sociological science

Social Status and the Moral Acceptance of Artificial Intelligence

Patrick Schenk,[a] Vanessa A. Müller,[a] Luca Keiser[b]

a) University of Lucerne; b) gfs.bern

Abstract: The morality of artificial intelligence (AI) has become a contentious topic in academic and public debates. We argue that AIs moral acceptance depends not only on its ability to accomplish a task in line with moral norms but also on the social status attributed to AI. Agent type (AI vs. computer program vs. human), gender, and organizational membership impact moral permissibility. In a factorial survey experiment, 578 participants rated the moral acceptability of agents performing a task (e.g., cancer diagnostics). We find that using AI is judged less morally acceptable than employing human agents. AI used in high-status organizations is judged more morally acceptable than in low-status organizations. No differences were found between computer programs and AI. Neither anthropomorphic nor gender framing had an effect. Thus, human agents in high-status organizations receive a moral surplus purely based on their structural position in a cultural status hierarchy regardless of their actual performance.

Keywords: morality; artificial intelligence; organizational status; gender; bias; transparency

Reproducibility Package: A reproduction package with data, codebook, and statistical code is available through the following link: https://doi.org/10.5281/zenodo.13850548

NOBODY likes a lazy doctor. Doctors should work conscientiously. They should be reliable. They should give everyone their best attention irrespective of the patient's religious beliefs or ethnic background. But let us ask you: given a doctor's performance is fully in line with the standards of their profession, do you think their actions are more permissible because they are male or female? Does it matter if they work in a highly respectable hospital? Does it matter if they are a human being or an artificial intelligence (AI)? For (most) moral philosophers, the answer is evident: the moral evaluation of an agent should not depend on their social position. Status sensitivity in moral judgment is an epistemic distortion (Argetsinger 2022; Willemsen et al. 2023). Hence, given equal performance, it should be irrelevant whether the doctor is male or female, works in a prestigious hospital, or even is human or AI. In this article, we show that it does.

We develop the argument that an agent's social status causally affects their moral acceptance. Agent type, gender, and organizational rankings influence moral judgments irrespective of an agent's actual ability to accomplish a task in a concrete situation. We propose a theoretical framework integrating these three status dimensions. An agent's social status causally affects its moral acceptance due to the membership in a hierarchically ordered cultural category associated with competence expectations. Similar to other agents, AI could be evaluated based on its social status. For instance, whether using AI for cancer diagnostics is considered morally questionable depends not only on the violation of some ethical norms but

Citation: Schenk, Patrick, Vanessa A. Müller, Luca Keiser. 2024. "Social Status and the Moral Acceptance of Artificial Intelligence" Sociological Science 11: 989-1016.

Received: August 20, 2024

Accepted: September 29, 2024

Published: October 29, 2024

Editor(s): Ari Adut, Stephen Vaisey

DOI: 10.15195/v11.a36

also on social categories attributed to AI, such as male gender or organizational membership.

We argue that the type of agent, for example, human agents and artificial entities, can be fruitfully theorized as an additional status characteristic besides gender and organizational ranking, extending traditional sociological theories of social status (Berger, Ridgeway, and Zelditch 2002; Podolny 1993; Ridgeway 1991; Sauder 2005). AI is not just a new technology but serves as a cultural category tied to imaginaries, myths, and expectations surrounding its potential and danger (Gamez et al. 2020; Sartori and Theodorou 2022; Shank et al. 2021; Suchman 2023). This conceptual move allows us to empirically compare the relative importance of agent type, gender, and organizational ranking in a theoretically coherent way. It allows us to contrast status effects with observable performance—a long-standing interest of status research (Biegert, Kühhirt, and van Lancker 2023; Sauder, Lynn, and Podolny 2012). Finally, we are able to assess whether the effects of gender and organizational status differ between human agents and AI, taking up current discussions on the applicability of established social theories to the emerging phenomenon of AI (Kelly, Kaye, and Oviedo-Trespalacios 2023).

This argument is put to the test with a factorial survey experiment (FSE) on the moral permissibility of employing AI in three concrete situations: cancer diagnostics in a hospital, hiring by a recruitment agency, and fact-checking in the editorial office of a newspaper. These cases depict scenarios where AI systems are already in use, raising ethical questions on algorithmic discrimination, explainability, and accountability (Da Motta Veiga, Figueroa-Armijos, and Clark 2023; Kelly et al. 2023; Larkin, Drummond Otten, and Árvai 2022; Lavanchy et al. 2023; Schaap, Bosse, and Hendriks Vettehen 2024; Shulner-Tal, Kuflik, and Kliger 2023). In line with our theoretical model, we constructed vignettes with three status dimensions: the type of agent (AI vs. simple computer program vs. human), the agent's gender (male vs. female), and organizational status (among the top vs. bottom positions in a ranking) (Ridgeway 2014; Sauder et al. 2012). To compare the impact of these status dimensions with observable performance, we additionally manipulate the outcome of the task, transparency, and bias (Kieslich, Keller, and Starke 2022; Shin and Park 2019). The survey was administered online to 578 participants on the crowdsource platform Prolific.

AI has become a timely and contentious topic in the ethical, political, and sociological debates (Bostrom 2014; Joyce et al. 2021). Several arguments have been made about the conditions under which the use of AI systems is morally justified (Kieslich et al. 2022; Shin and Park 2019). We show how the moral acceptability of AI also rests on status characteristics. In contrast to the vast majority of research in engineering, psychology, and AI ethics, where an agent's behavior is understood independently of social context (Bigman and Gray 2018; Kieslich et al. 2022; Shin and Park 2019; Shulner-Tal et al. 2023), we hence underscore the social embeddedness of moral judgments (Haidt and Baron 1996; Willemsen et al. 2023). Although some research on attitudes toward AI has looked into the effects of agent type (e.g., Bigman and Gray 2018; Dietvorst, Simmons, and Massey 2015; Gamez et al. 2020, Schaap et al. 2024) and agent's gender (e.g., Ahn, Kim, and Sung 2022; Bernotat, Anne Eyssel, and Sachse 2021; Borau et al. 2021), no research so far

has analyzed organizational status (Sauder 2005) and barely any contrasted AI to simple computer programs (Shank et al. 2020, 2021), let alone comparing various dimensions of social status and actual performance in one coherent framework.

We first specify a general theoretical mechanism for the conversion of social status into moral acceptance, applicable to human agents, collective actors, and technological artifacts alike. We then discuss in detail the status beliefs related to AI, gender, and organizational ranking, deriving hypotheses on their relationship with moral acceptance. Afterward, we present the data and methods, followed by the results. In the next section, we discuss the findings, situating them into research on social status and research on AI, before summarizing the main results, acknowledging limitations, and pointing out practical implications in the concluding section.

Theory

The Conversion of Social Status into Moral Acceptance

We define status as prestige, honor, or esteem attributed to an entity as a function of its position in a system of hierarchically ranked social categories (Sauder et al. 2012; Willer et al. 2012). Sociologists have conceptualized very different things to have social status. Classical theories referred to the social standing of individuals within a social group (Ridgeway 1991; Weber 2019). Economic sociologists referred to the status of collective actors, such as firms or market intermediaries, to explain economic outcomes (Espeland and Sauder 2007; Podolny 1993). Status has also been extended to material objects such as prestigious works of art or fine wines (Beckert and Rössel 2013; Bourdieu 2005). Clearly, all of these entities (individuals, collective actors, and material artifacts) differ in theoretically relevant aspects. For example, in contrast to human agents, material artifacts do not have motivations and beliefs guiding their performance when signaling status in social encounters (Anderson and Kilduff 2009; Sauder 2005). Similarly, collective actors do not derive social status from face-to-face interactions within a group (Ridgeway 1991). Yet, the sociological literature is ripe with the idea that social status yields similar effects on judgments, irrespective of the origins of social status.

To explain this common causal effect of social status on evaluative judgments, we are in need of a general status conversion mechanism. We argue that social status derives from the membership in hierarchically ordered cultural categories, activating generalized performance expectations, leading to judgments of differential worth (cf. Berger et al. 2002; Ridgeway 2014). Such an approach is agnostic about many of the assumptions regarding intentionality, motivations, or other qualities of the entity attributed social status. It is a subject-centered approach, focusing on the processes within the perceiver judging objects of varying status (Jasso 2020). Status as an explanatory concept for moral judgments is thus equally applicable to human agents, organizations, or material artifacts. All that is necessary is the triad of categorization, expectations, and differential worth. This move has the considerable advantage of enabling the analysis of technological artifacts from the perspective of social status theories, integrating them into a coherent framework

with established status dimensions. Technological entities, such as AI, may also have social status, given the notion of AI represents a ranked cultural category (we will return to this below in the next section) (Beer 2016). Naturally, such an approach neglects interactions on the microlevel (Ridgeway 1991; Sauder 2005) and broader dynamics on the macrolevel (Biegert et al. 2023; Espeland and Sauder 2007), only capturing one important moment of an encompassing societal status system.

The status conversion mechanism is fueled by status beliefs (Berger et al. 2002). Status beliefs define relevant attributes to categorize an object called status characteristics. Often, status characteristics refer to social groups, such as gender or organizational membership, but they may also refer to other social categories, such as high-status or popular culture. They also define the relevant set of states of characteristics and their rank ordering, for example, male or female gender (Anderson and Kilduff 2009; Campos-Castillo 2018). Most importantly, status beliefs give rise to competency expectations (Berger et al. 2002). In previous research, competence has been conceptualized as the ability to offer outcomes of high quality (Lynn, Podolny, and Tao 2009). It referred to the instrumental aspects of a task, to effectively bring about a certain outcome, let's say, conducting a correct diagnosis in the medical context (Campos-Castillo 2018). However, when theorizing moral acceptance, competence also includes the ability to accomplish a task in accordance with moral values, let's say, adhering to norms of non-discrimination when treating patients (see especially Willer et al. [2012] or Anderson and Kilduff [2009] for work on status and ethical behavior). Moral judgments refer to the common good of a group, taking into account the legitimate interests of fellow (human) beings and transcending narrow self-interest (Durkheim 2009). Something is morally acceptable (or permissible or ethically approved) if it is judged as morally good and right in this sense. It is morally questionable if it is judged as bad and wrong. Yet, as empirical research shows, instrumental and moral competence is often intertwined: the latter works as an indication for the former and vice versa (Figueroa-Armijos, Clark, and da Motta Veiga 2023). For instance, a firm's innovativeness is taken as evidence for its ability to accomplish ethical ideals (Da Motta Veiga et al. 2023). Individuals rated higher in fairness are also rated higher in general ability.

At least two different explanations for the impact of status beliefs on judgments have been put forward in the literature. First, once activated in a situation, competency expectations might lead to confirmation bias, resulting in more positive assessments of the observable performance of high-status entities (Biegert et al. 2023; Ridgeway 2014). For example, individuals might expect high-status agents to be highly competent at a task, such as medical diagnostics. Thus, they subconsciously focus on information that is consistent with their expectations, resulting in selective attention and selective recall, for instance, in terms of information about the recovery of the patient. Second, status has been conceptualized as a signal for unobservable qualities (Gambetta 2011; Podolny 1993). For instance, individuals might reason that status is a signal for the general resourcefulness, skills, experience, or trustworthiness of a user of a new technology. Status should therefore be especially consequential in circumstances characterized by incomplete information, high uncertainty, and high-stake consequences. This is surely the case with AI, entailing unknown potentials and risks, unsolved ethical challenges, resulting in general

optimism or fear (Zhang et al. 2021). It follows from the second explanation that status effects should be especially pronounced for AI compared to more familiar situations, such as using a simple computer program or employing human agent to perform a task.

None of these explanations assume that competency expectations are accurate. Quite the opposite, they might be completely false and unrelated to innate differences between individuals or objects (Lynn et al. 2009; Ridgeway 1991). Indeed, status only has explanatory value if competency expectations do not perfectly mirror actual competence or quality (Sauder et al. 2012). Although competency expectations might partially rest on past performance, various ways have been described in which actual ability/quality and status become decoupled, for example, through cumulative advantage (Biegert et al. 2023). In addition, status beliefs possess substantial inertia once established, especially if they are widely shared, being part of the institutional fabric of a society (Bourdieu 2005; Sauder 2005; Willer et al. 2012). Hence, more than a simple affirmation of actual ability or quality, competency expectations reflect historically formed cultural beliefs that are, to various extents, socially constructed (Lynn et al. 2009). Status beliefs constitute an enduring, transsituational framework of typifications, giving rise to expectations in concrete encounters (Berger et al. 2002). It follows that an individual's judgments are independently and directly influenced by status beliefs (Benjamin and Podolny 1999; Ridgeway 2014). Status has structural effects in a proper sense. These effects solely derive from an agent's or object's position within a cultural status hierarchy, independent of the actual abilities of the position's incumbent (Bourdieu 2005; Sauder 2005).

Three Dimensions of Social Status: Agent Type, Gender, and Organizational Status

As cultural categories and symbols (Sauder 2005), status characteristics need to be understood in a specific historical context. In this section, we turn to the content of status beliefs for three status dimensions: agent type, gender, and organizational status. These characteristics are of great interest to sociological theory, see, for example, the discussions on agent type (Rammert 2016; Spillman 2023; Suchman 2023), gender (Auspurg, Hinz, and Sauer 2017; Ridgeway 1991), and organizational status (Benjamin and Podolny 1999; Podolny 1993; Sauder 2005).

We define agent type as the type of entity fulfilling a task, for example, diagnosing a patient. The agent dimension encompasses human and non-human entities, namely human agents, AI, and simple computer programs (Mays 2021; Shank et al. 2021). For our purpose, we broadly define AI as an entity "capable of displaying [. . .] behaviors that we consider to be intelligent" (Arkoudas and Bringsjord 2014, P. 34). In contrast to prominent accounts in science and technology studies, such as the actor–network theory, agent type does not make any ontological claims, for example, assuming that non-human actants are equipped with the same agency as human actors (Latour 2007). As a status characteristic, it refers to a cultural category. It relates to the idea or notion of AI, not the underlying code or some other ontological attributes (Beer 2016; Gamez et al. 2020; Suchman 2023). Individ-

uals react to an entity based on its category membership. Experimental evidence supports this perspective. Krach et al. (2008) found that participants attributed more intelligence and showed stronger brain activity in cortical regions (i.e., the classical theory-of-mind network) when told they were interacting with a human agent instead of a computer or an anthropomorphic robot, although the interaction partner always played a random sequence.

Such an approach is in accordance with the Computers Are Social Actors paradigm, stating that people respond with social scripts to advanced technological artifacts (Nass and Moon 2000). These technological entities are perceived as sources rather than mere transmitters of social interaction. Using social scripts means that people actively construct the social nature of the other, thereby categorizing them as human or non-human agents. The tendency to attribute humanlike attributes might even be stronger with self-learning and autonomous AI. Due to its complexity and obscurity, AI might be perceived and treated with an intentional stance, as if it behaves rationally and consistently, with beliefs and desires (Dennett 1971). This hints at the possibility that individuals split simple computer programs and advanced AI into different status categories.

Agent types and their associated beliefs vary historically and socially (Spillman 2023). Three examples might help to illustrate this point. First, the category of the firm as a collective actor with rights and responsibilities has only evolved in modern times (Rammert 2016). Firms as legal agents are a product of modernization in western societies. Second, in medieval times, non-human animals were subjected to animal trials, assuming the ability to understand legal parlance, contrary to our current view of the responsibilities and capabilities of non-human animals (Dinzelbacher 2002). Finally, the very definition of AI has shifted over the past century, encompassing more and more aspects of human intelligence, such as logical thinking, expertise, strategy, and more lately creativity (Woolgar 1985). In short, the boundaries of the social world are contingent. Who counts as social agent varies between historical epochs and cultures.

Agent types are associated with competency expectations, leading to a hierarchical ordering of status categories (Mays 2021). The category of AI is linked to technology myths (Elish and boyd 2017) and narratives (Sartori and Theodorou 2022). However, neither public discourse nor empirical research is consistent regarding the status value of AI. On the one hand, people attribute high general competence to AI. AI is believed to be more performant, expert, accurate, objective, unbiased, and rational compared to humans. On the other hand, people are skeptical about an AI system's reliability, moral capabilities, and its potential for eliminating human bias (Borau et al. 2021; Figueroa-Armijos et al. 2023; Zajko 2022). In addition, fear and distrust combined with uncertainties about AI regulation and performance contribute to the hesitation in adopting AI technologies (Zhang et al. 2021).

The majority of empirical findings on the perception of AI points to an aversion toward AI giving advice (Dietvorst et al. 2015; Larkin et al. 2022) or making moral decisions (Bigman and Gray 2018). Research also finds a positive correlation between human likeness and trust (Xu, Chen, and You 2023). In addition, humans are generally attributed higher levels of fairness (Lavanchy et al. 2023) and virtuous

character than AI (Gamez et al. 2020). Yet, there is some counterevidence with respondents actually preferring advice from AI systems (Logg, Minson, and Moore 2019), choosing an AI doctor over a human counterpart (Bigman and Gray 2018), or believing humans are weaker than AI (Shank et al. 2021). Finally, Shank, DeSanti, and Maninger (2019) and Schaap et al. (2024) found no statistical differences in moral permissibility between human and AI agents. There is far less research contrasting self-learning AI and simple computer programs. Yet, in two studies, respondents perceived simple computer programs as more powerful, more active, and better than AI (Shank et al. 2020, 2021).

In sum, we hypothesize that agent type as a status characteristic impacts the entity's moral acceptance. Although inconclusive, the empirical evidence rather points to AI aversion (Bigman and Gray 2018; Dietvorst et al. 2015; Gamez et al. 2020; Larkin et al. 2022; Lavanchy et al. 2023; Shank et al. 2020, 2021). By extension, we expect that human likeness will increase moral acceptance (Krach et al. 2008; Xu et al. 2023). Thus, an anthropomorphic framing of AI should additionally impact moral evaluation. It follows:

> H1: The use of an AI system to perform a task is judged to be less morally acceptable than a human agent performing the task.
>
> H2: The use of an AI system to perform a task is judged to be less morally acceptable than a simple computer program performing the task.
>
> H3: Anthropomorphic framing of AI increases the moral acceptability of using the AI system.

Second, an agent's gender is a highly salient status characteristic, structuring many human interactions (Ridgeway 1991). It rests on consensual cultural assumptions of gender role expectations (Auspurg et al. 2017). It commonly takes the form of a binary category in western societies with the states "men" and "women." As research on gender stereotypes has shown (Eagly 2013), male gender roles are associated with higher competence, assertiveness, and power, especially in the public sphere and socially highly valued tasks. Female gender is associated with domains of feeling, warmth, empathy, and compassion. Corroborating this research, previous studies found that female AI (or robots) evoke more affective trust (Bernotat et al. 2021; Siegel, Breazeal, and Norton 2009), are rated higher in warmth (Ahn et al. 2022), and generate more user acceptance (Borau et al. 2021). Yet, there is less support for gender differences in terms of task competence, agency, or human likeness (Ahn et al. 2022; Bernotat et al. 2021; Borau et al. 2021; Mays 2021). Hence, it is difficult to make a precise prediction for the effect of an agent's gender on moral acceptance. Although male gender generally confers higher status in western societies (Campos-Castillo 2018), gender is a diffuse status characteristic (Ridgeway 2014), being associated with a wide array of instrumental and moral competences. Therefore, we formulate an undirected hypothesis:

> H4: Gender framing of an AI impacts the moral acceptability of using the AI system.

The final dimension is organizational status (Espeland and Sauder 2007; Podolny 1993; Sauder 2005). It impacts moral acceptance through an agents' classification

as an organizational member (Sauder 2005). At least in modern western societies, rankings by third parties have become a major determinant for the status of organizations (Espeland and Sauder 2007). Third parties include but are not limited to rating agencies, critics, or social movements. Their impact has been widely documented in economic sociology, especially in markets for cultural goods, such as art (Beckert and Rössel 2013) or fine wines (Podolny 1993). Organizational status conferred by third-party ranking is a hierarchical, graduated, and achieved status characteristic (Ridgeway 1991). It necessarily comprises an ordinal scale from better to worse and is commonly understood as an outcome of past performance in line with meritocratic ideals (Biegert et al. 2023; Lynn et al. 2009).

High-status organizations should therefore be seen as highly competent in accomplishing a valued task. For example, status might serve as a signal for the unobservable abilities of an organization's members to use new technologies, such as AI, skillfully and beneficially. It might also be read as a signal for an organization's resourcefulness in acquiring high-quality technology. Whether high-status organizations are also perceived as more ethical is unclear. On the one hand, high-status agents might be seen as more egoistic, less compassionate, and more extrinsically motivated to gain fame and money (Cheng, Weidman, and Tracy 2014). On the other hand, recent experimental research shows that individuals only acquire status if they act generously and in line with the group's interest (Anderson and Kilduff 2009; Baker and Bulkley 2014; Willer et al. 2012). Hence, achieved status might serve as a sign of general competence and the willingness to act ethically at the same time (Campos-Castillo 2018; Figueroa-Armijos et al. 2023). Organizational status could therefore be especially crucial to alleviate potential fears surrounding AI and generate trust.

To the best of our knowledge, there is no research on the causal effects of organizational status on moral acceptability of AI. However, there is related empirical research. Some studies manipulated the status of a virtual agent in terms of its level of expertise by a specific title (expert or Dr.) and appearance (e.g., wearing a lab coat) (Horstmann, Gratch, and Krämer 2021; Obaid et al. 2016). None of these studies found effects of social status on the (moral) evaluation of the agent. Another strand of recent studies investigates the effects of an agent's status within an organization's hierarchy on moral judgments. Artificial agents took either the position of supervisor, peer, or subordinate in a collaborative task. Although participants attributed more blame and responsibility to AI with increasing status in a study by Lei and Rau (2021), Mays (2021) found no differences in the attribution of humanlike qualities and likeability. A last strand of empirical studies theorized the opposite causal direction from moral acceptability to organizational status. They found the ethical use of AI in hiring practices to be positively correlated with ratings of organizational attractiveness (Acikgoz et al. 2020; Da Motta Veiga et al. 2023). In general, then, although there is some evidence for the relevance of hierarchical social status for the moral acceptability of AI, the specific causal effect of organizational status is unknown. Based on the previous theoretical arguments, we formulate the hypothesis:

> H5: Organizational status increases the moral acceptability of using the AI system.

Performance Criteria and Moral Acceptability of AI: Outcome, Transparency, and Bias

We argue that the social status of a human or artificial agent impacts their moral acceptability independent of the agent's observable performance in a situation. In line with recent studies comparing the perceived importance of normative criteria for the moral assessment of AI in the general population (Kieslich et al. 2022; Shin and Park 2019), we include three criteria for observable instrumental and moral performance with an especially strong impact on moral evaluations in our study.

First, outcome refers to the agent's instrumental performance. It depicts the ability to accomplish a task without making an error, resulting in favorable consequences. For example, making a correct medical diagnosis leading to a patient's recovery (Kieslich et al. 2022). Seminal studies by Dietvorst et al. (2015) have shown that how users quickly lose confidence if they see algorithmic systems making mistakes. Similarly, studies found reliability and usefulness to be major determinants of the trustworthiness and perceived fairness of AI systems (Figueroa-Armijos et al. 2023; Glikson and Woolley 2020; Lavanchy et al. 2023; Park 2020; Schaap et al. 2024; Shulner-Tal et al. 2023).

Second, transparency and bias refer to the moral aspects of an agent's performance. Transparency means that an agent provides sufficient information for laypeople to reasonably understand how a task was achieved. For example, on what grounds an agent arrived at a diagnosis (Shulner-Tal et al. 2023). Otherwise, the process is opaque. Transparency has been identified as a crucial ingredient of trust and fairness perception by users of AI systems in various studies (Glikson and Woolley 2020; Hoff and Bashir 2015; Park 2020; Shin and Park 2019). Finally, bias refers to systematic discrimination based on group membership (Zajko 2022). An example would concern differential risks of misdiagnosis among patients with or without migration background. Research found that algorithmic bias causes uncanny feelings toward AI (Sullivan, de Bourmont, and Dunaway 2022).

Data and Methods

To test our hypotheses on social status and the moral acceptability of AI systems, we used a Factorial Survey Experiment (FSE). In FSEs, respondents evaluate detailed descriptions of hypothetical situations (vignettes), in which theoretically relevant attributes (dimensions) are systematically varied in their levels (Auspurg and Hinz 2015; Jasso 2020; Wallander 2009). FSEs are particularly suited to test the causal effects of social status on the moral acceptability of AI. First, they put us in a better position to isolate the specific causal effect of social status on moral acceptance than studies employing observational ex-post designs. They allow us to disentangle status bias from the effect of actual quality (Benjamin and Podolny 1999; Biegert et al. 2023) and rule out alternative causal directions such as ethically acceptable behavior leading to high status (Acikgoz et al. 2020; Da Motta Veiga et al. 2023; Willer et al. 2012). Second, responses in FSEs are less prone to social desirability bias, which is crucial when researching moral acceptability and highly sensitive topics such as gender stereotypes (Auspurg et al. 2017). Finally, unlike conventional survey items,

An **artificial intelligence** *[agent]* **named Paul** *[anthropomorphization & gender]*, able to learn by itself, is used at a hospital's oncology unit. This hospital occupies **one of the first places** on a ranking of all German hospitals *[organizational status]*.

A patient is checked for cancer in this hospital. Without human supervision, the artificial intelligence Paul makes the **diagnosis**. The diagnosis is that the patient has cancer. She **does not provide any further information**, so the patient cannot understand the reasons for the diagnosis *[transparency]*.

Based on this diagnosis by the artificial intelligence, the patient was treated. Since the diagnosis was **correct**, the patient was eventually **completely cured** *[outcome]*.

An independent analysis of all diagnoses revealed the following: There was **no evidence** that the diagnoses by the artificial intelligence Paul were more often wrong in patients with a migration background *[bias]*.

Figure 1: Sample vignette for the hospital situation. Dimensions in brackets.

vignettes provide detailed and vivid descriptions of concrete situations, which is particularly useful for a topic as abstract as the morality of AI.

Our vignettes describe hypothetical situations in which an agent carries out a task (Table 1 gives an overview of the vignette dimensions, Figure 1 presents a sample vignette, and additional sample vignettes are included in the online supplement). In total, we vary seven vignette dimensions, in line with methodological recommendations (Auspurg and Hinz 2015). First, the situation varies between three levels: (1) cancer diagnosis at a hospital, (2) hiring by a recruitment agency, and (3) fact-checking in the editorial office of a newspaper (Bigman and Gray 2018; Da Motta Veiga et al. 2023; Kelly et al. 2023; Larkin et al. 2022; Lavanchy et al. 2023; Shulner-Tal et al. 2023).

In terms of social status, we include a dimension for the type of agent. It varies between three levels: simple computer program, self-learning AI, and human. Because we are interested in status characteristics as cultural categories (Gamez et al. 2020; Suchman 2023), we intentionally did not provide a definition of AI to respondents, not to impose our own categorization. In addition, we included a dimension manipulating anthropomorphization and the agent's gender with three levels: a functional framing by a serial number ("G4-PLV"), an anthropomorphic framing by a male name ("Paul"), and an anthropomorphic framing by a female name ("Claudia") (Darling 2015; Mays 2021; Onnasch and Roesler 2019). To reinforce the anthropomorphic framing, we added a last name for the human agent ("Müller"). These names clearly signal gender, while being uncorrelated with expectations of general competence, for example, in terms of intelligence (Rudolph, Böhm, and Lummer 2007). Finally, we manipulate organizational status by varying the relative position of the organization employing the agent (i.e., hospital, company, and newspaper) in a ranking by a third party. It varies between two levels: among the bottom positions versus among the top positions. The operationalization follows experimental paradigms using awards to manipulate social status (Cheng

Table 1: Vignette dimensions and levels

Dimension		Levels	
1	Situation	1	Cancer diagnosis in a hospital
		2	Fact-checking in the editorial office of a newspaper
		3	Hiring by a recruitment agency
2	Agent	1	Self-learning AI
		2	Simple computer program
		3	Human
3	Organizational status	1	Among top positions in ranking
		2	Among bottom positions in ranking
4	Anthropomorphization and gender	1	Functional framing by serial number
		2	Anthropomorphic framing by male name
		3	Anthropomorphic framing by female name
5	Outcome	1	Negative outcome
		2	Positive outcome
6	Transparency	1	Opaque
		2	Transparent
7	Bias	1	Biased
		2	Unbiased

et al. 2014) and observational studies using expert rankings as status measures (Podolny 1993).

Three additional dimensions describe the agent's observable performance in a situation. These dimensions pertain to the agent's actual performance at a given point in time, not to its previous, future, or general performance. First, instrumental performance with the outcome of the task. It has two levels: a positive or a negative outcome. For example, a correct diagnosis leading to a patient's recovery or a mistaken diagnosis leading to her death. Two additional dimensions describe whether the agent's behavior was in line with moral standards, that is, their moral performance. Transparency has two levels. In the transparent condition, the agent provides additional information necessary to understand how the task was fulfilled, for example, how the agent arrived at a diagnosis. In the opaque condition, the agent provides no additional information. Finally, we included bias. In the biased condition, the agent's behavior was less reliable among people with a migration background[1], for example, doctor's diagnoses are more often wrong for this group. In the unbiased condition, there is no evidence of a higher frequency of mistakes among people with migration background.

Finally, all vignettes explicitly stated that the AI and the simple computer program accomplish the task without human supervision, keeping the level of an artificial agent's autonomy constant (Kim and Hinds 2006; Schaap et al. 2024). We confirmed the validity, the comprehensibility, and the plausibility of the vignettes in cognitive pretests with participants from various age groups. Readability scores confirm that the vignettes are easy to read. The SMOG scores (Simple Measure of

Gobbledygook; McLaughlin 1969) are between 8 and 11, which is around the same value as for daily newspapers. The mean response time per vignette is around 80 seconds.

At the end of each vignette, we asked participants to rate the moral acceptability of employing the agent in the situation. The item was adapted to the agent and the situation. For example, in the case of a cancer diagnosis by the AI, it reads: "The fact that the hospital is using this artificial intelligence seems to me. . . ," with response categories from 1 "extremely morally questionable" to 7 "not morally questionable at all." Thus, higher values indicate that respondents consider it less morally questionable and more morally acceptable to employ the human or artificial agent. The mean of the variable is exactly at the scale's midpoint of 3.5, suggesting that respondents had enough freedom to discriminate between vignettes.

In total, we have a $3^3 2^4$ design. After excluding one logically impossible combination of levels (functional framing and human agent), the universe comprises 384 vignettes. To build a vignette sample, we used a fractionalized experimental design with D efficiency as target criterion (Auspurg and Hinz 2015). D-efficient designs maximize the orthogonality between dimensions and the variance of the levels, resulting in smaller standard errors. They are clearly preferable to random sampling. The resulting sample comprises 132 vignettes with a D efficiency of 99, which is well above the recommended threshold of 90 and close to the maximum of 100. According to our results in the cognitive pretests, six vignettes per person were optimal. Therefore, we partitioned the vignette sample into 22 decks, again using D efficiency as target criterion.

The factorial survey was administered online to 578 participants on the crowdsource platform Prolific. The experiment was not preregistered. The survey was titled "new technologies—new responsibilities." Although the topic of the survey was deliberately chosen to be broader than AI, we can't rule out a certain amount of self-selection of participants especially interested in new technologies. Yet, the distribution of the dependent variable does not indicate strong self-selection in terms of the moral acceptance of AI. The responses spread over the entire range of the scale with a mean close to the midpoint. There is clearly enough variation in the dependent measure to test our hypotheses. We used quota sampling according to gender and age. Because the vignettes were in German, only respondents with German as first language were eligible. Table 2 provides a description of the sample. Most notably, we find that respondents are rather young on average and well educated. The vast majority lives in a German-speaking country, namely Germany, Switzerland, or Austria. FSEs do not presuppose representative samples to ensure internal validity (Auspurg et al. 2017). Moreover, methodological research shows that findings from online experiments with crowdsource recruitment exhibit rather high data quality and population validity, especially when controlling for socio-demographics (Weinberg, Freese, and McElhattan 2014). Participants were paid a fair amount for their valuable time.

Given the number of participants, we arrive at 3,648 vignette judgments and 23–28 respondents per deck. Respondents were randomly assigned to one of the decks. Within the deck, we randomized the order of the situations (hospital, recruitment, and newspaper), and within the situation, we randomized the order of

Table 2: Description of the sample

Variable	*n*	
Gender		
Female	285	(49%)
Male	289	(50%)
Other	4	(1%)
Years of schooling		
Mean		16
Age in years		
Mean		32
Household income in dollars (PPP)		
Mean		2190
Region		
Germany	426	(74%)
Switzerland/Austria	70	(12%)
Other	82	(14%)

the vignettes. To control for carryover effects, we include the position of the vignette in the regression analysis. In addition, we computed models with interactions between the status dimensions and vignette position. None were statistically significant at the 5 percent level, indicating that effect sizes do not depend on vignette position. To ensure comprehensibility and text flow, we presented the vignettes in running text with dimensions in a fixed order. Düval and Hinz (2020) only found weak effects for the order of the dimensions and only for respondents with shorter response time. To check the robustness of the findings, we computed models with and without response time as an additional control. Response time was never significant, and the results were virtually identical. We also computed models with interaction terms between the status dimensions and response time. None were statistically significant. For these reasons, we exclude response time in the final analysis below.

The structure of our data is hierarchical with three levels: vignette judgments, decks, and respondents. The intraclass correlation coefficients (ICCs) from a multilevel regression model with random intercepts are 0.08 at the respondent level and 0.006 at the deck level. According to Snijders and Bosker (2012), only when ICC exceeds 0.05, the hierarchical structure should be taken into account. Hence, we use multilevel regressions with random intercepts and two levels: vignettes and respondents. All models include socio-demographic variables as controls, namely the respondent's gender (male, female, and other), age, years of schooling, household equivalence income adjusted for purchase power, and region (Germany, Switzerland/Austria, and other). We employ multiple imputations to handle missing values. We use 10 imputations, corresponding to the highest number of missing values (which is for income) (Enders 2010). We checked convergence and the distributions of the imputed data. There were no anomalies. We have sufficient power to detect even very small effects of $f^2 = 0.02$ at the 5 percent level in regression models with all controls. Only the model for the human agent is very slightly underpowered with $\beta = 0.77$, although we have enough power to detect very small

effect in a simpler model, sufficient to reliably test the hypotheses (see model M6 in footnote 2). All calculations were performed using R 4.3.0.

Results

The first model M1, shown in Table 3, regresses moral acceptability on the vignette characteristics and controls. It includes all vignettes for all three types of agents. Looking at agent type, only the coefficient for the human agent is statistically significant. Respondents evaluate human agents more favorably than AI. We do not find a statistically significant difference between AI and the simple computer program. A post-hoc test on the equality of the coefficients for the simple computer program and the human agent yields a highly significant result ($F = 326.67, p < 0.001$). Hence, respondents also evaluate the computer program differently from the human agent. In the next step, we want to briefly discuss the remaining effects, keeping in mind that they are based on all agent types, and not only AI.

We find a highly significant positive effect of organizational status on moral acceptability. An organization with a top position in a ranking significantly increases the moral acceptability of employing the human or artificial agent compared to an organization with a bottom ranking. In contrast, none of the variables for anthropomorphization or gender framing reach statistical significance. Hence, respondents do not attribute higher moral acceptability to a female agent in contrast to a male agent. Nor does a functional framing in contrast to framing by a human name decrease moral acceptability.

All of our dimensions for the agent's observable performance in a situation are highly significant covariates of moral acceptability. As expected, a negative outcome and bias decrease moral acceptability, whereas transparency increases it. Comparing the effects of agent type to situational performance, we see that the human agent effect is slightly stronger in absolute magnitude ($b = 1.17$) than the outcome ($b = -1.10$) and bias effects ($b = -1.05$) and more than twice as larger than the transparency effect ($b = 0.53$).

In models 2–4 (Table 3), we analyze the vignettes for each level of the agent dimension separately. Figure 2 plots the effects together with their 95 percent confidence intervals. These models exclude the variables for the agent dimension (M2–M4) and the variable for the functional framing in the model for the human agent (M4).

M2 for the AI is of special interest. The results are rather consistent compared to the model with all vignettes (M1). Organizational status has a positive impact on the moral acceptability of AI. If the organization employing the AI is among the top positions in a ranking, the use of AI to perform a task is rated more morally acceptable. This result applies not only to the AI but is also consistent across all three agents. Organizational status is clearly significant at the 0.1 percent level for the simple computer program and it is just significant at the 5 percent level for the human agent. As before, we do not find significant effects of male gender framing or a functional framing by a serial number in contrast to female gender framing. A post-hoc test on the equality of the coefficients for male gender framing and functional framing turns out insignificant ($F = 0.4674, p > 0.1$). Hence, male

Table 3: Multilevel regression of moral acceptability on vignette dimensions, respondent characteristics, and experimental design

	M1: All	M2: AI	M3: CP	M4: Human	M6: All
Vignette dimensions					
Agent: Human[1]	1.17***				0.77***
	(0.07)				(0.22)
Agent: Simple computer program[1]	−0.06				−0.18
	(0.06)				(0.20)
Organizational status	0.30***	0.20*	0.28***	0.22*	0.20*
	(0.05)	(0.09)	(0.09)	(0.11)	(0.09)
Male gender framing[2]	0.00	−0.05	0.07	0.11	0.00
	(0.06)	(0.11)	(0.12)	(0.11)	(0.06)
Functional framing[2]	−0.01	−0.13	0.08		−0.01
	(0.07)	(0.11)	(0.11)		(0.07)
Negative outcome	−1.1***	−1.12***	−0.85***	−1.36***	−1.10***
	(0.05)	(0.08)	(0.09)	(0.10)	(0.05)
Transparency	0.53***	0.52***	0.49***	0.60***	0.52***
	(0.05)	(0.09)	(0.09)	(0.10)	(0.05)
Bias	−1.05***	−0.77***	−0.89***	−1.74***	−1.04***
	(0.05)	(0.09)	(0.09)	(0.11)	(0.05)
Situation: Diagnosis[3]	−0.34***	−0.49***	−0.63***	0.39**	−0.36***
	(0.06)	(0.11)	(0.11)	(0.13)	(0.06)
Situation: Recruitment[3]	0.21***	0.00	0.10	0.50***	0.21***
	(0.06)	(0.11)	(0.11)	(0.13)	(0.06)
Organizational status × human					0.26†
					(0.14)
Organizational status × CP					0.08
					(0.13)
Respondent characteristics					
Gender: Male[4]	0.30***	0.30**	0.33**	0.22†	0.30***
	(0.08)	(0.11)	(0.11)	(0.12)	(0.08)
Gender: Other[4]	0.21	0.87	−0.44	0.05	0.22
	(0.46)	(0.63)	(0.64)	(0.68)	(0.46)
Years of schooling	0.04*	0.05*	0.03	0.06*	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age	0.00	0.00	−0.01	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Household income	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Region: Germany[5]	0.03	0.20	−0.18	0.00	0.03
	(0.12)	(0.16)	(0.16)	(0.18)	(0.12)
Region: Switzerland/Austria[5]	0.12	0.26	0.04	−0.09	0.12
	(0.15)	(0.21)	(0.21)	(0.23)	(0.15)
Experimental design					
Vignette order	0.00	−0.03	−0.02	0.08**	0.00
	(0.01)	(0.03)	(0.03)	(0.03)	(0.01)
Constant	4.42***	4.27***	4.46***	6.06***	4.58***
	(0.35)	(0.49)	(0.54)	(0.57)	(0.36)
Number of vignettes	3468	1262	1261	945	3468
Number of respondents	578	578	578	578	578

Note: Unstandardized coefficients with standard errors in parentheses. AI = self-learning artificial intelligence and CP = simple computer program. [1]Ref. cat. agent: AI, [2]ref. cat. female gender framing, [3]ref. cat. situation: newspaper, [4]ref. cat. gender: female, and [5]ref. cat. region: other.
†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

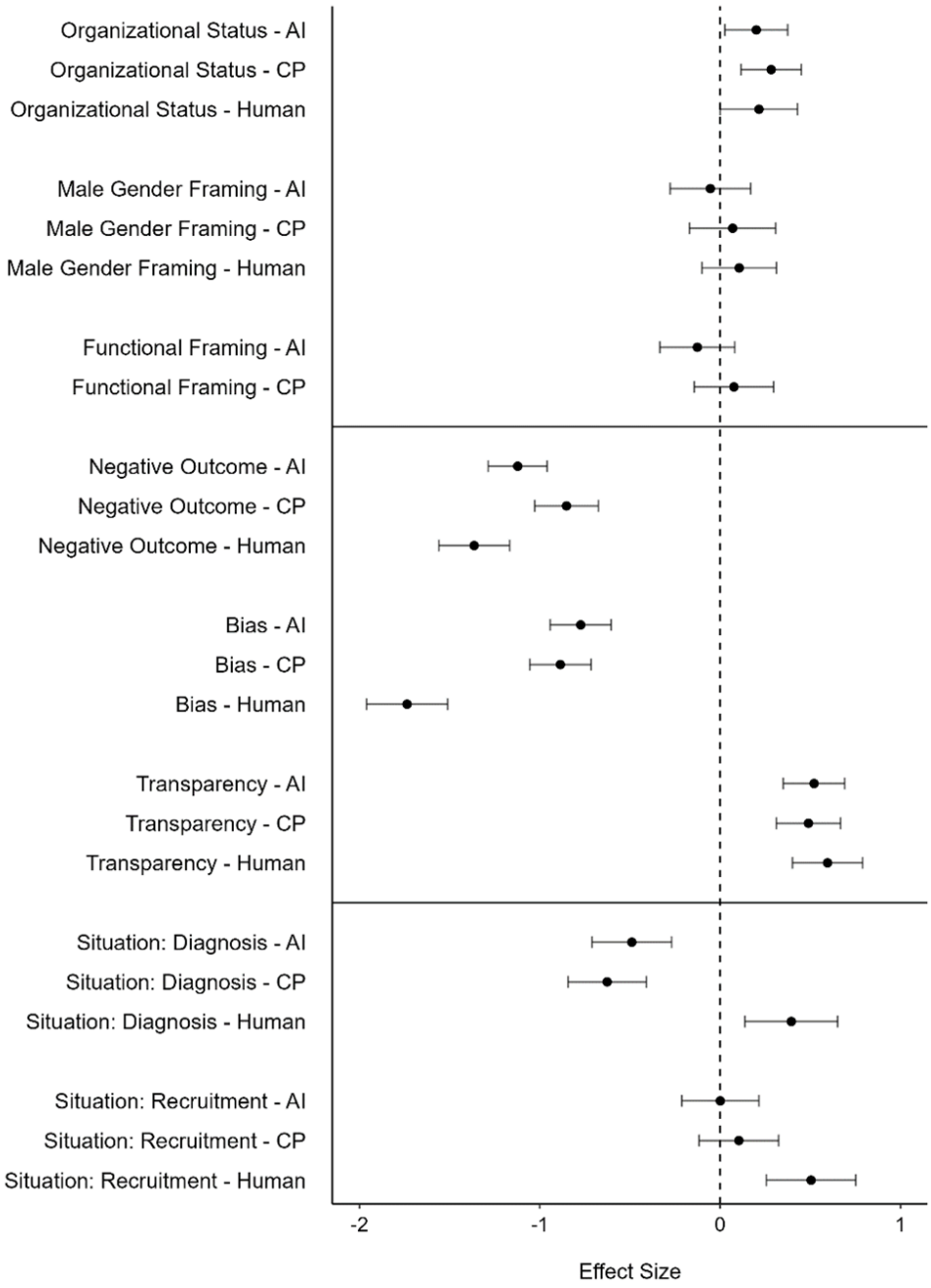

Figure 2: Coefficient plots of multilevel regression models 2–4. Unstandardized regression coefficients with 95 percent confidence intervals. AI = self-learning artificial intelligence and CP = simple computer program.

gender framing also doesn't increase or decrease moral acceptance in contrast to a functional framing by a serial number. These results are also consistent across agent types (M2–M4)[2].

To explicitly test whether the strength of the status effects varies between entities, we computed an additional model M5, including an interaction term between agent type and organizational status. We focus on organizational status because it is the only statistically significant status effect in the separate models M2–M4. We find that neither the interaction term with the simple computer program nor that with the human agent is statistically significant at the 5 percent level (Table 3). In addition, we use Akaike's information criterion (AIC) to compare these models. AIC is a measure for relative model fit, considering the number of independent variables. A lower AIC of 13,131 for M1 indicates a better fit compared to an AIC of 13,136 for M5. Hence, the simpler model M1, assuming that the effect size of organizational status does not differ between agent types, is preferable to the more complex model M5, assuming different effects of organizational status according to entity. This is further confirmed graphically in Figure 2. Looking at the effects of organizational status, the 95 percent confidence intervals of all three types of agents overlap, showing that there are no statistically significant differences. The same conclusions hold for gender and functional framing, which are, however, insignificant covariates in the first place.

Turning to the observable performance of AI, we again find that outcome, bias, and transparency are significant predictors. The absolute magnitude of these effects ($b = -1.12, b = -0.77, b = 0.52$, respectively) are substantially stronger than the effect of the organizational ranking ($b = 0.20$). Hence, actual performance has a comparatively strong impact on the moral acceptability of an AI system. This conclusion holds true for the other two agents, human and simple computer program (M3 and M4).

Looking at the three situations as the final dimension of our vignettes, we do find statistically highly significant effects, although the pattern is not consistent across agent types in this case. The use of AI or a simple computer program is considered less morally acceptable in the medical context than in the context of fact-checking for a newspaper (M2 and M3). For the human agent, it is the other way round. Human agents are judged more morally acceptable in the medical context and personal recruitment in contrast to fact-checking for a newspaper (M4).

On a final note, we want to point out that most of our control variables are insignificant. Looking at the respondent's characteristics, only male gender and years of schooling yield significantly positive effects. Neither age, household income, nor the region is a significant covariate, which speaks to the generalizability of the results to a broader population. Except for M4, we do not find any order effects.

Discussion

We find clear evidence for the unique causal influence of social status on the moral acceptability of using AI systems. Agent type and organizational status both yield statistically significant effects on moral permissibility. The moral acceptance of an

agent cannot be solely explained by its ability to accomplish a task in line with moral norms (Dietvorst et al. 2015; Glikson and Woolley 2020; Hoff and Bashir 2015; Kieslich et al. 2022; Lavanchy et al. 2023; Park 2020; Shin and Park 2019; Shulner-Tal et al. 2023; Sullivan et al. 2022). Social status exerts an independent, irreducible influence on moral evaluation (Willemsen et al. 2023). Extending sociological research on market outcomes (Beckert and Rössel 2013; Benjamin and Podolny 1999; Biegert et al. 2023; Podolny 1993) or justice perceptions (Auspurg et al. 2017), our study corroborates the explanatory importance of social status for moral acceptability.

We find pronounced differences in the moral permissibility of human and artificial agents. A human performing a task is clearly more morally acceptable to respondents than an AI system or a simple computer program performing the same task, irrespective of the outcome and the compliance with ethical standards. This supports our first hypothesis. The results are in line with previous findings on algorithm aversion (Bigman and Gray 2018; Dietvorst et al. 2015) and a deficit of trustworthiness and moral character for AI (Gamez et al. 2020; Larkin et al. 2022). In contrast to a study by Schaap et al. (2024), we do find strong effects of the agent category even when controlling for vital aspects of observable performance. Hence, the effect of agent type is not simply explainable by the "concrete benefits" of using AI (Schaap et al. 2024) or poor performance, that is, "seeing algorithms err" (Dietvorst et al. 2015). Moreover, human category membership has a slightly stronger effect than outcome and bias and clearly stronger than transparency. Hence, status beliefs associated with socially constructed agent categories exert an effect on moral acceptability over and above instrumental or moral performance (Shank et al. 2021; Spillman 2023; Suchman 2023). Therefore, theories of social status are well advised to integrate agent type as an additional and very powerful status characteristic. To respondents, it is more important who does the task than how they do the task (Willemsen et al. 2023).

Organizational status also yields a positive causal effect on the moral acceptability of using AI systems. If an AI is employed in an organization with a top position in a ranking, its use is deemed more morally permissible compared to an AI in an organization with a low ranking. This finding supports our fifth hypothesis, extending the scarce research on achieved graduated status characteristics and moral judgments of AI. The latter is influenced not only by positions within organizations, that is, supervisor and subordinate (Lei and Rau 2021; Mays 2021), but also by the organization's position in an organizational field (Bourdieu 2005; Podolny 1993; Sauder 2005). Because we manipulated status in an experimental design, we are confident in claiming that organizational status serves as a cause and not (only) as a consequence of the moral perception of using AI (Acikgoz et al. 2020; Da Motta Veiga et al. 2023). To respondents, it is of importance where an agent is employed and not only how an agent performs the task.

When asked to make a moral judgment about the use of AI, people may consider an organization's status as an indication of its general competence and trustworthiness in handling AI (Gambetta 2011; Podolny 1993). High-status organizations may be perceived as having the necessary resources and skills to implement AI effectively and beneficially, as well as the expertise to judge its performance. In contrast,

low-status organizations tend to be generally associated with poor performance, and the use of AI in such settings may be viewed as exacerbating existing deficiencies. Given the fear, uncertainty, and limited knowledge pertaining to AI (Cugurullo and Acheampong 2024; Sartori and Theodorou 2022), people may feel safer and more confident when high-status organizations are using it, also because these organizations may have more at stake if things go wrong. Organizational status can hence mitigate uncertainty when it comes to the implementation of AI. Uncertainty is heightened in safety-critical areas like medical diagnosis (Zhang et al. 2021), which aligns with our result that using AI in this setting is generally seen less morally acceptable.

Having said this, our models also show that organizational status yields a consistent and similar effect across all three agent types. Indeed, statistical analysis rather supports a more parsimonious model, combining all three entities, than separate models for each entity. This somewhat weakens the argument that AI is uniquely uncertain and risky, making status signals more important to cope with this emerging technology (Podolny 1993). The result is more in line with confirmation bias (Berger et al. 2002; Ridgeway 1991). Organizational status evokes expectations of greater instrumental and moral competence, subconsciously biasing the interpretation of a given performance in the direction of these expectations. Future research could explore these alternative explanations for the status effects more directly by including measures for trust, competence expectations, or even information processing.

The effect size of organizational status is smaller compared to the other determinants of moral acceptability. Yet, statistical effects must also be theoretically contextualized. Agent type and observable performance are directly related to the agent and its behavior in a concrete situation. Organizational status is causally much more distant by comparison. Status needs to "flow" from the ranking to the organization and from the organization to its members, only then impacting moral judgments in particular situations (Beckert and Rössel 2013; Benjamin and Podolny 1999). Moreover, organizational status might be more consequential in practice. In contrast to performance in a single situation, organizational status has an effect on all members of an organization simultaneously and across various situations (Sauder 2005). A change in organizational status is therefore amplified by the sheer number of agents it affects, whereas the impact of an unfavorable outcome is limited to one particular situation. Hence, the theoretical and practical ramifications of organizational status are larger than the statistical estimates might suggest.

Not all status characteristics are relevant for moral permissibility. A self-learning AI is not evaluated differently than a simple computer program. Either respondents lump these two types of agents together in one category (Suchman 2023) or these categories are not salient for the evaluation of moral acceptability (Berger et al. 2002). In contrast to the few studies showing that AI is perceived differently to simple computer programs (Shank et al. 2020, 2021), self-learning AI does not have a lower (or higher) status value than simple computer programs when it comes to moral acceptability. Thus, H2 is not corroborated. Likewise, we cannot confirm our third hypothesis on anthropomorphic framing. Although functional framing leads to a slight decrease in moral acceptability in contrast to female framing, the

effect does not reach any conventional level of statistical significance. This is not entirely surprising given the mixed findings on anthropomorphization and trust (Glikson and Woolley 2020) or empathy toward robots (Onnasch and Roesler 2019) in previous research, with overall small effect sizes of social cues (Xu et al. 2023).

An agent's gender does not have an impact on moral acceptability either. Consequently, we must also reject H4. One might argue that respondents do not attribute human gender to artificial agents despite having a male or female name (Fortunati et al. 2022). However, this does not explain why the gender effect is also absent in the model for the human agent. One might also refer to social desirability. Yet, FSEs are less prone to this type of bias (Auspurg and Hinz 2015). Finally, the result could be specific to the three situations studied in the experiment. As research suggests, gender status beliefs are more relevant in professional contexts with high levels of existing gender inequality (Ahn et al. 2022; Auspurg et al. 2017). The shares of women in oncology, journalism, and employment agencies are around 45 percent to 50 percent in Germany, Austria, and Switzerland (Bundesärztekammer 2022; Keil and Dorer 2019; statista 2023). This might be a reason for the absence of gender discrimination in the experiment.

The effects of agent type and organizational status do not imply that actual performance is irrelevant for moral acceptability. Quite the contrary, in line with previous research, the ability to fulfill a task without making mistakes, resulting in a favorable outcome, is highly relevant (Dietvorst et al. 2015; Figueroa-Armijos et al. 2023; Lavanchy et al. 2023; Park 2020; Schaap et al. 2024). Bias is slightly less important (Kieslich et al. 2022). Transparency turns out to be least important, in contrast to findings by Shin and Park (2019). In addition, we observe a substantial main effect of the situation. The moral acceptability of using AI for cancer diagnostics is generally lower compared to fact-checking in a newspaper or personal recruiting. The latter two might be seen as more mechanical, standardized, and low-stake tasks, demanding less interpersonal skills than medical treatment, thus making AI generally more permissible in these situations (Glikson and Woolley 2020; Larkin et al. 2022).

High status leads to a substantial increase in the moral acceptability of an agent. We might tentatively quantify the total amount of the status effect, comparing it to an agent's instrumental and moral performance. According to the results in model 1, the moral acceptability of a human agent in an organization with a top ranking is estimated to be around 1.5 scale points higher on average than an AI in an organization with a bottom ranking. Yet, this is insufficient to cover the performance impact. An agent producing a positive outcome in a fair and transparent way is rated around 2.6 scale points higher on average than an agent producing a negative outcome in a biased and opaque way. Hence, while social status puts an agent in a clearly better position, it is insufficient to fully make up for poor performance. Of course, these calculations need to be taken with a grain of salt, not to suggest a false sense of precision of statistical effect sizes.

Conclusion: Social Status as Moral Credit

An agent's moral acceptability depends not only on its ability to fulfill a task in line with the normative expectations put forward by philosophers and policymakers but also on its social status. In a Factorial Survey Experiment (FSE), respondents judged the moral permissibility of employing AI systems, simple computer programs, and human agents for cancer diagnostics, personnel recruitment, and fact-checking in a newspaper. Irrespective of the entity's ability to produce a favorable outcome in a fair and transparent way (Kieslich et al. 2022; Shin and Park 2019), human agents are judged as more morally acceptable than AI and simple computer programs (Bigman and Gray 2018; Dietvorst et al. 2015; Gamez et al. 2020; Larkin et al. 2022). AI and simple computer programs do not differ in their status value (Shank et al. 2020, 2021). Moreover, the use of AI in organizations with a top position in a ranking is perceived as more morally acceptable compared to a low-ranking organization.

Theories of social status are well advised to integrate agent type as an additional and very powerful status dimension. The effects of agent type are better explained by the activation of status beliefs fueled by imaginaries, myths, and narratives surrounding the capabilities and dangers of AI than by the overserved performance of an algorithm in a concrete situation (Dietvorst et al. 2015; Schaap et al. 2024). Moreover, the effect of organizational ranking is highly consistent across human agents, AI, and simple computer programs. These findings are hence also relevant for research on status effects among human agents in sociology and social psychology (Haidt and Baron 1996; Sauder et al. 2012; Willemsen et al. 2023). Established theories of organizational status are equally applicable to human and technological agents (Kelly et al. 2023; Nass and Moon 2000; Podolny 1993). Surprisingly, gender did not show similar status effects, despite being commonly theorized as crucial structuring principle of social interactions (Ahn et al. 2022; Ridgeway 1991). Anthropomorphizing the technological agent by a human name also did not produce a statistically significant difference in moral permissibility (Xu et al. 2023).

We want to point out two limitations of our study. First, we administered the survey online with a non-representative sample of German-speaking respondents on Prolific. Surely, as methodological research has shown, results based on samples from crowdsource platforms are readily generalizable to a broader population (Weinberg et al. 2014). Still, future research should try to replicate our results with more representative samples of the general population. Second, we had to make a selection of situations for the vignettes. We chose three situations of special interest to sociological and ethical debates. Future studies should include additional situations. This is especially pertinent for research on gender status beliefs, which might be more salient in contexts with high levels of gender inequality (Auspurg et al. 2017).

Although the total statistical effect of social status is smaller than the combined effects of instrumental and moral performance, it is sociologically and practically highly relevant. Status generates a moral surplus based on social category membership alone and irrespective of individual behavior. It grants leeway in meeting performance standards (Sauder et al. 2012), power to shape the rules of the game (Ridgeway 2014), and facilitates the accumulation of economic and non-economic

resources over time (Biegert et al. 2023). Human agents in high-ranking organizations are able to convert their social status into a type of moral credit, putting them in an advantageous structural position (Bourdieu 2005; Pellandini-Simányi 2014; Willer et al. 2012).

This is not necessarily warranted. AI algorithms have been shown to outperform humans at various tasks (Dietvorst et al. 2015). AI is under close scrutiny with developers and policymakers auditing these systems for ethical deficiencies, such as algorithmic discrimination (Akinrinola et al. 2024). In addition, sociologists have shown how rankings become decoupled from actual performance because, among others, organizations invest in practices that are rewarded by ranking agencies but are of little relevance to actual quality, such as marketing efforts (Espeland and Sauder 2007). As Argetsinger (2022) puts it, status therefore works as an "epistemic distorter," disrupting sober moral assessments. Moral judgments should be grounded on the agent's abilities to produce favorable consequences and to comply to moral norms, rather than on an agent's social position. As a consequence, laypeople, policymakers, and ethicists need to be made aware of their status sensitivity in moral judgments, whether it is in terms of agent type or organizational ranking. They have an epistemic responsibility to be cautious and aware of status bias.

Notes

1 Migration background refers to people who either migrated into the country of residence or have at least one parent who previously migrated into the country. Migration background is a highly salient criterion in political and public debates in German-speaking countries, linked to ethnic discrimination in various areas, such as the housing market or hiring.

2 We also do not find a significant effect of male gender framing in a simpler model for the human agent (see M6 in Table A1 in the online supplement), only including vignette characteristics and significant controls (respondent's gender, years of schooling, and vignette order; 11 parameters), which offers sufficient statistical power.

References

Acikgoz, Yalcin, Kristl H. Davison, Maira Compagnone, and M. Laske. 2020. "Justice perceptions of artificial intelligence in selection." *International Journal of Selection and Assessment* 28(4):399–416. https://doi.org/10.1111/ijsa.12306

Ahn, Jungyong, Jungwon Kim, and Yongjun Sung. 2022. "The effect of gender stereotypes on artificial intelligence recommendations." *Journal of Business Research* 141:50–59. https://doi.org/10.1016/j.jbusres.2021.12.007

Akinrinola, Olatunji, Chinwe Chinazo Okoye, Onyeka Chrisanctus Ofodile, and Chinonye Esther Ugochukwu. 2024. "Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability." *GSC Advanced Research and Reviews* 18(3):50–058. https://doi.org/10.30574/gscarr.2024.18.3.0088

Anderson, Cameron and Gavin J. Kilduff. 2009. "The Pursuit of Status in Social Groups." *Current Directions in Psychological Science* 18(5):295–8. https://doi.org/10.1111/j.1467-8721.2009.01655.x

Argetsinger, Henry. 2022. "Blame for me and Not for Thee: Status Sensitivity and Moral Responsibility." *Ethical Theory and Moral Practice* 25(2):265–82. https://doi.org/10.1007/s10677-022-10274-z

Arkoudas, Konstantine and Selmer Bringsjord. 2014. "Philosophical Foundations." Pp. 34–63 in *The Cambridge Handbook of Artificial Intelligence*, edited by K. Frankish and W. M. Ramsey. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139046855.004

Auspurg, Katrin and Thomas Hinz. 2015. *Factorial Survey Experiments.* Thousand Oaks, CA: SAGE. https://doi.org/10.4135/9781483398075

Auspurg, Katrin, Thomas Hinz, and Carsten Sauer. 2017. "Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments." *American Sociological Review* 82(1):179–210. https://doi.org/10.1177/0003122416683393

Baker, Wayne E. and Nathaniel Bulkley. 2014. "Paying It Forward vs. Rewarding Reputation: Mechanisms of Generalized Reciprocity." *Organization Science (Linthicum)* 25(5):1493–510.

Beckert, Jens and Jörg Rössel. 2013. "The Price of Art." *European Societies* 15(2):178–95. https://doi.org/10.1080/14616696.2013.767923

Beer, David. 2016. "The Social Power of Algorithms." *Information, Communication & Society* 20(1):1–13. https://doi.org/10.1080/1369118X.2016.1216147

Benjamin, Beth A. and Joel M. Podolny. 1999. "Status, Quality and Social Order in the California Wine Industry." *Administrative Science Quarterly* 44(3):563–89. https://doi.org/10.2307/2666962

Berger, Joseph, Cecilia L. Ridgeway, and Morris Zelditch. 2002. "Construction of Status and Referential Structures." *Sociological Theory* 20(2):157–179. https://doi.org/10.1111/1467-9558.00157

Bernotat, Jasmin, Friederike Anne Eyssel, and Janik Sachse. 2021. "The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots." *International Journal of Social Robotics* 13(3):477–89. https://doi.org/10.1007/s12369-019-00562-7

Biegert, Thomas, Michael Kühhirt, and Wim van Lancker. 2023. "They Can't All Be Stars: The Matthew Effect, Cumulative Status Bias, and Status Persistence in NBA All-Star Elections." *American Sociological Review* 88(2):189–219. https://doi.org/10.1177/00031224231159139

Bigman, Yochanan E. and Kurt Gray. 2018. "People are Averse to Machines Making Moral Decisions." *Cognition* 181:21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Borau, Sylvie, Tobias Otterbring, Sandra Laporte, and Samuel Fosso Wamba. 2021. "The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI." *Psychology & Marketing* 38(7):1052–1068. https://doi.org/10.1002/mar.21480

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies.* 1st ed. Oxford, England: Oxford University Press.

Bourdieu, Pierre. 2005. "Principles of an Economic Anthropology." Pp. 75–89 in *The Handbook of Economic Sociology*. 2nd ed., edited by N. J. Smelser and R. Swedberg, Princeton, NJ: Princeton University Press.

Bundesärztekammer. 2022. *Medizin ist weiblich*. Accessed October 16, 2024 (https://www.bundesaerztekammer.de/presse/aktuelles/detail/berlin-medizin-ist-weiblich)

Campos-Castillo, Celeste. 2018. "Trust in Health Care: Understanding the Role of Gender and Racial Differences between Patients and Providers." Pp. 151–74 in *Gender, Women's Health Care Concerns and Other Social Factors in Health and Health Care*, edited

by J. J. Kronenfeld. Bingley: Emerald Publishing Limited. https://doi.org/10.1108/S0275-495920180000036009

Cheng, Joey T., Aaron C. Weidman, and Jessica L. Tracy. 2014. "The Assessment of Social Status: A Review of Measures and Experimental Manipulations." Pp. 347–62 in *The Psychology of Social Status*, edited by Joey T. Cheng, Jessica L. Tracy, Cameron Anderson, New York: Springer New York. https://doi.org/10.1007/978-1-4939-0867-7_16

Cugurullo, Federico and Ransford A. Acheampong. 2024. "Fear of AI: An Inquiry into the Adoption of Autonomous Cars in Spite of Fear, and a Theoretical Framework for the Study of Artificial Intelligence Technology Acceptance." *AI & Society* 39:1569–84. https://doi.org/10.1007/s00146-022-01598-6

Da Motta Veiga, Serge P., Maria Figueroa-Armijos, and Brent B. Clark. 2023. "Seeming Ethical Makes You Attractive: Unraveling How Ethical Perceptions of AI in Hiring Impacts Organizational Innovativeness and Attractiveness." *Journal of Business Ethics* 186(1):199–216. https://doi.org/10.1007/s10551-023-05380-6

Darling, Kate. 2015. "'Who's Johnny?'. Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy." In *Robot Ethics 2.0*, edited by P. Lin, G. Bekey, K. Abney, and R. Jenkins. Oxford: Oxford University Press.

Dennett, Daniel C. 1971. "Intentional Systems." *The Journal of Philosophy* 68(4):87–106. https://doi.org/10.2307/2025382

Dietvorst, Berkeley, Joseph P. Simmons, and Cade Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." *Journal of Experimental Psychology: General* 144(1):114–26. https://doi.org/10.1037/xge0000033

Dinzelbacher, Peter. 2002. "Animal Trials. A Multidisciplinary Approach." *Journal of Interdisciplinary History* 32(3):405–21. https://doi.org/10.1162/002219502753364191

Durkheim, Émile. 2009. *Sociology and Philosophy.* London: Routledge. https://doi.org/10.4324/9780203092361

Düval, Sabine and Thomas Hinz. 2020. "Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments." *Field Methods* 32(1):23–37. https://doi.org/10.1177/1525822X19886827

Eagly, Alice H. 2013. *Sex Differences in Social Behavior.* Psychology Press. https://doi.org/10.4324/9780203781906

Elish, M.C. and danah boyd. 2017. "Situating Methods in the Magic of Big Data and AI." *Communication Monographs* 85(1):57–80. https://doi.org/10.1080/03637751.2017.1375130

Enders, Craig K. 2010. *Applied Missing Data Analysis.* New York: Guilford.

Espeland, Wendy Nelson and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *AJS; American Journal of Sociology* 113(1):1–40. https://doi.org/10.1086/517897

Figueroa-Armijos, Maria, Brent B. Clark, and Serge P. da Motta Veiga. 2023. "Ethical Perceptions of AI in Hiring and Organizational Trust: The Role of Performance Expectancy and Social Influence." *Journal of Business Ethics* 186(1):179–97. https://doi.org/10.1007/s10551-022-05166-2.

Fortunati, Leopoldina, Autumn P. Edwards, Anna M. Manganelli, Chad Edwards, and Federico de Luca. 2022. "Special Issue: Gender and Human-Machine Communication." *HMC* 5:75–97. https://doi.org/10.30658/hmc.5.3

Gambetta, Diego. 2011. *Codes of the Underworld. How Criminals Communicate.* Princeton, NJ: Princeton University Press.

Gamez, Patrick, Daniel B. Shank, Carson Arnold, and Mallory North. 2020. "Artificial Virtue: The Machine Question and Perceptions of Moral Character in Artificial Moral Agents." *AI & Society* 35(4):795–809. https://doi.org/10.1007/s00146-020-00977-1

Glikson, Ella and Anita Williams Woolley. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research." *Academy of Management Annals* 14(2):627–60. https://doi.org/10.5465/annals.2018.0057

Haidt, Jonathan and Jonathan Baron. 1996. "Social Roles and the Moral Judgement of Acts and Omissions." *European Journal of Social Psychology* 26(2):201–18. https://doi.org/10.1002/(SICI)1099-0992(199603)26:2<201::AID-EJSP745>3.0.CO;2-J

Hoff, Kevin and Masooda Bashir. 2015. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors* 57(3):407–34. https://doi.org/10.1177/0018720814547570

Horstmann, Aike C., Jonathan Gratch, and Nicole C. Krämer. 2021. "I Just Wanna Blame Somebody, Not Something! Reactions to a Computer Agent Giving Negative Feedback Based on the Instructions of a Person." *International Journal of Human-Computer Studies* 154:102683. https://doi.org/10.1016/j.ijhcs.2021.102683

Jasso, Guillermina. 2020. *Factorial Survey*. SAGE Research Methods Foundations. London: SAGE Publications Ltd.

Joyce, Kelly, Laurel Smith-Doerr, Sharla Alegria, Susan Bell, Taylor Cruz, Steve G. Hoffman, Safiya Umoja Noble, and Benjamin Shestakofsky. 2021. "Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change." *Socius* 7:237802312199958. https://doi.org/10.1177/2378023121999581.

Keil, Susanne and Johanna Dorer. 2019. "Medienproduktion: Journalismus und Geschlecht." Pp. 1–16 in *Handbuch Organisationssoziologie*, edited by Maja Apelt, Ingo Bode, Raimund Hasse, Uli Meyer, Victoria V. Groddeck, Maximiliane Wilkesmann, and Arnold Windeler. Wiesbaden: Springer Fachmedien Wiesbaden.

Kelly, Sage, Sherrie-Anne Kaye, and Oscar Oviedo-Trespalacios. 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

Kieslich, Kimon, Birte Keller, and Christopher Starke. 2022. "Artificial Intelligence Ethics by Design. Evaluating Public Perception on the Importance of Ethical Design Principles of Artificial Intelligence." *Big Data & Society* 9(1):205395172210929. https://doi.org/10.1177/20539517221092956

Kim, Taemie and Pamela Hinds. 2006. "Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction." Pp. 80–85 in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, Univ. of Hertfordshire, Hatfield, UK, September 6–8, 2006. IEEE. https://doi.org/10.1109/ROMAN.2006.314398

Krach, Sören, Frank Hegel, Britta Wrede, Gerhard Sagerer, Ferdinand Binkofski, and Tilo Kircher. 2008. "Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI." *PLoS ONE* 3(7):e2597. https://doi.org/10.1371/journal.pone.0002597

Larkin, Connor, Caitlin Drummond Otten, and Joseph L. Árvai. 2022. "Paging Dr. JARVIS! Will People Accept Advice from Artificial Intelligence for Consequential Risk Management Decisions?" *Journal of Risk Research* 25(4):407–22. https://doi.org/10.1080/13669877.2021.1958047

Latour, Bruno. 2007. *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lavanchy, Maude, Patrick Reichert, Jayanth Narayanan, and Krishna Savani. 2023. "Applicants' Fairness Perceptions of Algorithm-Driven Hiring Procedures." *Journal of Business Ethics* 188(1):125–50. https://doi.org/10.1007/s10551-022-05320-w

Lei, Xin and Pei-Luen Patrick Rau. 2021. "Effect of Relative Status on Responsibility Attributions in Human–Robot Collaboration: Mediating Role of Sense of Responsibility and Moderating Role of Power Distance Orientation." *Computers in Human Behavior* 122:106820. https://doi.org/10.1016/j.chb.2021.106820

Logg, Jennifer M., Julia A. Minson, and Don A. Moore. 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

Lynn, Freda B., Joel M. Podolny, and Lin Tao. 2009. "A Sociological (De)Construction of the Relationship between Status and Quality." *American Journal of Sociology* 115(3):755–804. https://doi.org/10.1086/603537

Mays, Kate K. 2021. "Humanizing Robots? The Influence of Appearance and Status on Social Perceptions of Robots." Dissertation, Boston University. College of Communication. Accessed May 2, 2024 (https://open.bu.edu/handle/2144/41877.)

McLaughlin, G. H.1969. "SMOG Grading: A New Readability Formula." *Journal of Reading* 12(8):639–46.

Nass, Clifford, and Youngme Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56(1):81–103. https://doi.org/10.1111/0022-4537.00153

Obaid, Mohammad, Maha Salem, Micheline Ziadee, Halim Boukaram, Elena Moltchanova, and Majd Sakr. 2016. "Investigating Effects of Professional Status and Ethnicity in Human-Agent Interaction." Pp. 179–86 in *Proceedings of the Fourth International Conference on Human Agent Interaction. HAI '16: The Fourth International Conference on Human Agent Interaction*, Biopolis Singapore, October 4–7, 2016, New York: ACM. https://doi.org/10.1145/2974804.2974813

Onnasch, Linda, and Eileen Roesler. 2019. "Anthropomorphizing Robots: The Effect of Framing in Human-Robot Collaboration." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63(1):1311–5. https://doi.org/10.1177/1071181319631209

Park, Sangwon. 2020. "Multifaceted Trust in Tourism Service Robots." *Annals of Tourism Research* 81:102888. https://doi.org/10.1016/j.annals.2020.102888

Pellandini-Simányi, Léna. 2014. "Bourdieu, Ethics and Symbolic Power." *The Sociological Review* 62(4):651–74. https://doi.org/10.1111/1467-954X.12210

Podolny, Joel M. 1993. "A Status-Based Model of Market Competition." *The American Journal of Sociology* 98(4):829–72. https://doi.org/10.1086/230091

Rammert, Werner. 2016. *Technik - Handeln - Wissen.* Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-11773-3

Ridgeway, Cecilia. 1991."The Social Construction of Status Value: Gender and Other Nominal Characteristics." *Social Forces* 70(2):367. https://doi.org/10.2307/2580244

Ridgeway, Cecilia L. 2014. "Why Status Matters for Inequality." *American Sociological Review* 79(1):1–16. https://doi.org/10.1177/0003122413515997

Rudolph, Udo, Robert Böhm, and Michaela Lummer.2007. "Ein Vorname sagt mehr als 1000 Worte." *Zeitschrift für Sozialpsychologie* 38(1):17–31. https://doi.org/10.1024/0044-3514.38.1.17

Sartori, Laura and Andreas Theodorou. 2022. "A Sociotechnical Perspective for the Future of AI: Narratives, Inequalities, and Human Control." *Ethics and Information Technology* 24(1). https://doi.org/10.1007/s10676-022-09624-3

Sauder, Michael. 2005. "Symbols and Contexts: An Interactionist Approach to the Study of Social Status." *The Sociological Quarterly* 46(2):279–98. https://doi.org/10.1111/j.1533-8525.2005.00013.x

Sauder, Michael, Freda Lynn, and Joel M. Podolny. 2012. "Status: Insights from Organizational Sociology." *Annual Review of Sociology* 38(1):267–83. https://doi.org/10.1146/annurev-soc-071811-145503

Schaap, Gabi, Tibor Bosse, and Paul Hendriks Vettehen. 2024. "The ABC of Algorithmic Aversion: Not Agent, But Benefits and Control Determine the Acceptance of Automated Decision-Making." *AI & Society* 39(4):1947–60. https://doi.org/10.1007/s00146-023-01649-6

Shank, Daniel B., Madison Bowen, Alexander Burns, and Matthew Dew. 2021. "Humans are Perceived As Better, But Weaker, Than Artificial Intelligence: A Comparison of Affective Impressions of Humans, AIs, and Computer Systems in Roles on Teams." *Computers in Human Behavior Reports* 3:100092. https://doi.org/10.1016/j.chbr.2021.100092

Shank, Daniel B., Alexander Burns, Sophia Rodriguez, and Madison Bowen. 2020. Software Program, Bot, or Artificial Intelligence? Affective Sentiments across General Technology Labels. *Current Research in Social Psychology* 28(4):32–41.

Shank, Daniel B., Alyssa DeSanti, and Timothy Maninger. 2019. "When Are Artificial Intelligence versus Human Agents Faulted for Wrongdoing? Moral Attributions after Individual and Joint Decisions." *Information, Communication & Society* 22(5):648–63. https://doi.org/10.1080/1369118X.2019.1568515

Shin, Donghee, and Yong J. Park. 2019. "Role of Fairness, Accountability, and Transparency in Algorithmic Affordance." *Computers in Human Behavior* 98:277–84. https://doi.org/10.1016/j.chb.2019.04.019

Shulner-Tal, Avital, Tsvi Kuflik, and Doron Kliger. 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):1455–82. https://doi.org/10.1080/10447318.2022.2095705

Siegel, Mikey, Cynthia Breazeal, and Michael I. Norton. 2009. "Persuasive Robotics. The influence of robot gender on human behavior." Pp. 2563–8 in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*, St. Louis, MO, USA, IEEE. https://doi.org/10.1109/IROS.2009.5354116

Snijders, Tom A. B. and Roel J. Bosker. 2012. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. 2nd ed. Los Angeles: SAGE.

Spillman, Lyn. 2023. "Morality, Inequality, and the Power of Categories." Pp. 373–85 in *Handbook of the Sociology of Morality*. Vol. 2, edited by S. Hitlin, S. M. Dromi, A. Luft. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-32022-4_26

statista. 2023. Anzahl der Beschäftigten im Bereich Vermittlung von Arbeitskräften in der Schweiz nach Geschlecht von 2011 bis 2021. Accessed October 16, 2024 (https://de.statista.com/statistik/daten/studie/462893/umfrage/beschaeftigte-im-bereich-personalvermittlung- in-der-schweiz-nach-geschlecht/.)

Suchman, Lucy. 2023. "The Uncontroversial 'Thingness' of AI." *Big Data & Society* 10(2). https://doi.org/10.1177/20539517231206794

Sullivan, Yulia, Marc de Bourmont, and Mary Dunaway. 2022. "Appraisals of Harms and Injustice Trigger an Eerie Feeling That Decreases Trust in Artificial Intelligence Systems." *Annals of Operations Research* 308(1-2):525–48. https://doi.org/10.1007/s10479-020-03702-9

Wallander, Lisa. 2009. "25 Years of Factorial Surveys in Sociology: A Review." *Social Science Research* 38(3):505–20. https://doi.org/10.1016/j.ssresearch.2009.03.004

Weber, Max. 2019. *Economy and Society. A New Translation.* Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674240827

Weinberg, Jill, Jeremy Freese, and David McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsource-Recruited Sample." *Sociological Science* 1:292–310. https://doi.org/10.15195/v1.a19

Willemsen, Pascale, Albert Newen, Karolina Prochownik, and Kai Kaspar. 2023. "With Great(er) Power Comes Great(er) Responsibility: An Intercultural Investigation of the Effect of Social Roles on Moral Responsibility Attribution." *Philosophy of Psychology* 1–27. https://doi.org/10.1080/09515089.2023.2213277

Willer, Robb, Reef Youngreen, Lisa Troyer, and Michael J. Lovaglia. 2012. "How Do the Powerful Attain Status? The Roots of Legitimate Power Inequalities." *MDE Managerial and Decision Economics* 33(5-6):355–67. https://doi.org/10.1002/mde.2554

Woolgar, Steve. 1985. "Why not a Sociology of Machines? The Case of Sociology and Artificial Intelligence." *Sociology* 19(4):557–72. https://doi.org/10.1177/0038038585019004005

Xu, Kun, Mo Chen, and Leping You. 2023. "The Hitchhiker's Guide to a Credible and Socially Present Robot: Two Meta-Analyses of the Power of Social Cues in Human–Robot Interaction." *International Journal of Social Robotics* 15(2):269–95. https://doi.org/10.1007/s12369-022-00961-3

Zajko, Mike. 2022. "Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates." *Sociological Compass* 16(3). https://doi.org/10.1111/soc4.12962

Zhang, Qiyuan, Christopher D. Wallbridge, Dylan Marc Jones, and Phil Morgan. (2021). "The Blame Game: Double Standards Apply to Autonomous Vehicle Accidents." Pp. 308–14 in *Advances in Human Aspects of Transportation. AHFE 2021. Lecture Notes in Networks and Systems*. Vol. 270, edited by N. Stanton. Springer. https://doi.org/10.1007/978-3-030-80012-3_36

Acknowledgments: We thank Gabriel Abend, Michael Sauder, the editor of *Sociological Science*, and an anonymous reviewer for their valuable comments. Earlier versions of this article were presented at the Congress of the Academy of Sociology in Bern, Switzerland, and the Conference of the European Sociological Association in Porto, Portugal.

Funding: This study was funded by the Swiss National Science Foundation (grant number 100017_200750/1).

Patrick Schenk: Department of Sociology, University of Lucerne.
E-mail: patrick.schenk@unilu.ch.
Vanessa A. Müller: Department of Sociology, University of Lucerne.
E-mail: vanessa.mueller2@unilu.ch.
Luca Keiser: gfs.bern. E-mail: luca.keiser@gfsbern.ch.