative to humans, rather than per se fairer decision makers. Finally, our results put attitudes towards ADM by AI within media in context, by contrasting with scenarios in (public) health and in the judicial sector - showing similar trends when it comes to usefulness and risk in higher impact decisions, while in general less difference in attitudes towards human experts and AI as decision makers.

While these findings make important contributions to our understanding of algorithmic appreciation in contemporary societies in general, some limitations need to be acknowledged. Firstly, when it comes to the scenarios, the comparison between human expert and ADM by AI as decision-makers were done with *between group* comparisons. This presents advantages for getting in general unbiased evaluations of each decision-maker separately, but further research should also evaluate *preference* by explicitly asking respondents to score one type of decision-maker against the other, and the extent to which respondents would be *willing* to or *accept* becoming subjects of ADM by AI. Secondly, the decisions were presented as scenarios, with the participants being asked to imagine that they, or other people, were in that specific situation, with the same questions (fairness, usefulness and risk) asked across all scenarios and decisions for consistency and comparability. Future research should extend and complement these findings by exploring the perceptions of people who were *actually* subject of these decisions and use research designs that are even more realistic when it comes to the manipulation as well as context-specific measures. Finally, the participants were from the Netherlands, which has extremely high internet penetration rates and technology adoption, and is part of the European Union - which, especially given the attention provided to the General Data Protection Regulation - may influence perceptions and expectations about privacy and individual data protection. Future research should therefore extend our findings with comparative studies in other countries with different privacy expectations and regulations, such as the US, China or Brazil.

Conclusion

By means of a survey-experiment with a high-quality national sample the present study brings mixed views about perceptions of risk, fairness and usefulness of automated-decision making by AI. When evaluating the societal consequences of ADM by AI in general, the attitudes that emerged point to concerns about risk, and mixed opinions about its fairness and usefulness. However, when respondents had to evaluate the potential fairness, usefulness and risk of specific decisions taken automatically by AI in comparison to human experts, ADM by AI was often evaluated *on par* or even *better* for high-impact decisions. Notably domain-specific knowledge, as well as belief in equality and online self-efficacy were associated with more positive general attitudes about the usefulness, the fairness and the risk of decisions taken by AI, whereas increased levels of privacy concerns had a negative association.

In this sense, privacy is a pivotal aspect as is human agency. The research showed that people who felt more in control (online efficacy) were more likely to consider ADM as fair and useful, but for this feeling of being in control to not become a fallacy, finding ways of increasing true control and the ability of the receivers of the decision to exercise agency is an important aspect. Agency, i.e. the ability to take an active part in ADM means more than getting an explanation, or being able to demand another human to reconsider the decision one the decision has been made. Agency, in the sense of true self-efficacy would start already earlier in the process by being for example able to choose for which decisions ADM is preferable, or review the accuracy and completeness of the information that goes into that processes or being able to bring in additional or alternative argumentation

Overall, these findings are somewhat contrasted by the rather critical and pessimistic tone that is often prevalent in media reporting but also the academic literature, highlighting fears over bias, loss in human dignity and autonomy, and more generally concerns about 'AI taking over' and replacing human decision makers. The research findings seem to suggest that the Dutch population is at least not blind to the potential benefits of ADM, in terms of usefulness and fairness, even though they do see risks. Interestingly, the findings do seem to suggest that humans as decision makers are not per se perceived as being irreplaceable – at least when comparing perceptions about decisions taken by humans and by ADM in specific scenarios. And yet, caution is in place before interpreting these findings as an encouragement for current government initiatives in the Netherlands, but also elsewhere in the EU to explore the potential of ADM in various sectors of society. Public perceptions about the potential usefulness and fairness of ADM are not the same as societal and individual acceptance of actual automated decisions. Insofar, more research is needed into whether, and the conditions under which people not only perceive ADM decisions fairer, but are also willing to accept an automated decision. All in all, the findings of this paper contribute to understanding of algorithmic appreciation in contemporary societies in general, and invite more in-depth reflection on the conditions under which ADM can and should play a role in decision making processes.

References

Agarwal R, Gao G (Gordon), DesRoches C, Jha AK (2010) Research Commentary—The Digital Transformation of Healthcare: Current Status and the Road Ahead. Inf Syst Res 21:796–809. doi: 10.1287/isre.1100.0327

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine Bias. In: ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 18 Jun 2018

Baek TH, Morimoto M (2012) Stay Away From Me. J Advert 41:59–76. doi: 10.2753/JOA0091-3367410105

Bickmore T, Utami D, Matsuyama R, Paasche-Orlow MK (2016) Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. J Med Internet Res 18:. doi: 10.2196/jmir.5239

Boerman SC, Kruikemeier S, Borgesius FJZ (2017) Online Behavioral Advertising: A Literature Review and Research Agenda. J Advert 0:1–14. doi: 10.1080/00913367.2017.1339368

Boerman SC, Kruikemeier S, Zuiderveen Borgesius FJ (2018) Exploring Motivations for Online Privacy Protection Behavior: Insights From Panel Data. Commun Res 0093650218800915. doi: 10.1177/0093650218800915

Bol N, Kruikemeier S, Boerman SC, et al (2018) Understanding the Effects of Personalization as a Privacy Calculus: Analyzing Self-Disclosure Across Health, News, and Commerce Contexts. J Comput-Mediat Commun

Carlson M (2018) Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. New Media Soc 20:1755–1772. doi: 10.1177/1461444817706684

Chu Z, Gianvecchio S, Wang H, Jajodia S (2012) Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? IEEE Trans Dependable Secure Comput 9:811–824. doi: 10.1109/TDSC.2012.75

Cox D, Cox AD (2001) Communicating the Consequences of Early Detection: The Role of Evidence and Framing. J Mark 65:91–103. doi: 10.1509/jmkg.65.3.91.18336

Davis FD (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 13:319–340. doi: 10.2307/249008

Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. Science 243:1668–1674. doi: 10.1126/science.2648573

Diakopoulos N, Koliska M (2017) Algorithmic Transparency in the News Media. Digit Journal 5:809–828. doi: 10.1080/21670811.2016.1208053

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144:114–126. doi: 10.1037/xge0000033

Dijkstra JJ, Liebrand WBG, Timminga E (1998) Persuasiveness of expert systems. Behav Inf Technol 17:155–163. doi: 10.1080/014492998119526

Dilsizian SE, Siegel EL (2013) Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. Curr Cardiol Rep 16:441. doi: 10.1007/s11886-013-0441-8

Dineen BR, Noe RA, Wang C (2004) Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. Hum Resour Manage 43:127–145. doi: 10.1002/hrm.20011

Dodge M, Kitchin R (2007) The automatic management of drivers and driving spaces. Geoforum 38:264–275. doi: 10.1016/j.geoforum.2006.08.004

Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv 4:eaao5580. doi: 10.1126/sciadv.aao5580

Elish MC, boyd danah (2018) Situating methods in the magic of Big Data and AI. Commun Monogr 85:57–80. doi: 10.1080/03637751.2017.1375130

European Commission (2018) Data protection working party

Ferrara E, Varol O, Davis C, et al (2016) The Rise of Social Bots. Commun ACM 59:96–104. doi: 10.1145/2818717

Field A (2013) Discovering statistics using IBM SPSS statistics. sage

Gillespie T (2014) The relevance of algorithms. Media Technol Essays Commun Mater Soc 167:

Graefe A, Haim M, Haarmann B, Brosius H-B (2018) Readers' perception of computer-generated news: Credibility, expertise, and readability. Journalism 19:595–610. doi: 10.1177/1464884916641269

Grolleman J, Dijk B van, Nijholt A, Emst A van (2006) Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation. In: IJsselsteijn WA, Kort YAW de, Midden C, et al. (eds) Persuasive Technology. Springer Berlin Heidelberg, pp 133–141

Hajian S, Bonchi F, Castillo C (2016) Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16. ACM Press, San Francisco, California, USA, pp 2125–2126

Hudlicka E (2013) Virtual training and coaching of health behavior: Example from mindfulness meditation training. Patient Educ Couns 92:160–166. doi: 10.1016/j.pec.2013.05.007

Inglehart R, Puranen B, Pettersson T, et al (2005) The world values survey

Jha S, Topol EJ (2016) Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA 316:2353–2354. doi: 10.1001/jama.2016.17438

Kennedy LW, Caplan JM, Piza E (2011) Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. J Quant Criminol 27:339–362. doi: 10.1007/s10940-010-9126-2

Kitchin R (2017) Thinking critically about and researching algorithms. Inf Commun Soc 20:14–29. doi: 10.1080/1369118X.2016.1154087

Lee MK (2018) Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data Soc 5:2053951718756684. doi: 10.1177/2053951718756684

Lee MK, Baykal S (2017) Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. ACM Press, pp 1035–1048

Logg J (2017) Theory of Machine: When Do People Rely on Algorithms? SSRN Electron J. doi: 10.2139/ssrn.2941774

Logg J, Minson J, Moore DA (2018) Algorithm Appreciation: People Prefer Algorithmic To Human Judgment. Social Science Research Network, Rochester, NY

Madhavan P, Wiegmann DA (2007) Effects of Information Source, Pedigree, and Reliability on Operator Interaction With Decision Support Systems. Hum Factors J Hum Factors Ergon Soc 49:773–785. doi: 10.1518/001872007X230154

McQuillan D (2015) Algorithmic states of exception. Eur J Cult Stud 18:564–576. doi: 10.1177/1367549415577389

Nass C, Steuer J, Tauber ER (1994) Computers Are Social Actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp 72–78

Newell S, Marabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification.' J Strateg Inf Syst 24:3–14. doi: 10.1016/j.jsis.2015.02.001

Nissan E (2017) Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. AI Soc 32:441–464. doi: 10.1007/s00146-015-0596-5

Nysveen H (2005) Intentions to Use Mobile Services: Antecedents and Cross-Service Comparisons. J Acad Mark Sci 33:330–346. doi: 10.1177/0092070305276149

O'Neil C (2017) Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books

Pierson E (2017) Demographics and discussion influence views on algorithmic fairness. ArXiv171209124 Cs

Siddiqui H, Healy E, Olmsted A (2017) Bot or not. In: 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST). pp 462–463

Smith A (2018) Public Attitudes Toward Computer Algorithms. Pew Research Center

Stanford University (2016) One Hundred Year Study on Artificial Intelligence (AI)

Sundar SS (2008) The MAIN model: A heuristic approach to understanding technology effects on credibility. Digit Media Youth Credibil 73100:

Sundar SS, Nass C (2000) Source Orientation in Human-Computer Interaction Programmer, Networker, or Independent Social Actor. Commun Res 27:683–703. doi: 10.1177/009365000027006001

Sundar SS, Nass C (2001) Conceptualizing sources in online news. J Commun 51:52–72

Thurman N, Moeller J, Helberger N, Trilling D (2018) My Friends, Editors, Algorithms, and I. Digit Journal 0:1–23. doi: 10.1080/21670811.2018.1493936

Thurman N, Schifferes S (2012) The Future of Personalization at News Websites. Journal Stud 13:775–790. doi: 10.1080/1461670X.2012.664341

van Dijck J, Poell T, de Waal M (2018) The platform society: public values in a connective world. Oxford University Press

Yeomans M, Shah A, Mullainathan S, Kleinberg J (2016) Making sense of recommendations

Yu K-H, Kohane IS (2018) Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf bmjqs-2018-008551. doi: 10.1136/bmjqs-2018-008551

Zarsky T (2016) The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. Sci Technol Hum Values 41:118–132. doi: 10.1177/0162243915605575

Tables and Figures

Table 1

Overview of the scenarios presented to the participants

	Low impact	High impact
Media	[AI / human editor] decides about news recommendations at the end of an article	[AI / human editor] decides whether to block access to Facebook and to news site
Health	[AI / human analyst] decides about fitness recommendations	[AI / human doctor] decides whether to provide a special medical treatment
Justice	[AI / human administrative employee] decides on whether a parking ticket should be given	[AI / human public prosecutor] decides whether a criminal lawsuit should be started

Table 2

Influence of personal characteristics on AI perceptions

	Usefulness [A]	Fairness	Risk
Intercept	3.64 (0.28)***	3.57 (0.35)***	3.13 (0.26)***
Age	-0.01 (0.002)***	-0.004 (0.002)	0.005 (0.002)*
Gender (Female)	-0.16 (0.07)*	0.01 (0.09)	0.11 (0.07)†
Education	0.07 (0.02)**	0.04 (0.03)	0.03 (0.02)
Knowledge	0.21 (0.03)***	0.17 (0.03)***	0.03 (0.03)
Belief in Equality	0.04 (0.02)*	0.05 (0.02)*	-0.02 (0.02)
Self-Efficacy	0.06 (0.03)*	0.11 (0.04)**	-0.09 (0.03)**
Privacy Concerns	-0.08 (0.03)*	-0.13 (0.04)***	0.25 (0.03)***
Adjusted R2	0.14	0.07	0.11

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; N = 958
[A] Same pattern (direction of the coefficient and significance) seen for usefulness including the reversed item, except for self-efficacy (not significant).

Table 3.

Multilevel models for scenario evaluations of Fairness, Risk and Usefulness

	Usefulness	Fairness	Risk
Fixed effects			
Decision-maker			
AI (vs. Human)	0.18 (0.12)	0.14 (0.11)	-0.06 (0.11)
Context			
Health (vs. Justice)	-0.24 (0.11)*	-0.15 (0.11)	0.16 (0.1)
Media (vs. Justice)	0.36 (0.11)**	0.44 (0.11)***	0.07 (0.11)
Decision-maker X Context			
AI X Health	0.2 (0.17)	0.15 (0.16)	-0.17 (0.16)
AI X Media	0.13 (0.16)	0.01 (0.16)	-0.25 (0.16)
Impact of the decision			
High impact (vs. Low)	0.09 (0.07)	0.09 (0.07)	-0.37 (0.06)***
Subject of the decision			
Self (vs. Other)	-0.13 (0.07)†	-0.05 (0.07)	0.06 (0.06)
Age	0.001 (0.002)	0.004 (0.002)†	-0.003 (0.002)
Gender (Female)	-0.04 (0.09)	0.03 (0.08)	-0.04 (0.08)
Education	-0.01 (0.03)	-0.01 (0.02)	0.04 (0.02)†
Knowledge	-0.05 (0.03)	-0.02 (0.03)	0 (0.03)
Belief in Equality	0.01 (0.02)	0.03 (0.02)	-0.02 (0.02)
Self-efficacy	0.05 (0.04)	0.03 (0.03)	0.09 (0.03)**
Privacy Concerns	0.09 (0.04)*	0.09 (0.03)**	-0.07 (0.03)*
Intercept	2.91 (0.35)***	2.72 (0.34)***	4.27 (0.32)***
Random-effects Parameters			
Var (u_j)	0.76 (0.05)	0.68 (0.05)	0.57 (0.05)
Var (intercept e_{0j})	1.38 (0.03)	1.39 (0.03)	1.37 (0.03)
ICC	0.23 (0.03)	0.19 (0.03)	0.15 (0.03)
AIC	7654.68	7598.60	7461.47
BIC	7750.36	7694.28	7557.15
Contrasts with Bonferroni adjustments: AI vs. Human decision-maker			
Justice	0.18 (0.12)	0.14 (0.11)	-0.06 (0.11)
Health	0.38 (0.12)*	0.29 (0.12)	-0.23 (0.11)
Media	0.3 (0.12)	0.15 (0.11)	-0.31 (0.11)†

Notes: Standard Errors in parenthesis. Respondents set as contextual (group) level as each respondent saw more than one scenario. Only responses that correctly identified the decision-maker in the manipulation check included (N = 2056). With all evaluations included (N = 2874), AI has significantly higher levels of usefulness and fairness than humans in all contexts, and lower levels of risk for media.
† $p < 0.1$, * $p < 0.5$, ** $p < 0.01$, *** $p < 0.001$

Table 3.

Linear regression models for Usefulness, Fairness and Risk of ADM by AI per context

	Usefulness			Fairness			Risk		
	Justice	Health	Media	Justice	Health	Media	Justice	Health	Media
				OLS Regression					
Decision-maker									
AI (vs. Human)	-0.4 (0.21)†	0.18 (0.2)	0.23 (0.23)	-0.45 (0.21)*	0.14 (0.2)	0.12 (0.22)	0.38 (0.2)†	0.09 (0.21)	-0.24 (0.21)
Subject of the decision									
Other (vs. Self)	-0.08 (0.16)	0.01 (0.15)	-0.2 (0.18)	-0.29 (0.16)†	0.24 (0.15)	-0.13 (0.17)	0.18 (0.15)	-0.07 (0.15)	0.16 (0.16)
Decision-maker X Subject									
AI X Other	-0.05 (0.24)	0.02 (0.23)	-0.2 (0.25)	0.24 (0.24)	-0.29 (0.23)	0.05 (0.24)	-0.1 (0.23)	-0.08 (0.23)	0 (0.22)
Impact of the decision									
High (vs. Low)	-0.52 (0.16)**	0.05 (0.15)	-0.09 (0.18)	-0.36 (0.17)*	-0.03 (0.15)	-0.13 (0.17)	-0.18 (0.16)	-0.31 (0.15)*	0.16 (0.16)
Decision-maker X Impact									
AI X High	1.17 (0.24)***	0.33 (0.23)	0.22 (0.25)	0.91 (0.24)***	0.67 (0.23)**	0.01 (0.24)	-0.89 (0.23)***	-0.56 (0.23)*	-0.18 (0.22)
Age	0.003 (0.004)	-0.003 (0.003)	0.004 (0.004)	0.003 (0.004)	0 (0.003)	0.01 (0.004)**	0.001 (0.004)	0 (0.004)	-0.01 (0.003)**
Gender (Female)	-0.14 (0.13)	-0.13 (0.12)	0.2 (0.14)	-0.02 (0.13)	-0.05 (0.12)	0.18 (0.13)	0.1 (0.12)	0 (0.12)	-0.26 (0.12)*
Education	0.01 (0.04)	-0.07 (0.04)*	0.04 (0.04)	0 (0.04)	-0.06 (0.04)	0.03 (0.04)	0.02 (0.04)	0.06 (0.04)	0.03 (0.04)
Knowledge	-0.07 (0.05)	-0.12 (0.05)*	0 (0.05)	-0.01 (0.05)	-0.08 (0.05)	0.01 (0.05)	0.04 (0.05)	0.04 (0.05)	-0.06 (0.05)
Belief in Equality	0.07 (0.04)†	-0.02 (0.03)	-0.02 (0.04)	0.07 (0.04)†	0 (0.03)	0.03 (0.04)	0.02 (0.03)	-0.04 (0.03)	-0.04 (0.03)
Self-efficacy	-0.03 (0.05)	0.11 (0.05)*	0.07 (0.05)	-0.04 (0.05)	0.08 (0.05)	0.06 (0.05)	0.08 (0.05)†	0.05 (0.05)	0.11 (0.05)*
Privacy Concerns	-0.07 (0.05)	0.11 (0.05)*	0.21 (0.06)***	-0.04 (0.06)	0.06 (0.05)	0.24 (0.05)***	-0.03 (0.05)	-0.03 (0.05)	-0.14 (0.05)**
Intercept	4.03 (0.53)***	3.12 (0.49)***	2.19 (0.55)***	3.82 (0.54)***	3.2 (0.49)***	1.85 (0.52)***	3.55 (0.5)***	4.15 (0.49)***	5.01 (0.49)***
Adjusted R2	0.03	0.03	0.02	0.02	0.03	0.03	0.05	0.04	0.04

Contrasts with Bonferroni adjustments: AI vs. Human Decision-Maker									
Low impact	-0.42 (0.17)†	0.19 (0.16)	0.13 (0.19)	-0.33 (0.17)	-0.02 (0.17)	0.15 (0.17)	0.33 (0.16)	0.05 (0.17)	-0.24 (0.16)
High impact	0.75 (0.17)***	0.52 (0.16)**	0.36 (0.17)	0.58 (0.17)**	0.65 (0.16)***	0.16 (0.16)	-0.56 (0.16)**	-0.5 (0.16)*	-0.41 (0.15)*

Notes: Standard Errors in parenthesis. Respondents set as contextual (group) level as each respondent saw more than one scenario. Only responses that correctly identified the decision-maker in the manipulation check included for Justice (N = 695), Health (N = 668) and Media (N = 693). With all evaluations included for (N = 958 per context), AI compared to humans has significantly higher levels of usefulness for Justice in high-impact scenarios, for Health in both Low and High impact scenarios, and for Media in high-impact scenarios; for fairness, AI scores significantly higher than humans for Justice (high-impact), Health (high-impact) and Media (high-impact); for risk, AI scores significantly lower than humans for Justice (high-impact), Health (high-impact) and Media (high-impact).

† $p < 0.1$, * $p < 0.5$, ** $p < 0.01$, *** $p < 0.001$