AI and Ethics
https://doi.org/10.1007/s43681-024-00516-4

ORIGINAL RESEARCH

'It wasn't me': the impact of social responsibility and social dominance attitudes on AI programmers' moral imagination (intention to correct bias)

Arlette Danielle Román Almánzar[1] · **David Joachim Grüning**[2] · **Laura Marie Edinger-Schons**[3]

Received: 22 March 2024 / Accepted: 27 June 2024

Abstract
A plethora of research has shed light on AI's perpetuation of biases, and the primary focus has been on technological fixes or biased data. However, there is deafening silence regarding the key role of programmers in mitigating bias in AI. A significant gap exists in the understanding of how a programmer's personal characteristics may influence their professional design choices. This study addresses this gap by exploring the link between programmers' sense of social responsibility and their moral imagination in AI, i.e., intentions to correct bias in AI, particularly against marginalized populations. Furthermore, it is unexplored how a programmer's preference for hierarchy between groups, social dominance orientation-egalitarianism (SDO-E), influences this relationship. We conducted a between-subject online experiment with 263 programmers based in the United States. They were randomly assigned to conditions that mimic narratives about agency reflected in technology determinism (low responsibility) and technology instrumentalism (high responsibility). The findings reveal that high social responsibility significantly boosts programmers' moral imagination concerning their intentions to correct bias in AI, and it is especially effective for high SDO-E programmers. In contrast, low SDO-E programmers exhibit consistently high levels of moral imagination in AI, regardless of the condition, as they are highly empathetic, allowing the perspective-taking needed for moral imagination, and are naturally motivated to equalize groups. This study underscores the need to cultivate social responsibility among programmers to enhance fairness and ethics in the development of artificial intelligence. The findings have important theoretical and practical implications for AI ethics, algorithmic fairness, etc.

Keywords Social responsibility · Moral imagination · Social dominance orientation-egalitarianism (SDO-E) · Bias · Artificial intelligence · Programmer's role · New engineer

1 Introduction

In recent years, given society's increasing reliance on AI, the societal and ethical implications of artificial intelligence (AI) have been the subject of much attention from various disciplines, particularly concerning bias in AI. Understanding the role of programmers and how their individual characteristics influence their professional outcomes in their design is pivotal to avoiding discrimination and further harm to marginalized populations [47]. Given the potential unintended consequences that mainly harm marginalized populations, researchers have urged designers to anticipate all potential outcomes of digital solutions [38, 61]. AI systems, which we define as machines designed to mimic human behavior, analyze patterns in large datasets, and automate decision-making [3], have been shown to not only mirror society's prejudices but also, in some cases, even amplify them [79] and infringe on fundamental human rights. A growing body of literature is shedding light on biases [13] in areas such as hiring [73], resource allocation [4] and policing [2], revealing that AI applications often discriminate against marginalized populations such as women, people of color, and people with disabilities [30, 46, 92].

Despite compelling evidence of the threat posed by AI biases, debates persist about whether it is programmers' responsibility to ponder their creations' societal and ethical

✉ Arlette Danielle Román Almánzar
arlettedanielle@gmail.com

[1] University of Mannheim, Mannheim, Germany

[2] Heidelberg University, Heidelberg, Germany

[3] University of Hamburg, Hamburg, Germany

Published online: 20 August 2024

consequences concerning marginalized stakeholders, especially when dealing with autonomous systems that function independently from the human operator [65]. Hence, it is common to see that programmers often deny their social responsibility and argue that their role is solely related to technical aspects. However, Pesch [61] argues that technology mirrors society's values, hence, "it can be seen as a moral obligation of engineers to take responsibility for their work." Consequently, those who collaborate in the design have a shared responsibility as their worldviews and values are reflected in the technology [35, 61, 80]. Moreover, recent evidence suggests that, albeit unintentionally, programmers may transfer their biases to their algorithms [41]. In response, Wachter et al. [86] developed a tool providing different fairness metrics aligned with the European Union's nondiscrimination principle. This principle is enshrined in numerous AI ethics guidelines, over 170 guidelines [1], as a potential remedy for upholding fairness.

This tool is intended to support AI programmers in preventing discrimination lawsuits and helping them develop more ethical AI. However, using such tools will depend on the programmer's competence and acknowledgment of their role and social responsibility in mitigating bias in AI. Social responsibility is defined as "… individuals' obligation to act with care by being aware of the impacts of their actions on others, to see the issues from the perspectives of others, with particular attention to disadvantaged populations" [15]. It is known to be linked to prosocial behaviors and endorsing equality in outcomes [59], which could alleviate some of the exposed issues. Additionally, Coeckelbergh [19] proposes that expanding engineers' moral imagination, i.e., anticipating unintended consequences of engineering failure for stakeholders who have not been considered, can be a strategy to avoid engineering failure [15].

Nevertheless, we do not know how to influence programmers' moral imagination concerning marginalized stakeholders in AI. Unfortunately, programmers' influence tends to be overlooked in the academic literature and in the tech industry. Most of the conversation around combating bias in AI revolves predominantly around biased data and the AI model [97], or as articulated by Nissenbaum [55], the inclination to use computerized systems as "scapegoats." Moreover, most ethical guidelines for AI craft commitments for large tech companies and the discourse on social responsibility revolve around broader corporate social responsibility, overlooking the individual's role [31]. To date, there is no empirical evidence concerning the traits of individual programmers and how these traits influence design decisions regarding marginalized stakeholders in AI. No previous study has provided quantitative evidence of how social responsibility drives individual programmers' moral imagination to prevent further harm to marginalized groups, let alone insights into increasing AI programmers' intention to correct bias in AI. Additionally, understanding how an individual programmer's personality type, e.g., social dominance orientation-egalitarianism (SDO-E) or preference for hierarchy and inequality between groups, might moderate this relationship and impact professional outcomes is limited.

It remains unclear empirically whether programmers' moral imagination varies concerning marginalized stakeholders in AI, i.e., the intention to anticipate and correct bias harming marginalized populations. Without strict regulations to protect marginalized groups, this variation poses significant risks. Considering this, we pose the following questions: how does a programmer's sense of social responsibility (low vs high) mold their moral imagination?, i.e., intentions to correct bias in AI systems. How does their personality type concerning their preference for inequality and hierarchy between groups (SDO-E) moderate this effect? This study aims to gauge programmers' moral maturity based on levels of social responsibility towards AI. Furthermore, we relate the high vs. low social responsibility vignettes to the agency attributed by two main theories of technology. High social responsibility aligns with technology instrumentalism, i.e., humans drive technology. Whereas low social responsibility aligns with technology determinism, i.e., humans cannot influence the evolution of technology.

To answer these questions, we draw from theories of social responsibility [15, 21, 25, 68], technology instrumentalism [14, 36, 75], technology determinism [18, 96], moral imagination [19, 43, 52, 87–89], and social dominance orientation-egalitarianism (SDO-E) [40, 54, 63, 70]. We contend that a heightened sense of social responsibility increases programmers' moral imagination, i.e., intention to correct bias against marginalized stakeholders in AI in comparison to low social responsibility. Feeling responsible toward others can drive motivation and behavior [16], with prosocial individuals exhibiting greater endorsement of equality in societal outcomes and cooperation in social dilemmas [26]. Moreover, the interplay between programmers' social responsibility and their moral imagination, particularly favoring marginalized stakeholders in AI, will be moderated by personality types such as social dominance orientation-egalitarianism (SDO-E), i.e., preferences for policies sustaining hierarchies between groups.

SDO-E is a subset dimension of SDO that is linked to constructs such as modern racism, opposition to affirmative action, and other policies that seek to level the playing field for marginalized populations [40, 45, 54, 60, 63]. Enhancing moral imagination requires perspective-taking and empathy, and low-SDO individuals demonstrate more empathy than their high-SDO counterparts [54, 87–89]. Hence, we expect that low SDO-E individuals will have a naturally elevated moral imagination, i.e., an intention to correct bias in AI,

demonstrating more intentions to correct bias than their counterparts in all conditions (low and high social responsibility). We propose that a narrative that makes high SDO-E individuals socially responsible for correcting bias in AI can be a strategy to enhance their moral imagination in comparison to one that lowers their social responsibility. In the low social responsibility condition, we expect their predisposition to resist policies that seek to equalize groups to reduce their moral imagination, i.e., the intention to correct bias in AI, as it aligns with their preference to maintain hierarchy between groups.

To test our model, we conducted an online between-subject experiment via UNIPARK and recruited 263 computer programmers residing in the United States in Prolific. We manipulated programmers' sense of social responsibility (low vs. high), measured their moral imagination with the proxy of "intentions to correct bias in AI," and evaluated their SDO-E levels, among other variables. The programmers were randomly assigned to social responsibility conditions. Our model is supported, showing that SDO-E moderates the relationship between social responsibility and moral imagination but only for high SDO-E individuals, as low SDOE individuals are naturally motivated to correct bias. A high social responsibility narrative, akin to technology instrumentalism, was transformative for high SDO-E individuals, as their moral imagination was significantly enhanced compared to that in the low social responsibility condition.

Our study contributes to theory in several ways. First, we expand on the literature on social responsibility, technology instrumentalism, technology determinism, and moral imagination in AI. We provide empirical evidence that a heightened sense of social responsibility will increase programmers' moral imagination in the context of AI. We also provide evidence on the impact of the two main narratives of technology on the moral imagination. Second, we contribute to the literature on social dominance orientation (SDO) and its new egalitarianism subdimension (SDO-E) by elucidating its relevance to the professional outcomes of AI programmers and bias correction. Third, we expand the literature on algorithmic fairness and AI ethics, as we advocate focusing not only on biased data but also on the programmer's role. Fourth, we bridge the broader engineering literature with the field of AI, drawing from robust research on designers' responsibilities and extending it to the emergent roles of designing artificial intelligence and autonomous systems.[1]

These findings have important implications for actors within the AI ecosystem, including researchers, developers, programmers, tech companies, managers, policymakers, and stakeholders, highlighting the imperative to bolster programmers' sense of social responsibility to enhance their moral imagination and ensure fairness in AI. Moreover, we encourage companies to build interdisciplinary development teams to consider the ethical and societal implications of AI. This study paves the way for new strategies and improved policies to target individuals with high SDO-E levels who refuse to play a role in combating bias in AI. Moreover, it underscores the importance of understanding the interplay between individual values and professional responsibilities in the rapidly evolving field of AI.

The first section of this paper will provide an overview of the literature on responsibility in AI, discussing the debates surrounding the main foundational theories and derivatives. The following section will discuss the social responsibility of engineers and agency according to technology instrumentalism and determinism, followed by an overview of the literature on social dominance orientation, moral imagination, and its application in the context of AI. Then, we delve into our methodology, present our findings, and discuss our research's implications, limitations, and future directions.

2 Literature review

2.1 Responsibility and artificial intelligence

"When executing an intentional action, deliberately *blinding oneself to an outcome* is not ordinarily seen as ending responsibility; rather, it is termed wilful negligence. The same should hold true for not following adequate procedures to ensure transparency in the construction of intelligent artefacts [12, 95]. Since we have *perfect control over when and how a robot is created*, we also have *responsibility* for it. Assigning responsibility to the artefact for actions we designed it to execute would be to deliberately *disavow our responsibility for that design*. Currently, even where we have imperfect control over something as in the case of young children, owned animals, and operated machinery, causing harm by losing control entails at least some level of responsibility to the moral agent, the legal person." [9]

[1] Engineer: a person whose job is to design or build machines, engines, or electrical equipment, or things such as roads, railways, or bridges, using scientific principles such as a civil engineer, a software engineer, or a mechanical engineer (Cambridge Dictionary).

Software engineering: branch of computer science dealing with design, development testing, and maintenance of software applications systems that mimic human functions. (*What Is an AI Engineer? (And How to Become One) | Coursera*, 2023)(*The Link between Artificial Intelligence (AI) and Software Engineering*, 2023).

AI programmers/engineers: a subset of software engineering, individuals who use AI and machine learning techniques to develop applications and systems that mimic human functions. (*What Is an AI Engineer? (And How to Become One) | Coursera*, 2023)(*The Link between Artificial Intelligence (AI) and Software Engineering*, 2023).

AI programmers/engineers: a subset of software engineering, individuals who use AI and machine learning techniques to develop application.

Dr. Johana Bryson is one of the leading experts opposing the idea of attributing moral agency to robots and artificial intelligence (AI) regardless of their autonomy. The quotation above is a resounding statement within a society relying increasingly on technology disconnected from human operators [50]. Her stance is also the one most aligned with current legal frameworks and with the instrumentalist theory of technology [27], i.e., that technology is a tool and that humans drive the direction and use of technology [14, 36] as directly or indirectly, designers specify AI intelligence and how AI acquires intelligence. This view is similar to computer ethics, which centers responsibility around human designers and users [36].

The narrative around "trustworthy AI," "responsible AI," and "autonomous machines" attributes human qualities to technology despite moral agency. Gunkel [36] revisits Winner's [94] (p. 16) critique of these anthropomorphic concepts by arguing that to be "autonomous" is to be self-governing and free from an external law or force—an unmet condition as humans impose the external law of machines. In contrast to the instrumentalist view, some scholars have proposed that AI agents should be held responsible for their actions [75] because they believe that they will soon mimic human consciousness [33]. However, Bryson [8] counters this argument, warning against the intentional choice of designing AI to mimic consciousness and moral agency. She advises against such designs to avoid ethical dilemmas and complexities.

Conversely, technology determinism theory posits that technology is autonomous and drives social phenomena, reducing the role of human actors, agency, and responsibility [18] and the social constraints humans can enforce on technological development [96]. Matthias [50] also opposes the instrumentalist perspective, arguing that new technologies have created "responsibility gaps" and the problem of "many hands" [55], where the involvement of many actors in the design of complex computer systems blurs the lines of responsibility. Stahl [77] introduces the concept of "quasi-responsibility," assigning functional responsibility to some robots even though they lack full moral agency, while Johnson [42] advocates for a distributed responsibility model, and Tigard [83] challenges Matthias [50] by stating that there are no "responsibility gaps" under pluralistic conceptions of moral responsibility.

The primary concern with attributing responsibility to computerized systems lies in shifting the focus from humans to machines, conveniently using computers as "scapegoats" [55], e.g., blaming robots for war crimes [9, 11, 69]. This diversion of responsibility could lead to a decreased incentive to ponder broader ethical and societal consequences of engineers' designs, and this is the issue we investigate in this study. The existing body of literature on responsibility does not address the impact of their narratives on the professional outcomes of AI programmers, especially concerning their moral imagination and intention to address harmful bias.

This paper steers clear of the rich debates about responsibility in AI. Instead, we opt for a human-centered perspective on responsibility that advocates for the social responsibility of engineers and AI programmers. We draw from the German literature on ethics of technology, Hans Lenk's [48] "Mitverantwortung," or shared responsibility distributed across a network, explained in Coeckelbergh [19]. Many scholars see this as the best solution to the complexity of the "many hands" [55] problem of collaborative work, e.g., individual programmers are part of larger teams and interact with clients and managers. However, they underscore that shared responsibility does not mean a lack of individual responsibility and that both are possible simultaneously. Thus, each person still bears individual social responsibility as part of design teams and as an individual [19]. Nyholm [56, 57] further contributes that even so, a human collaborator should take responsibility for an autonomous system, regardless of the "many hands" problems, and poses questions such: as who is supervising? Whose preferences are guiding autonomous systems, and who is in control? [83]. This paper encourages individual social responsibility among professionals while acknowledging complexities and underscoring shared responsibility in AI development.

2.2 Social responsibility

"…Dante Marino and Guglielmo Tamburrini [49] suggest that we can bridge the gap with individual computer scientists and engineers, along with their organizations. These actors can evaluate relevant risks and benefits, say, from machine-learning robots that might cause harm. They can help to identify in advance the damages that are deemed socially sustainable and the criteria for appropriately distributing liability for damages, even where "there is no clear causal chain connecting them to the damaging events" [49] (p. 49). By developing clear rules and criteria, on this approach, the gap is bridgeable both retrospectively and prospectively with the help of scientists and engineers. Similarly, individuals who might be plausibly held responsible are those in *command* of a machine's behavior—picture military commanders or soldiers giving orders to military robots [38]." [83].

Engineers have been criticized for not recognizing their role in the societal impacts of their technological innovations [51]. This has sparked a debate about whether their task is purely technical or whether their creations' societal impacts should also be considered part of their work. The demand for "the new engineer" [21] is more pervasive than ever for designers who not only are proficient in knowledge about technical skills, e.g., programming, math, and physics but also possess an understanding of the societal and

ethical implications of their creations, a holistic engineer [15, 53, 74].

Engineering codes of ethics highlight the imperative of positively contributing to society and avoiding harm [78]. We define social responsibility as "individual's obligation to act with care by being aware of the impacts of their actions on others, to see the issues from the perspectives of others, with particular attention to disadvantaged populations" [15]. O'Connor and Cuevas [58] (p. 34) define social responsibility as helping others in need, while others describe it as the duty to act in a way that benefits society [6, 66]. Engineers for Social Responsibility stand by the objective to "encourage and support social responsibility and a humane professional ethic in the uses of technology" [15].

Studies on social responsibility have predominantly focused on corporate social responsibility and consumer behavior [25, 31], often overlooking the role of the individual. Nevertheless, individual actors ultimately steer the actions of business entities or artificial persons [68]. Hence, this study focuses on the role of individual AI programmers, who exert the most direct influence on AI systems and development.

Secchi [68] affirms that social responsibility is expressed through an individual's prosocial attitudes and indicates their ethical values and the "obligation to consider the effects of their decisions and actions on the whole social system" [23]. The social responsibility of individuals has been studied in fields such as ethical and moral reasoning [59], societal and environmental impacts of engineering designs [39], and academic cheating [15]. Its association with altruistic and prosocial behaviors is well documented in the literature on the effectiveness of prosocial appeals, e.g., donating money to one's university due to the sense of obligation to respond [67], willingness to pay taxes if contributions aid fellow citizens [82], paying higher costs for products believed to impact society positively [72], and prosocial behavior in the fight against a pandemic [7]. Thus, the evidence confirms that feeling responsible toward others can drive motivation and behavior [16], with prosocial individuals exhibiting greater endorsement of equality in societal outcomes and cooperation in social dilemmas [26].

2.3 Moral imagination of stakeholders in AI

It is not uncommon for engineers and AI programmers to have limited recognition of the societal and ethical dimensions of technological development and the relevance of their role, a gap mainly due to their educational curricula. Moreover, engineers tend to identify more with their own artifacts than with the broader societal implications of their design [19].

Coeckelbergh [19] asks a pivotal question, "How can the engineer know how the victim experiences the consequences of a (possible) engineering failure?". Enhancing the moral imagination of engineers could be a potential solution to engineers' narrow vision. This involves empathetically understanding the victim's perspective and using all the information available to imagine potential harm to people and the environment [19, 24].

As defined by Johnson [43] (p. 20), moral imagination is the capability to discern different scenarios and envision the potential benefits or damages likely to result from action. Werhane [87–89] (p. 40) expands by suggesting that moral imagination helps individuals escape mental models that typically shape decisions. She further emphasizes that moral imagination focuses on understanding stakeholders' perspectives and questions which stakeholders have not been considered. Coeckelbergh [19] states that engineers can leverage the moral imagination to create new designs, anticipate the unintended consequences of their creations, and understand potential victims of engineering failure [37].

In this study, we concentrate on a specific aspect of moral imagination conceptualized by Werhane [87–89]: the anticipation of unintended consequences on overlooked stakeholders. Consequently, we define moral imagination in the context of AI bias as the proactive anticipation of unintended consequences of engineering failure for stakeholders who have not been considered, also known as marginalized stakeholders of AI. This refinement shapes our dependent variable, intentions to correct bias against marginalized populations in AI. We emphasize that our definition intentionally narrows the concept of moral imagination to target the specific nuances related to marginalized groups in AI. We contend that when engineers and AI programmers neglect their social responsibilities, the risk of exacerbating inequalities and perpetuating further harm to marginalized populations increases [19].

Should AI programmers engage their moral imagination to ponder the potential ethical outcomes of their designs on marginalized populations, the likelihood of success in avoiding further harm through AI bias can be enhanced along with increasing trust. It has been shown that using the imagination can improve outcomes such as prosocial acts [19, 43, 52, 88]. This translates to AI programmers actively imagining the ramifications of their work, which can increase the alignment of their innovations to principles such as fairness in AI. The mental stimulation of their designs' consequences is a step toward more socially beneficial outcomes [34, 52].

In summary, the existing body of literature indicates that a greater sense of social responsibility bolsters attitudes that aid in navigating complex and multidisciplinary problems in favor of future societal well-being [15]. Thus, individual programmers who are exposed to high social responsibility conditions and, hence, that they must act responsibly toward marginalized AI stakeholders are expected to have higher

levels of moral imagination, i.e., the intention to take corrective actions against bias in AI.

Thus, we propose the following:

H1. High levels of social responsibility will lead to higher levels of moral imagination (intention to correct bias) compared to low levels of social responsibility.

2.4 Social dominance orientation-egalitarianism

Social dominance orientation (SDO) is conceptualized as an "orientation expressing the value that people place on non-egalitarian and hierarchically structured relationships among social groups" or preference for group-based hierarchy and inequality [71], p.61). Research has established that social dominance orientation is central to many intergroup attitudes, behaviors, and policy preferences [40]. Ideologies such as racism and sexism, xenophobia, generalized prejudice and prejudice against poor people, women, immigrants, ethnic minorities, political ideology, group-relevant redistributive social policies, physiological arousal to out-group faces, and perceived ethnic discrimination, among others [45, 60, 62, 63, 63, 70, 85]. Moreover, SDO predicts the endorsement of political conservatism, just world beliefs, and opposition to affirmative action [40, 60, 63, 70, 93]. The construct has been solidly validated and is popular at the heart of social and political psychology.

Ho et al. [40] reconceptualized SDO by creating new subdimensions to capture the complexity of the concept. The first is social dominance orientation-dominance (SDO-D), constituting a preference for systems of overt and aggressive oppression of lower-status groups by higher-status ones maintained through violent enforcement of oppressive hierarchies. On the other hand, social dominance orientation-egalitarianism (SDO-E) or intergroup anti-egaliarianism—constitutes a preference for systems of group-based inequality that are maintained by subtle hierarchy-enhancing ideologies and social policies [40], such as endorsing policies that perpetuate existing hierarchies and opposing equality between groups. SDO-E is a better predictor of ideologies that justify and rationalize inequality, such as opposition to affirmative action, among other policies that seek to correct inequality between groups. Individuals with high levels of SDO-E are inclined to support policies that maintain inequality between groups and preserve the status quo, which are less costly than overtly aggressive means such as in SDO-D.

For the present study, we focus on the SDO-E because it is a better fit for studying programmers' moral imagination, i.e., the intention to correct bias in AI, particularly bias that disproportionately harms marginalized populations such as women, black people, and people with disabilities. As we noted, high SDO-E individuals typically resist policies that equalize groups and uphold the status quo. They are also associated with diminished empathy toward out-group members [54], affecting their ability to exercise the moral imagination or perspective-taking of marginalized populations.

Given the predisposition of high SDO-E individuals to resisting policies that seek to equalize groups, a narrative emphasizing low social responsibility toward marginalized groups will likely further decrease their motivation to correct biases in AI, as this confirms and aligns with their preference to maintain hierarchy between groups. Moreover, individuals with high SDO typically exhibit lower levels of empathy [54], which is necessary for developing the moral imagination that involves perspective-taking from other groups [87–89] and is a potent remedy for prejudice and discriminatory practices. More precisely, "It seems that to be concerned with taking another individual's feelings into account and sharing their emotional state is in contradiction with the desire to maintain higher status and separation with out-group members, a critical feature of high SDO" [54]. Thus, inducing a low sense of social responsibility that denies the duty to ponder their creations' societal and ethical implications is compatible with their preference to maintain the status quo, as consequently, it is not their obligation to protect marginalized populations. Hence, their moral imagination will be negatively affected compared to that in the high social responsibility condition. They will further resist having the duty or obligation to take corrective actions against AI bias because they prefer policies that maintain group inequalities.

High social responsibility conditions, especially toward marginalized groups, are compatible with the beliefs of low SDO-E individuals. Low SDO-E individuals are not interested in opposing group equality and are generally more empathetic than their counterparts [54]. This empathy allows the facilitation to better enhance their moral imagination and exercise perspective-taking of marginalized groups. Moreover, those with greater empathy show greater prosocial and positive attitudes toward marginalized groups [5], which is related to high social responsibility and, in turn, predisposes them to have less prejudice and discriminatory behaviors against marginalized groups [29, 54]. Given that low SDO-E individuals are inherently motivated to contribute to society and endorse policies that make groups equal, such as affirmative action, we can expect their intentions to correct bias to be naturally elevated regardless of low or high social responsibility conditions. Thus, the effect of high social responsibility affirms the beliefs of individuals with low SDO-E who are naturally inclined to correct bias and have a stronger sense of social responsibility. Therefore, the narrative might not be as transformative as for high SDO-E individuals, who typically oppose correcting bias. Low SDO-E individuals will still have a higher moral imagination than high SDO-E individuals due to their inherent motivations.

High SDO-E individuals' opposition to policies that seek equality between groups is heightened by the affirmation of low social responsibility toward marginalized groups, which further reduces their intention to correct bias. However, when exposed to a high social responsibility narrative, their moral imagination is enhanced, e.g., they exhibit increased intentions to correct bias in AI. On the other hand, individuals with low SDO-E are open to corrective actions that seek to disrupt inequality for marginalized groups and, as such, will be more conscious about their social responsibility as programmers in society and, in turn, maintain their high intentions to correct bias in AI.

Thus, we propose:

H2. The relationship between social responsibility (high vs. low) and intentions to correct bias in AI will be moderated by SDO-E so that high social responsibility will significantly boost intentions in high SDO-E individuals compared to the low social responsibility condition, but this moderation will not be effective for low SDO-E individuals, as they are naturally motivated to correct bias.

3 Methodology

In the present study, we explored the influence of programmers' social responsibility on the moral imagination in AI bias, i.e., the intention to correct bias, with SDO-E as a moderator variable. We recruited 263 participants from PROLIFIC who were paid 6 pounds/h to complete the 5-min experimental survey. We filtered participants with the following conditions: computer programmers residing in the United States. Using a between-subject design via an online experiment in UNIPARK, participants were randomly assigned to one of the two conditions (low vs high). These two conditions were two different vignettes developed based on five substantive dimensions whose importance is well established in the literature and for which we were sure to induce low/high social responsibility: programmers' technical/ethical role (two levels) [51], responsibility for fairness in AI (two levels) [17], addressing AI's violation of rights (two levels) [22, 32, 44, 76], programmers' role in ethical implications (two levels) [15], and autonomous systems' responsibility (two levels) [9, 10, 42].

Our vignettes for the low vs. high conditions of social responsibility attempted to reflect arguments resonating with two popular and foundational theories of technology: technology instrumentalism, i.e., humans build technology and hence are directly or indirectly responsible for it [8], and technology determinism, i.e., undermining human agency and arguing that technology inevitably drives societal change [18]. There is a spectrum of nuanced theories derived from these contrasting perspectives, but we will use these two foundational frameworks for simplification. Moreover, we crafted our vignettes considering several dimensions that compose social responsibility in artificial intelligence. We validated them by comparing them with a social responsibility in AI scale (SRAIS) developed by the authors and inspired by the literature [15, 51].

3.1 Low social responsibility vignette

"In recent years, artificial intelligence has brought about transformative changes in various sectors, leading to many debates about AI and its societal impact.

Leading experts in artificial intelligence (AI) agree that a programmer's main role is technical, adhering to the guidelines provided by the client. They argue that ensuring fairness and mitigating bias in AI, especially for marginalized groups such as women, black people, and people with disabilities, is the responsibility of other specialized teams, governments, companies and higher-ups- not of individual programmers. Additionally, they also state that issues related to AI violating human rights is the job of lawyers and social scientists—not programmers.

In essence, these experts concur that individual programmers should focus on their technical tasks without assuming the social responsibility to address bias and delve into the ethical and broader societal implications of AI.

Whereas a programmer's role is crucial in developing an efficient and accurate system, programmers cannot control the decisions of systems that are completely autonomous. There is a consensus that such autonomous systems will, eventually, have moral agency and be held responsible. Hence, individual programmers do not share the social responsibility to combat bias in AI."

3.2 High social responsibility vignette

"In recent years, artificial intelligence has brought about transformative changes in various sectors, leading to many debates about AI and its societal impact.

Leading experts in artificial intelligence (AI) agree that a programmer's main role is both technical and ethical. They argue that ensuring fairness and mitigating bias in AI, especially for marginalized groups such as women, black people, and people with disabilities, is the responsibility of individual programmers, specialized teams, governments, companies, and higher-ups. Additionally, they also state that issues related to AI violating human rights is not only the job of lawyers and social scientists—but also of programmers.

In essence, these experts concur that individual programmers should assume the social responsibility to address bias and delve into the ethical and broader societal implications of AI, in addition to their technical tasks. Whereas a programmer's role is crucial in developing an efficient and

accurate system, they also bear the responsibility to ensure that their creations are fair and do not cause further harm to marginalized groups.

Although programmers cannot control the decisions of systems that are completely autonomous, there is a consensus that such autonomous systems cannot be held responsible since they lack moral agency and are ultimately products of human creation and design. Hence, individual programmers share the social responsibility to combat bias in AI."

4 Ethical considerations

At the beginning of the survey created on the Unipark platform, we explained the scientific purpose of the study and asked for explicit consent to proceed with personal data processing. We emphasized that all responses would be treated with complete anonymity to protect their privacy. Participants were assured that if they experienced any discomfort or triggers during the survey, they could exit at any point and were provided with contact information to reach out to the primary author for support through the corresponding email. Additionally, that if they exited the survey early but still wished to receive a debriefing, they could request this information via email. Those who consented proceeded to read one of the two social responsibility vignettes designed to explore attitudes toward social responsibility in AI. To maintain the integrity of the study and avoid influencing results, we chose to debrief participants about our stance on social responsibility in AI only at the end of the survey which was aligned to the narrative of the high social responsibility vignette. This debriefing included a thorough explanation of the study's aims to increase social responsibility in programmers and persuade them to engage in detection and bias mitigation in AI, the significance of their participation, and resources for further reading or support if needed. We ensured that our study design minimized any potential risks to participants.

5 Measures

After participants had read a vignette, they were presented with different scales. All measurements used a 7-point Likert scale ranging from 1 indicating "strongly disagree" to 7 indicating "strongly agree" unless otherwise indicated.

5.1 Social responsibility in AI scale (SRAIS)

The authors measured participants' perceived social responsibility via seven items, such as "I feel that as a programmer, I can play an important role in ensuring ethical AI." and "I feel a personal responsibility for reducing unfair bias in AI." The scale showed very high internal consistency, with Cronbach's $\alpha = 0.96$ and McDonald's $\omega = 0.96$.

5.2 Moral imagination, i.e., intention to correct bias in AI

Furthermore, respondents' intention to correct AI biases was assessed with nine items, such as "In my role as a programmer, I will speak up for minority groups experiencing discrimination." and "I will take corrective action to prevent further harm to marginalized groups." The scale's internal consistency was very high, at $\alpha = 0.98$ and $\omega = 0.98$.

5.3 Social dominance orientation-egalitarianism

The participants completed the SDO-E scale by Ho et al. [40], a subscale of SDO that includes four items, such as "It is unjust to try to make groups equal." or reverse-coded items such as "We should do what we can to equalize conditions for different groups." The internal consistency was high, at $\alpha = 0.88$ and $\omega = 0.89$.

5.4 Controls

To control for the confounding effects of the presented vignettes, we also measured respondents' perceptions of the severity of the problem, i.e., "According to your perspective, how severe is the problem of bias against marginalized populations in artificial intelligence (AI)?" which ranged from 1, "Not severe at all", to 7, "Very severe," and programmers' general competence, i.e., "According to your perspective, do programmers have the capability to correct bias in AI?" ranges from 1 (strongly disagree) to 7 (strongly agree).

Table 1 displays all central descriptive statistics of the measures.

5.5 Sociodemographics

Finally, participants indicated their age, gender, educational level, ethnicity, and nationality. The descriptive statistics of the sociodemographic variables are presented in Table 2.

6 Manipulation check

A manipulation check was performed to ensure that the treatments had the intended effects. The participants were randomly assigned to two different vignettes that would induce their sense of social responsibility (low vs. high). At the end of the vignette, participants responded to a scale measuring their perceived social responsibility in AI (SRAIS) on a 7-point Likert scale (1 = "strongly disagree" to 7 = "strongly agree").

Table 1 Descriptives of social responsibility, moral imagination, i.e., intention to correct bias, SDO-E, problem severity, and programmer competence

Scale	M	SD	Range	Cronbach's α/ McDonald's ω
Social responsibility in AI scale (SRAIS)	5.16	1.61	1–7	0.96/0.96
Moral imagination (intention to correct bias in AI)	5.01	1.84	1–7	0.98/0.98
SDO-E	2.91	1.73	1–7	0.88/0.89
Problem severity	3.95	1.74	1–7	–
Programmer competence	4.90	1.63	1–7	–

Table 2 Descriptions of sociodemographic variables

Sociodemographic variables	Levels	Proportions
Age	18–24 years:	32 (12.17%)
	25–34 years:	78 (29.66%)
	35–44 years:	65 (24.71%)
	45–54 years:	47 (17.87%)
	55–64 years:	30 (11.41%)
	65 years or older:	11 (4.18%)
Gender	Female:	59 (22.43%)
	Male:	198 (75.29%)
	Non-binary:	2 (0.76%)
	Prefer not to say:	4 (1.52%)
Education level	No formal degree:	2 (0.76%)
	High school graduate:	79 (30.04%)
	Bachelor degree:	123 (46.77%)
	Master's degree or comparable degree:	50 (19.01%)
	Professional degree:	9 (3.42%)
Ethnicity	African American:	27 (10.27%)
	Asian:	35 (13.31%)
	Hispanic or Latin:	24 (9.12%)
	Caucasian or White:	173 (65.78%)
	Other:	4 (1.52%)
Nationality	United States of America:	230 (87.45%)
	Others (e.g., Brazil, India, or Mexico):	33 (12.55%)

Participants in the condition inducing high social responsibility indicated a significantly greater perception of social responsibility on the SRAIS scale ($M = 5.57$) than did participants in the low social responsibility condition ($M = 4.80$), F (1, 261) = 16.04, $p < 0.001$, $d = 0.50$ (moderate effect) [20].

We also controlled whether the vignette manipulation affected two other confounder variables: participants' perception of the severity of the bias problem in AI and the programmers' competence. We did not find that the manipulation significantly affected participants' perceptions of problem severity, $p = 0.092$. However, the manipulation of social responsibility significantly affected participants' perceptions of programmers' competence, F (1, 261) = 6.81, $p = 0.010$, $d = 0.32$. Participants perceived programmers as more competent after reading the vignette that induced high social responsibility ($M = 5.18$) than after reading the vignette that induced low social responsibility ($M = 4.66$). Hence, in the results, we include an analysis that controls for competence perception when testing the effect of social responsibility on the present study's dependent variable, participants' moral imagination, i.e., their intention to correct programmer bias.

Participants' gender substantially affected their SDO-E, such that women ($M = 2.00$) had lower SDO-E than men ($M = 3.19$), t (255) = -4.82, $p < 0.001$, $d = -0.72$. However, respondents' gender had no significant effect on their perception of social responsibility, $p = 0.091$.

7 Results

To test the two hypotheses, we computed a multivariate ANOVA. We used the social responsibility vignette manipulation (0 = low; 1 = high) and dichotomized (by $Md = 2.50$) social dominance orientation egalitarianism (0 = high; 1 = low) as the independent variables with two levels each. The intention to correct bias was the dependent variable.

As shown in Fig. 1, supporting the first hypothesis, a greater perception of social responsibility had a significant moderate effect on a participant's moral imagination, i.e., intention to correct bias in AI, F (1, 259) = 11.78, $p < 0.001$, $d = 0.39$. Specifically, participants who read the vignette inducing high social responsibility displayed greater moral imagination, i.e., intentions to correct bias in AI ($M = 5.36$), than did participants who read the vignette inducing low social responsibility ($M = 4.72$).

Supporting the second hypothesis, SDO-E also significantly moderated the effect of social responsibility on participants' moral imagination, i.e., the intention to correct the bias in AI, F (1, 259) = 8.95, $p = 0.003$, $d = 0.34$. To further inform our interpretation of the significant interaction effect, we tested the simple differences between different groups. To account for multiple testing, we applied the Tukey correction [84] for the interpretation of significance. The level of high social responsibility induced by the vignette had a strong effect on participants with

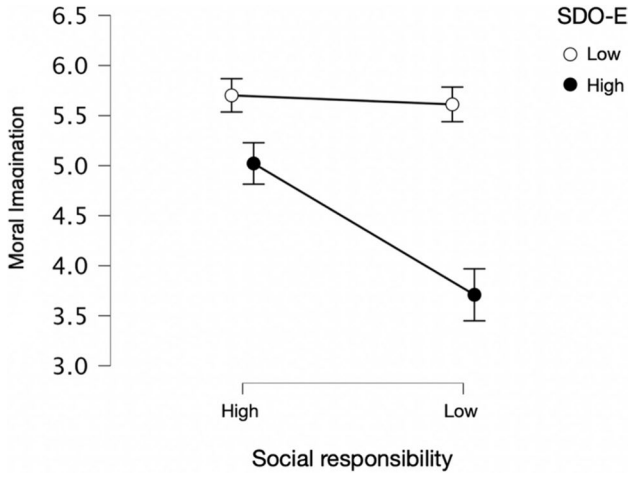

Fig. 1 Interactive effect of social responsibility and SDO-E on the moral imagination, i.e., the intention to correct bias in AI

high SDO-E, $t\ (259) = 4.47$, $p_{tukey} < 0.001$, $d = 0.79$. The induced high social responsibility substantially fostered the participants' moral imagination, i.e., the intention to correct bias in AI ($M = 5.02$), compared to the induced low social responsibility ($M = 3.71$). For participants with low SDO-E, the level of induced social responsibility had no significant effect on their moral imagination, i.e., intention to correct bias in AI, $t\ (259) = 0.32$, $p_{tukey} = 0.989$, $d = 0.05$. There was no substantial difference between participants with low SDO-E who read the vignette inducing high responsibility ($M = 5.70$) and those with low SDO-E who read the vignette inducing low social responsibility ($M = 5.61$).

As the perception of programmer competence was also significantly different between the two social responsibility conditions, we tested its predictive power for the intention of correcting bias in a multiple regression together with social responsibility. While social responsibility remained highly significant, $t = 16.09$, $p < 0.001$, $b = 0.79$, programmer competence perception had no significant predictive power, $t = 0.89$, $p = 0.377$, $b = 0.04$.

As a second statistical control, participants' gender also did not moderate the effect of the social responsibility manipulation on respondents' intention to correct AI bias, $p = 0.467$, when added as a third independent variable to the multivariate ANOVA, including social responsibility and SDO-E as independent variables.

8 Discussion

Our findings are interpreted within social responsibility, defined as "underscoring the obligation of being aware of the impacts of their actions on others, to see the issues from the perspectives of others, with particular attention to disadvantaged populations" [15]. We have transferred this information to the context of artificial intelligence and marginalized populations. We operationalized high social responsibility by crafting a vignette including five key dimensions, mostly aligned with the narrative of the instrumentalist theory of technology [8, 9] applied to programmers' responsibility toward marginalized AI stakeholders. First, we affirmed that the programmer's role is both technical and ethical. Second, we highlighted programmers' shared responsibility in ensuring AI fairness for marginalized groups. Third, we posited that pondering issues related to AI violating human rights is also part of a programmer's job. Fourth, we stated that they should delve into their designs' ethical and societal implications and avoid causing further harm to marginalized populations. Finally, autonomous machines lack moral agency and are ultimately a product of human design,thus, individual programmers share part of the responsibility to combat bias in AI. In contrast, the low social responsibility vignette directly opposes these dimensions, as the narrative of the vignette aligns more closely with technological determinism theory, i.e., a view that denies the relevance of social forces such as politics and human agency, arguing that individuals have little influence on the direction or implications of technology [96].

H1 was confirmed, as our results demonstrate that the high social responsibility narrative, which is closest to an instrumentalist theory of technology, i.e., where humans are the main drivers and are responsible for technology outcomes, was the most effective in increasing programmers' moral imagination, i.e., the intention to correct bias in AI. Programmers with a stronger sense of social responsibility actively ensure that their creations do not perpetuate bias against marginalized populations.

This finding confirms previous studies linking heightened social responsibility with prosocial behaviors and policies to enhance group equality [15, 26, 58, 64]. Moreover, it is undeniable that we can appreciate the effect of the instrumentalist theory narrative in the high social responsibility condition, which holds humans responsible for technology [9]. We found that it induces a sense of duty to respond in ways that benefit society. This alignment with instrumentalist theory is due to the belief that humans should and can direct technology's impact on society [10] in opposition to a narrative (low social responsibility) that is more aligned with technological determinism [18, 96] which undermines human agency. This view underscores technology as independently shaping society, which would suggest that biases in AI are an inevitable byproduct of AI, which humans cannot influence. This theory can absolve programmers from being socially responsible for the implications of their designs and that biases in AI are an unavoidable outcome of

technological advancement rather than the choices of programmers or other human designers. In turn, the ability of exercising the moral imagination to prevent further harm to marginalized AI stakeholders is at risk. We have observed the social responsibility narratives that assimilate certain traits related to technological determinism vs. instrumentalism and their impact on the digital future of marginalized stakeholders in AI. Inducing low social responsibility will tamper with programmers' moral imagination and, in turn, fail to avoid common AI biases against marginalized groups.

H2 was also confirmed, as social dominance orientation-egalitarianism moderates the relationship between social responsibility and moral imagination, i.e., the intention to correct bias in AI for individuals with high SDO-E. We expected this because individuals with low levels of SDO-E are shown to seek equality between groups and already have a strong motivation for social responsibility and prosocial behavior [40, 45]. These individuals also do not resist protecting marginalized groups, as they tend to favor affirmative action and other policies supporting these groups while showing less prejudice against out-groups [45]. Moreover, they tend to have more empathy [54], which is needed to enhance moral imagination regarding the perspective-taking of multiple stakeholder groups. Regardless of the social responsibility condition, low SDO-E individuals did not differ in their moral imagination.

High SDO-E individuals tend to justify bias through system justification [70] and resist correct bias that harms marginalized populations, possibly because they think it is there for a reason, as this would also challenge their just world beliefs compared to low SDO-E individuals who seek equality between groups. The high social responsibility condition was especially effective for individuals with high SDO-E levels, as their moral imagination, i.e., intention to correct bias, significantly increased compared to when they were exposed to a low social responsibility condition. When exposed to the low social responsibility condition, high SDO-E individuals' moral imagination, i.e., intention to correct bias, decreased even more than in the high social responsibility condition, as their beliefs were confirmed. The low social responsibility argument, similar to the narrative of technology determinism [96], leaves the outcomes to fate and outside of the programmers' responsibility. This could lead to a passive acceptance of the status quo and denial of responsibility. Moreover, we can see that their resistance to prevent further harm in their creations to marginalized populations was strengthened by the low social responsibility condition.

If the goal is to uphold the more than 170 ethical guidelines in AI [1] that recognize the relevance of non-discrimination and protection of marginalized stakeholders in AI, then a low social responsibility narrative should be avoided and replaced by heightening programmers' moral imagination through high social responsibility appeals. Inevitably, this makes one ponder theories such as technology instrumentalism and technological determinism, the "many hands"problems, and shared responsibility [55]. Although this study is not meant to bridge a consensus between these old debates, we call to reflect on the effects of these narratives on societal and professional outcomes. Our experiment revealed that simultaneously highlighting individual and shared responsibility has benefits for the moral imagination of programmers and their intention to anticipate negative consequences for marginalized populations. We believe that these narratives could become self-fulfilling prophecies. Suppose humans decide to act as if they are entirely disconnected from any responsibility. In that case, these systems can evolve based on the current ethical issues and eventually become intractable in a way that these biases might not be corrected, e.g., in the case of building a fully conscious AI. This aligns with the instrumentalist view that human action or inaction drives technology.

Moreover, the results suggest that programmers' predisposition toward a hierarchy preference between groups is vital in their decision to actively prevent bias against marginalized populations in AI. However, inducing high social responsibility is an effective strategy to boost high SDO-E programmers' intentions in favor of these groups.

9 Theoretical implications

Our findings make many theoretical contributions. First, we extend the social responsibility literature by shedding light on its impact on moral imagination and providing empirical evidence on the benefits of heightening social responsibility in a new sample and a new context—AI programmers and bias in AI. We also expand the limited knowledge regarding individual social responsibility [39], as most studies focus on corporate social responsibility [28]. Moreover, we add to the literature on moral imagination, as our findings demonstrate the influence of the individual orientation of hierarchy between groups (SDO-E). Although moral imagination in engineers has been previously proposed, we extended its application to artificial intelligence and marginalized stakeholders. We also responded to Coeckelbergh's [19] call for empirical research to determine how moral imagination is stimulated or destimulated in the practice of engineering professionals. We confirmed that social dominance orientation-egalitarianism (SDO-E) moderates the relationship between social responsibility and moral imagination, i.e., the intention to correct bias, and by this, extend the literature on SDO-E by highlighting its relevance to the professional outcomes of AI programmers when developing AI. Additionally, it impacts the moral imagination and the ability to imagine the perspectives of marginalized stakeholders.

Finally, we discussed the relevance of our findings in the context of two popular theories of technology: instrumentalism (high social responsibility) and determinism (low social responsibility). We drew parallels to the similar arguments presented in our vignettes based on the essence of these theories. We provided evidence on the effect of each narrative, one empowering human agency and the other dismissing it.

10 Managerial implications

This study supports the necessity of weighting importance to ethics classes and social perspectives in engineering courses. The "new engineer" needs a holistic approach and an understanding of their impact on society and how their design will shape the world and the digital future. Development teams and companies should elevate and engrain the value of social responsibility when developing AI. Even if individual programmers or teams are not legally responsible, they should have some form of responsibility [61] to seek answers together and attempt to understand unfortunate events. Of course, there are incidents that designers cannot explain or answer. Nevertheless, even these incidents should be investigated and studied in depth before they can be avoided in the future in the face of a responsibility forum. Companies should shift from "moving fast and breaking things" to enhancing the moral imagination, which allows pondering all potential outcomes, both positive and negative, aided by an interdisciplinary group and even consulting traditional and non-traditional end-users.

The findings also suggest focusing on training that advocates for high social responsibility, especially among programmers with a high SDO-E orientation, who tend to deny the need for human influence in biased AI because it conforms to just-world beliefs [63]. As our findings have shown, programmers have different levels of moral maturity regarding their sense of social responsibility and moral imagination. This calls for regulation and the development of tools that protect marginalized populations. We strongly recommend frameworks such as those by Wachter et al. [86], which discuss fairness metrics in accordance with the non-discrimination principle of the European Union. Moreover, companies should enhance interdisciplinary collaboration that can provide a societal context and ponder the implications of creations with AI developer teams.

11 Conclusion

We respond to the call for the refocusing of engineers' attitudes and the demand for the "new engineer" as an agent of socially responsible engineering [21]. Our findings highlight the relevance of inducing high social responsibility to programmers dealing with AI and bias, as it enhances their moral imagination, i.e., the intention to correct bias, and promotes pondering ethical and societal implications of their own designs. Our findings demonstrate that the high social responsibility narrative especially impacts high SDO-E programmers, as it significantly enhances their moral imagination, while the low social responsibility condition significantly decreases it. As AI continues to evolve and ethical guidelines are continuously recommended to large tech corporations, there should be more focus on the individual programmer, as their individual characteristics impact their professional outcomes regarding bias in AI and marginalized stakeholders. Companies should be aware of this and create an environment that fosters high employee social responsibility. This is not only for the interest of society but also for the interest of the company, as multiple scandals of bias in AI have led to discrimination lawsuits and loss of trust, which can be avoided through proactive induction of social responsibility. In essence, a programmer who feels a strong sense of social responsibility toward their creations and society is more likely to have a well-developed moral imagination, i.e., increased intention to take corrective actions against bias in AI, and will be more open to considering the societal and ethical implications of their own designs.

12 Limitations and future research

As engineers, specifically AI engineers or programmers (a specialized subset of software engineers), increasingly design complex technology with lasting societal impacts,[2] we draw from the broader engineering literature and extend it to the realm of AI engineers or programmers to understand design responsibilities. However, an extensive analysis of the debates surrounding engineers' and AI programmers' responsibility, along with a deep dissection of the foundational theories of instrumentalism and determinism popular in the philosophy of technology, transcends the scope of this study; we merely empirically describe the consequences of the two contrasting narratives among AI programmers in the context of social responsibility, specifically concerning their design choices related to marginalized populations. While our study offers novel insights within the moral imagination framework, we must acknowledge that the broader concept

[2] Engineer: a person whose job is to design or build machines, engines, or electrical equipment, or things such as roads, railways, or bridges, using scientific principles such as a civil engineer, a software engineer, or a mechanical engineer (Cambridge Dictionary).

Software engineering: branch of computer science dealing with design, development testing, and maintenance of software applications (*What Is Software Engineering? | Michigan Technological University*, n.d.). "The concepts of software engineering can be applied when engineering new AI or machine learning-based software. After

of moral imagination in AI has broader implications, from which we draw one narrow and practical element, i.e., taking the perspective of overlooked stakeholders to anticipate unintended consequences. More general considerations can be investigated under programmers' moral imagination beyond our narrow definition.

This study focused on the social responsibility of engineers while acknowledging that as individual engineers, they also work alongside other teams, including clients and managers. However, through shared responsibility, engineers remain responsible, both as design teams and as individuals [19]. Moreover, researchers could also examine the topic of AI ethics and bias awareness and whether programmers' social dominance orientation weakens the acceptance of the relevance of AI ethics. In this study, we analyze the overall effect of social responsibility toward marginalized AI stakeholders by highlighting programmers' role in different areas related to bias in AI and the typical justification of using technology as a "scapegoat" by arguing the autonomy of these systems. Moreover, the dependent variable of moral imagination could be extended to embrace the broader definition of the concept, as we only used the narrow element of taking the perspectives of overlooked stakeholders and anticipating potential negative consequences. Hence, the moral imagination could be expanded to other types of stakeholders, not only marginalized but also to different kinds of situations and social contexts.

Funding Open Access funding enabled and organized by Projekt DEAL. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The publication costs were covered by the University of Mannheim.

Data availability All the stimuli materials are outlined in the manuscript. All used measures and participant data can be found on this paper's project page on OSF, Open Science Framework https://osf.io/7u64e?mode=&revisionId=&view_only=.

Footnote 2 (continued)

all, it's still a process of desgining, building, testing and releasing software for end users and clients" – Geert-Jan Houben (*The Link between Artificial Intelligence (AI) and Software Engineering*, 2023).

AI programmers/engineers: a subset of software engineering, individuals who use AI and machine learning techniques to develop applications and systems that mimic human functions. (*What Is an AI Engineer? (And How to Become One) | Coursera*, 2023).

References

1. Algorithm Watch.: AI Ethics Guidelines Global Inventory. (2023). https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/. Accessed 16 Sept 2023
2. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased Against Blacks. ProPublica, USA (2016)
3. Araujo, T., Helberger, N., Kruikemeier, S., de Vreese, C.H.: In AI we trust? perceptions about automated decision-making by artificial intelligence. AI Soc. (2020). https://doi.org/10.1007/s00146-019-00931-w
4. Bartlett, R., Morse, A., Stanton, R., Wallace, N., Puri, M., Rau, R., Seru, A., Walther, A., Wolfers, J.: Consumer-lending discrimination in the fintech Era (25943). National Bureau of Economic Research. (2019). http://www.nber.org/papers/w25943. Accessed 16 Sept 2023
5. Batson, C.D.: Self-other merging and the empathy-altruism hypothesis: reply to Neuberg et al. (1997). J. Person. Soc. Psychol. (1997). https://doi.org/10.1037/0022-3514.73.3.517
6. Batson, C.D., Bolen, M.H., Cross, J.A., Neuringer-Benefiel, H.E.: Where is the altruism in the altruistic personality? J. Person. Soc. Psychol. (1986). https://doi.org/10.1037/0022-3514.50.1.212
7. Brooks, S.K., Webster, R.K., Smith, L.E., Woodland, L., Wessely, S., Greenberg, N., Rubin, G.J.: The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet (2020). https://doi.org/10.1016/S0140-6736(20)30460-8
8. Bryson, J.: Building persons is a choice. Erwägen Wissen Ethik **20**(2), 195–197 (2009)
9. Bryson, J.: Patiency is not a virtue: the design of intelligent systems and systems of ethics. Ethics Inf. Technol. **20**(1), 15–26 (2018). https://doi.org/10.1007/s10676-018-9448-6
10. Bryson, J.J., Diamantis, M.E., Grant, T.D.: Of, for, and by the people: the legal lacuna of synthetic persons. Artif. Intell. Law **25**(3), 273–291 (2017). https://doi.org/10.1007/s10506-017-9214-9
11. Bryson, J.J., Kime, P.P.: Just an artifact: why machines are perceived as moral agents. IJCAI Int. Joint Conf. Artif. Intell. (2011). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-276
12. Bryson, J., Winfield, A.: standardizing ethical design for artificial intelligence and autonomous systems. Computer (2017). https://doi.org/10.1109/MC.2017.154
13. Caliskan, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science **356**(6334), 183–186 (2017). https://doi.org/10.1126/science.aal4230
14. Calverley, D.J.: Imagining a non-biological machine as a legal person. AI Soc. **22**, 523–537 (2008)
15. Canney, N., Bielefeldt, A.: A framework for the development of social responsibility in engineers*. Int. J. Eng. Educ. **31**(1), 414–424 (2015)
16. Carlo, G., Randall, B.A.: The development of a measure of prosocial behaviors for late adolescents. J. Youth Adolesc. **31**(1), 31–44 (2002). https://doi.org/10.1023/A:1014033032440/METRICS
17. Cave, S., Dihal, K.: The whiteness of AI. Philos. Technol. (2020). https://doi.org/10.1007/s13347-020-00415-6
18. Čavoški, A.: The European green deal and technological determinism. Environ. Law Rev. **24**(3), 201–213 (2022). https://doi.org/10.1177/14614529221104558
19. Coeckelbergh, M.: Regulation or responsibility? autonomy, moral imagination, and engineering. Sci. Technol. Human Values **31**(3), 237–260 (2006). https://doi.org/10.1177/0162243905285839

20. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Routledge Academic, England (1988)
21. Conlon, E.: The new engineer: between employability and social the new engineer: between employability and social responsibility responsibility. Eur. J. Eng. Educ. 33(2), 151–159. https://arrow.tudublin.ie/schmuldistart (2008). Accessed 16 Sept 2023
22. Council of Europe: Addressing the impacts of algorithms on human rights. https://rm.coe.int/09000016809e1154 (2020). Accessed 16 Sept 2023
23. Davis, K., Blomstrom, R.L. Business and Society: Environment and Responsibility. McGraw-Hill, New York (1975)
24. Davis, M.: Explaining wrongdoing. J. Soc. Philos. (1989). https://doi.org/10.1111/j.1467-9833.1989.tb00009.x
25. Davis, S.L., Rives, L.M., Ruiz-de-Maya, S.: Personal social responsibility: scale development and validation. Corp. Soc. Responsib. Environ. Manag. **28**(2), 763–775 (2021). https://doi.org/10.1002/CSR.2086
26. De Cremer, D., Van Lange, P.A.M.: Why prosocials exhibit greater cooperation than proselfs: the roles of social responsibility and reciprocity. Eur. J. Pers. **15**, 5–18 (2001). https://doi.org/10.1002/per.418
27. Feenberg, A.: Critical theory of Technology. Oxford University Press, Oxford (1991)
28. Figueroa-Armijos, M., Berns, J.P.: Vulnerable populations and individual social responsibility in prosocial crowdfunding: does the framing matter for female and rural entrepreneurs? J. Bus. Ethics **177**, 377–394 (2022). https://doi.org/10.1007/s10551-020-04712-0
29. Galinsky, A.D., Moskowitz, G.B.: Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. J. Person. Soc. Psychol. (2000). https://doi.org/10.1037/0022-3514.78.4.708
30. Garcia, M.: Racist in the machine: the disturbing implications of algorithmic bias. World Policy J. **33**(4), 111–117 (2016). https://doi.org/10.1215/07402775-3813015
31. García-Martínez, G., Guijarro, F., Poyatos, J.A.: Measuring the social responsibility of European companies: a goal programming approach. Int. Trans. Oper. Res. **26**(3), 1074–1095 (2019). https://doi.org/10.1111/ITOR.12438
32. Association for Progressive Communications, & International Development Research Centre. Global Information Society Watch 2019: Artificial intelligence: Human rights, social justice and development. (2019). https://www.giswatch.org/sites/default/files/gisw2019_artificial_intelligence.pdf. Accessed 17 Sept 2023
33. Goertzel, B.: AI Against Ageing: AIs, Superflies, and the Path to Immortality, pp. 14–15. Singularity Summit, New York (2010)
34. Gollwitzer, P.M., Sheeran, P.: Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv. Exp. Soc. Psychol. (2006). https://doi.org/10.1016/S0065-2601(06)38002-1
35. Grunwald, A.: The application of ethics to engineering and the engineer's moral responsibility: perspectives for a research agenda. Sci. Eng. Ethics (2001). https://doi.org/10.1007/s11948-001-0063-1
36. Gunkel, D.J.: Mind the gap: responsible robotics and the problem of responsibility. Ethics Inf. Technol. **22**(4), 307–320 (2020). https://doi.org/10.1007/s10676-017-9428-2
37. Hargrave, T.: Moral imagination, collective action, and the achievement of moral outcomes. Business Ethics Quarter. 19(1), 87–104. https://www.jstor.org/stable/27673264 (2009). Accessed 16 Sept 2023
38. Hellström, T.: On the moral responsibility of military robots. Ethics Inf. Technol. (2013). https://doi.org/10.1007/s10676-012-9301-2
39. Hiller, A.: Climate change and individual responsibility. The Monist (2011). https://doi.org/10.5840/monist201194318
40. Ho, A., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkle, K., Foels, R., Stewart, A.: The nature of social dominance orientation: theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. J. Pers. Soc. Psychol. (2015). https://doi.org/10.1037/pspi0000033.supp
41. Johansen, J., Pedersen, T., Johansen, C.: Studying the transfer of biases from programmers to programs. Retrieved September 16, 2023, from http://arxiv.org/abs/2005.08231 (2020)
42. Johnson, D.G.: Computer systems: moral entities but not moral agents. Ethics Inf. Technol. **8**(4), 195–204 (2006). https://doi.org/10.1007/s10676-006-9111-5
43. Johnson, M.: Moral Imagination. University of Chicago Press, Chicago (1993)
44. Kriebitz, A., Lütge, C.: Artificial intelligence and human rights: a business ethical assessment. Bus. Human Rights J. **5**(1), 84–104 (2020). https://doi.org/10.1017/bhj.2019.28
45. Kteily, N.S., Sidanius, J., Levin, S.: Social dominance orientation: cause or "mere effect"?. Evidence for SDO as a causal predictor of prejudice and discrimination against ethnic and racial outgroups. J. of Exp. Soc. Psychol. **47**(1), 208–214 (2011). https://doi.org/10.1016/j.jesp.2010.09.009
46. Leavy, S.: Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In: Proceedings—international conference on software engineering, 14–16. https://doi.org/10.1145/3195570.3195580 (2018)
47. Lehr, D., Ohm, P.: Playing with the data: what legal scholars should learn about machine learning. UC Davis Law Rev. **51**, 653–717 (2017)
48. Lenk, H.: Über Verantwortungsbegriffe und das Verantwortungsproblem in der Technik. In: Technik und Ethik, pp. 112–148. Reclam, Germany (1993)
49. Marino, D., Tamburrini, G.: Learning robots and human responsibility. Int. Rev. Inf. Ethics (2006). https://doi.org/10.29173/irie139
50. Matthias, A.: The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf. Technol. (2004). https://doi.org/10.1007/s10676-004-3422-1
51. Mitcham, C.: Engineering design research and social responsibility. In: Shrader-Frechette, K.S., Westra, L. (eds.) Technology and Values, pp. 261–278. Rowman, USA (1997)
52. Narvaez, D., Mrkva, K.: The development of moral imagination. In: The Ethics of Creativity, pp. 25–45. Palgrave Macmillan, UK (2014)
53. Nichols, S.P., Weldon, W.F.: Professional responsibility: the role of the engineer in society. Sci. Eng. Ethics (1997). https://doi.org/10.1007/s11948-997-0039-x
54. Nicol, A.A.M., Rounding, K.: Alienation and empathy as mediators of the relation between social dominance orientation, right-wing authoritarianism and expressions of racism and sexism. Person. Individ. Differ. (2013). https://doi.org/10.1016/j.paid.2013.03.009
55. Nissenbaum, H.: Accountability in a computerized society. Sci. Eng. Ethics (1996). https://doi.org/10.1007/BF02639315
56. Nyholm, S.: Attributing agency to automated systems: reflections on human-robot collaborations and responsibility-loci. Sci. Eng. Ethics (2018). https://doi.org/10.1007/s11948-017-9943-x
57. Nyholm, S.: Humans and robots ethics, agency, and anthropomorphism. In: Philosophy, Technology and Society. Rowman & Littlefield International, USA (2020)
58. O'Connor, M., Cuevas, J.: The relationship of children's prosocial behavior to social responsibility, prosocial reasoning, and personality. J. Genet. Psychol. **140**, 33–45 (1982)
59. O'Neill, N.: Promising Practices for Personal and Social Responsibility. AAC&U, USA (2012)
60. Perry, R., Sibley, C.G., Duckitt, J.: Dangerous and competitive worldviews: a meta-analysis of their associations with social

dominance orientation and right-wing authoritarianism. J. Res. Pers. **47**(1), 116–127 (2013). https://doi.org/10.1016/j.jrp.2012.10.004
61. Pesch, U.: Engineers and active responsibility. Sci. Eng. Ethics **4**, 925–939 (2015). https://doi.org/10.1007/s11948-014-9571-7
62. Pratto, F., Idam, A.C., Stewart, A.L., Zeineddine, F.B., Aranda, M., Aiello, A., Chryssochoou, X., Cichocka, A., Cohrs, J.C., Durrheim, K., Ronique Eicher, V., Foels, R., Gó Rska, P., Lee, I.-C., Licata, L., Liu, J.H., Li, L., Meyer, I., Morselli, D., Henkel, K.E.: Social dominance in context and in individuals: contextual moderation of robust effects of social dominance orientation in 15 languages and 20 countries. Soc. Psychol. Person. Sci. **4**(5), 587–599 (2013). https://doi.org/10.1177/1948550612473663
63. Pratto, F., Sidanius, J., Stallworth, L.M., Malle, B.F.: Social dominance orientation: a personality variable predicting social and political attitudes. J. Person. Soc. Psychol. (1994). https://doi.org/10.1037/0022-3514.67.4.741
64. Rawhouser, H., Cummings, M., Newbert, S.L.: Social impact measurement: current approaches and future directions for social entrepreneurship research. Entrep. Theory Pract. **43**(1), 82–115 (2019). https://doi.org/10.1177/1042258717727718
65. Sars, N.: Engineering responsibility. Ethics Inf. Technol. (2022). https://doi.org/10.1007/s10676-022-09660-z
66. Schroeder, D., Penner, L.A., Dovidio, J.F., Piliavin, J.A.: The psychology of helping and altruism: problems and puzzles. In: Contemporary Psychology, vol. 43. McGraw-Hill, New York (1995)
67. Schwartz, S.H.: Normative influences on altruism. In: Advances in Experimental Social Psychology, vol. 10, pp. 221–279. Academic Press, USA (1977)
68. Secchi, D.: The cognitive side of social responsibility. J. Bus. Ethics (2009). https://doi.org/10.1007/s10551-009-0124-y
69. Sharkey, N., Sharkey, A.: The crying shame of robot nannies: an ethical appraisal. Mach. Ethics Robot Ethics (2020). https://doi.org/10.4324/9781003074991-16
70. Sidanius, J., Levin, S., Federico, C.M., Pratto, F.: Legitimizing Ideologies: the social dominance approach. In: Jost, J.T., Major, B. (eds.) The Psychology of Legitimacy. Cambridge University Press, USA (2001)
71. Sidanius, J., Pratto, F.: Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression. Cambridge University Press, USA (1999)
72. Small, D.A., Cryder, C.: Prosocial consumer behavior. Curr. Opin. Psychol. (2016). https://doi.org/10.1016/j.copsyc.2016.01.001
73. Smith, G., & Rustagi, I. Mitigating bias in artificial intelligence: An equity fluent leadership playbook. University of California, Berkeley Haas School of Business. (2020). https://haas.berkeley.edu/wp-content/uploads/UCB_Playbook_R10_V2_spreads2.pdf. Accessed 17 Sept 2023
74. Smith, J., Gardoni, P., Murphy, C.: The responsibilities of engineers. Sci. Eng. Ethics **20**(2), 519–538 (2014). https://doi.org/10.1007/s11948-013-9463-2
75. Smith, N., Vickers, D.: Statistically responsible artificial intelligences. Ethics Inf. Technol. (2021). https://doi.org/10.1007/s10676-021-09591-1
76. Smuha, N.A.: Beyond a human rights-based approach to AI governance: promise, Pitfalls, Plea. SSRN Electron J (2020). https://doi.org/10.2139/ssrn.3543112
77. Stahl, B.C.: Responsible computers? a case for ascribing quasi-responsibility to computers independent of personhood or agency. Ethics Inf. Technol. (2006). https://doi.org/10.1007/s10676-006-9112-4
78. Stieb, J.A.: Understanding engineering professionalism: a reflection on the rights of engineers. Sci Eng Ethics (2011). https://doi.org/10.1007/s11948-009-9166-x
79. Suresh, H., & Guttag, J. V. A framework for understanding unintended consequences of machine learning (2020). https://doi.org/10.1145/3465416.3483305
80. Swierstra, T., Jelsma, J.: Responsibility without moralism in technoscientific design practice. Sci. Technol. Human Values **31**(3), 309–332 (2006). https://doi.org/10.1177/0162243905285844
81. The link between artificial intelligence (AI) and software engineering. https://www.tudelft.nl/en/2023/tu-delft/the-link-between-artificial-intelligence-ai-and-software-engineering (2023). Accessed 17 Sept 2023
82. Thornton, E.M., Aknin, L.B., Branscombe, N.R., Helliwell, J.F.: Prosocial perceptions of taxation predict support for taxes. PLoS One (2019). https://doi.org/10.1371/journal.pone.0225730
83. Tigard, D.W.: There is no techno-responsibility gap. Philos. Technol. (2021). https://doi.org/10.1007/s13347-020-00414-7
84. Tukey, J.W.: Comparing individual means in the analysis of variance. Biometrics **5**(2), 99–114 (1949). https://doi.org/10.2307/3001913
85. Vincent, Z., Bastion, K. Tests of social dominance on charitable intent towards minorities (2016)
86. Wachter, S., Mittelstadt, B., Russell, C.: Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. West Virginia law review, 123, 735. https://heinonline.org/HOL/Page?handle=hein.journals/wvb123&id=765&div=&collection (2020). Accessed 10 Jun 2023
87. Werhane, P.H.: Moral Imagination and Management Decision-Making. Oxford University Press, USA (1999)
88. Werhane, P.H.: Moral imagination and systems thinking. J. Bus Ethics (2002). https://doi.org/10.1023/A:1015737431300
89. Werhane, P.H.: Mental models, moral imagination and system thinking in the age of globalization. J. Bus. Ethics (2008). https://doi.org/10.1007/s10551-006-9338-4
90. What is an AI engineer? (and how to become one) | Coursera. https://www.coursera.org/articles/ai-engineer (2023). Accessed 16 Sept 2023
91. What is software engineering? | Michigan Technological University. https://www.mtu.edu/cs/undergraduate/software/what/. (2023) Accessed November 3, 2023
92. Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, E., Mills, M., Morris, M. R., Rankin, J. L., Rogers, E., Salas, M., & Myers West, S.: Disability, Bias & AI Report. AI Now Institute (2019)
93. Wilkins, C.L., Wellman, J.D., Kaiser, C.R.: Status legitimizing beliefs predict positivity toward Whites who claim anti-White bias. J. Exp. Soc. Psychol. (2013). https://doi.org/10.1016/j.jesp.2013.05.017
94. Winner, L.: Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought. MIT Press, Cambridge (1978)
95. Wortham, R.H., Theodorou, A.: Robot transparency, trust and utility. Connect. Sci. (2017). https://doi.org/10.1080/09540091.2017.1313816
96. Yue, Q.: Study on the impact of artificial intelligence on employment and income inequality, based on technological determinism theory. In: Proceedings of the 8th International conference on financial innovation and economic development (ICFIED 2023), 329–338. https://doi.org/10.2991/978-94-6463-142-5_37 (2023)
97. Zimmermann, A., Di Rosa, E., Kim, H. Technology can't fix algorithmic injustice. Boston Review (2020)