

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Understanding Perceptions of Algorithmic Bias Through the Risk Perception and Attitude Framework

Hannah Overbye-Thompson, Erick A. Garcia, Xinyi Zhang & Laurent H. Wang

To cite this article: Hannah Overbye-Thompson, Erick A. Garcia, Xinyi Zhang & Laurent H. Wang (2026) Understanding Perceptions of Algorithmic Bias Through the Risk Perception and Attitude Framework, International Journal of Human–Computer Interaction, 42:8, 5854-5873, DOI: 10.1080/10447318.2025.2546661

To link to this article: https://doi.org/10.1080/10447318.2025.2546661

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 27 Aug 2025.

Submit your article to this journal

Article views: 4275

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hihc20

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2026, VOL. 42, NO. 8, 5854–5873
https://doi.org/10.1080/10447318.2025.2546661

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Understanding Perceptions of Algorithmic Bias Through the Risk Perception and Attitude Framework

Hannah Overbye-Thompson[a], Erick A. Garcia[a], Xinyi Zhang[b] and Laurent H. Wang[a]

[a]Department of Communication, University of California, Santa Barbara, CA, USA; [b]Department of Human Communication Studies, California State University, Fullerton, CA, USA

ABSTRACT

Drawing from risk perception attitude framework, this mixed-methods study investigated individuals' risk and efficacy perceptions of algorithmic bias. We examined algorithmic bias in both organizational algorithms (e.g., healthcare, hiring) and individual-use algorithms (e.g., search engines and facial filters). Quantitative analysis showed that participants perceived organizational algorithms as riskier than individual-use algorithms, while feeling less capable of addressing bias in organizational algorithms. Qualitative analysis revealed four key risk dimensions (i.e., mental health, privacy and exploitation, fairness and discrimination, and segmentation and polarization) and three types of efficacy beliefs (i.e., feelings of powerlessness, strategic consumption, and collective responsibility for mitigating bias). While quantitative data showed no significant racial/ethnic differences in risk perceptions, qualitative findings suggested that people of color may adopt avoidant attitudes toward algorithmic systems, potentially missing beneficial opportunities. Gender differences also emerged, with male participants reporting higher efficacy in managing individual-use algorithmic bias and perceiving greater risk in organizational algorithms.

KEYWORDS

Risk perception and attitudes framework; algorithmic bias; mixed methods; efficacy

1. Introduction

In an era dominated by algorithmic decision-making, from curating social media feeds to determining loan eligibility, the issue of algorithmic bias has become increasingly prevalent; facial recognition systems have led to wrongful arrests due to higher error rates for people with darker skin tones (Johnson, 2022), healthcare algorithms have been discovered systematically favoring White patients over Black patients (Obermeyer et al., 2019) and hiring algorithms continue to perpetuate discrimination based on gender, race, and other personal characteristics (Chen, 2023). These instances represent merely a fraction of the documented cases of algorithmic bias in contemporary systems (cf., Bhuiyan, 2024; Eubanks, 2018; Grother et al., 2019; O'Neil, 2016; Resende et al., 2024). Algorithmic bias generally includes two overlapping categories: *technical bias* emerges from the way algorithms are designed and implemented, leading to unintended disparities in how different users are treated (Overbye-Thompson et al., 2024), while *moral bias* refers to when algorithms produce discriminatory outcomes for specific groups or individuals, whether deliberately or not (Friedman & Nissenbaum, 1996). These categories classify bias manifestations rather than their origins, as bias can arise through multiple pathways including biased training data, flawed design, inadequate validation, and implementation contexts (Danks & London, 2017; Das et al., 2021; Olteanu et al., 2019).

When examining how risks associated with algorithmic biases manifest in practice, they present two distinct categories: (1) risks from *individual-use algorithms* (e.g., search engines and facial filters), such as limited access to information and technological benefits (e.g., being able to smoothly use facial

CONTACT Hannah Overbye-Thompson hoverbye@ucsb.edu Department of Communication, University of California, Santa Barbara, CA, USA

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

recognition technology), and (2) risks from organizational algorithms (e.g., healthcare and policing systems) which pose risks to people's fundamental well-being, including their health, livelihood, and financial stability. Despite increasing governmental attention to the risks posed by algorithmic bias (U.S. EEOC, 2024), there remains a gap in understanding how the public perceives algorithmic bias, leaving a fundamental question unanswered: To what extent do people believe they are vulnerable to the negative effects of algorithmic bias and how confident do they feel in their ability to manage this risk?

Using the Risk Perception and Attitude Framework (RPA; Rimal & Real, 2003), this concurrent triangulation mixed methods study aims to explore: (1) individuals' understanding of algorithmic bias with both organizational and individual-use algorithms, (2) individuals' risk and efficacy perceptions vis-à-vis such algorithms, and (3) how these perceptions are related to individuals' demographic characteristics. Theoretically, our study extends the RPA framework to the context of algorithmic bias and examines how risk and efficacy perceptions vary based on algorithm type as well as demographic factors such as gender and race. Understanding whether people perceive algorithmic bias as a risk to their wellbeing is crucial, not merely for fostering increased knowledge of the effects of algorithmic bias, but also for informing the adoption of strategies to protect against algorithmic bias. Given that algorithmic biases often reinforce existing societal inequities (cf. Noble, 2018; O'Neil, 2016) particularly those related to race and gender, our findings can inform policy to reduce algorithmic bias and inform the design of interventions to increase people's agency, all with the goal of ultimately decreasing inequities.

1.1. Exploring risk and efficacy perceptions of algorithmic bias using the RPA

The Risk Perception Attitude Framework (RPA; Rimal & Real, 2003) focuses on the role that risk perceptions and self-efficacy play in motivating protective behaviors. Perceived risk refers to people's interpretations and subjective judgements about risk (Paek & Hove, 2017) and includes the *perceived susceptibility* of a risk (i.e., how vulnerable someone perceives themselves to be) and the *perceived severity* of a risk (i.e., how susceptible someone perceives themselves to be) (Paek & Hove, 2017). Risk perceptions are influenced by a variety of factors, such as personal characteristics (e.g., gender, race, and age) and cognitive factors (e.g., knowledge) (Balog-Way et al., 2020). Self-efficacy refers to people's confidence in their ability to engage in risk-preventative behavior (Rimal & Real, 2003). The RPA posits that the effect of perceived risks on protection motivation is moderated by self-efficacy, such that individuals who perceive a high level of risk and have strong self-efficacy are more likely to engage in protective actions.

In particular, RPA identifies four distinct groups of people based on their risk and efficacy perceptions (Rimal & Real, 2003; Shi & Kim, 2020). When people perceive the risks are high, and feel efficacious in dealing with the risks, RPA refers to them as having a *responsive attitude*. This group is the most motivated to engage in self-protective behaviors. For example, in the context of algorithmic bias, people who perceive themselves at heightened risk for being misidentified by predictive policing algorithms—algorithms that use facial recognition to identify criminals—and who feel efficacious in preventing that from happening, may take action to remove their face from databases used by police (e.g., Clearview AI, 2025; Mueller, 2023).

Conversely, people can feel at heightened risk, but lack the efficacy to protect themselves; this group of people is characterized as having an *avoidant attitude* (Rimal & Real, 2003). Because this group feels particularly vulnerable to experiencing algorithmic bias, but do not feel like they have the ability to mitigate the risks, they are less likely to engage in self-protective behaviors than people with a responsive attitude. People with avoidant attitudes may perceive themselves at risk for being misdiagnosed by a healthcare algorithm, but because they do not feel they can protect themselves, they are less likely to take steps to mitigate this risk, such as talking to their doctor about the steps that went into their diagnosis.

When people perceive themselves at low risk, but feel efficacious, RPA (Rimal & Real, 2003) refers to them as having a *proactive attitude*. Although they are not motivated by perceived risk, they are still likely to engage in more self-protective behaviors than those with an avoidant attitude because they feel efficacious. For example, if people believe a search algorithm can produce racist or sexist results against *other groups* (e.g., Noble, 2018), they may not view the search algorithm as personally threatening their

well-being as it does not spread negative stereotypes about their group identity. However, because they feel capable of controlling how they engage with the algorithm, they may modify their search terms or switch to a different search engine. Indeed, Overbye-Thompson et al. (2024) found that when dealing with algorithmic bias in image recognition algorithms, an individual with a proactive attitude may engage in reinvention behaviors to adapt and manipulate the algorithm to better suit their needs, such as adjusting the lighting conditions of the environment that they are in.

Finally, RPA characterizes people who are low in both perceived risk and efficacy as having an *indifferent attitude*, as they are less likely to be motivated to protect themselves compared to the other three groups. For example, if women encounter a search engine that underrepresents them in professional occupations (e.g., Metaxa et al., 2021), they may assess the risk to their well-being as minimal. They may also feel powerless to address the algorithmic bias, and therefore, lack motivation to seek out solutions to eliminate the problem.

1.2. Applying the RPA in the algorithmic contexts

The RPA framework is particularly well-suited for examining responses to algorithmic bias for several reasons. First, algorithmic bias represents a contemporary risk that varies significantly in both personal relevance and individual capacity for mitigation, which can be explained by core constructs of RPA (i.e., risk perception and self-efficacy). Second, the RPA framework's emphasis on *subjective* risk perceptions aligns well with our focus on *perceived* algorithmic bias. Indeed, unlike risks that can be more objectively detected and assessed (e.g., health, natural disasters, etc.), algorithmic bias risks are often unevenly distributed across demographic groups and may be difficult for individuals to detect. Individuals also may have vastly different awareness and personal experiences with biased systems, making subjective risk assessment a critical factor in understanding individuals' protective behavior vis-à-vis such risks. Third, RPA's four-category typology (responsive, avoidant, proactive, and indifferent attitudes) provides a nuanced understanding of why different individuals may respond differently to the same algorithmic system, moving beyond binary approaches to risk response.

1.3. Risk and efficacy perceptions of algorithmic bias

Categorizing individuals into different RPA attitudinal groups requires a thorough understanding of what risk and efficacy perceptions people hold. In the context of algorithmic bias, one of the challenges in assessing individuals' risk perceptions related to algorithms is the limited understanding many people have about what algorithms are and how they can be biased. Although research has shed some light on how people form folk theories about algorithms (Ytre-Arne & Moe, 2021), there is still a dearth of information regarding how much people know about algorithms, and even more so when it comes to what people know about algorithmic bias specifically.

A key factor contributing to this knowledge gap is the opacity of most algorithmic systems. Burrell (2016) identifies three forms of opacity that hinder users' understanding: (1) intentional concealment by companies (i.e., to protect proprietary technology), (2) the general public's limited ability to comprehend coding languages, and (3) the fundamental disconnect between algorithmic processing of complex data and human intuitive reasoning. This opacity makes it challenging for users to grasp how algorithms are created, maintained, and how they arrive at conclusions.

Despite these barriers, users can develop a degree of algorithmic awareness and literacy through their interactions with algorithmic systems. Swart (2021) found that young people's algorithmic literacy, which includes their algorithm awareness, is often context-specific, varying depending on the platform or feature being discussed and triggered by expectancy violations or explicit personalization cues. However, this experiential knowledge does not always translate into an ability to articulate an understanding of algorithms or motivate users to intervene in algorithmic decisions.

This uncertainty may lead to either an underestimation of risks associated with algorithmic bias, potentially exposing users to harm, or an overestimation that breeds unwarranted fear and distrust in beneficial technologies. Moreover, limited algorithmic literacy can affect users' sense of efficacy in

managing algorithm-related risks. If individuals don't understand how algorithms operate or influence their digital experiences, they may feel powerless to protect themselves or make informed choices. This lack of perceived control could lead to resignation rather than active engagement with privacy settings or other protective measures.

In this study, we aim to address these gaps in the literature by obtaining deeper insights into understanding the risk and efficacy perceptions of algorithmic bias. Primarily, we aim to explore individuals' subjective personal experiences toward algorithms; in doing so we go beyond the current literature by making inferences to discern any notable key dimensions that arise when individuals discuss their thoughts on risk perceptions and efficacy toward algorithmic bias. Accordingly, we utilize RPA as a guiding theoretical framework to pose the following:

RQ1: What are individuals' perceptions of risks and efficacy associated with algorithmic bias?

1.3.1. Organizational versus individual algorithmic bias

When considering the types of risks that individuals may perceive regarding algorithms, and therefore their efficacy perceptions in handling such risks, algorithmic bias can be divided into two broad categories that can pose different risks to users: (1) algorithmic bias that occurs within algorithms that are used by an individual (e.g., Google search algorithms and the algorithms used by facial filters), and (2) algorithmic bias that occurs within algorithms designed for organizational use (e.g., algorithms used to make predictive policing or healthcare diagnoses). Organizational-based algorithmic bias can pose risks to an individual's livelihood, health, or other part of their life that would drastically affect their well-being. For example, a loan approval algorithm that exhibits gender bias (cf., Hassani, 2021) poses a risk to an individual's financial well-being, and a biased healthcare algorithm (cf., Obermeyer et al., 2019) poses a risk to an individual's health. Conversely, bias within individual-use algorithms rarely threatens one's livelihood or health directly. These algorithms do, however, pose a risk to one's ability to seek out relevant information (e.g., in the case of biased search engines, cf., Noble, 2018), and their ability to benefit from technology at the same rate as their peers (e.g., in the case of facial recognition algorithms not working for users with dark skin; Buolamwini & Gebru, 2018). If an algorithm does not function smoothly for a user, they may lose out on benefits that are available to other groups, become more frustrated when using the algorithm, or lose trust in the creators of that algorithm.

Thus, individual-use versus organizational-use algorithms may pose different risks to users and lead to varying perceptions of control and management capacity. Algorithms designed for organizational use may present a greater perceived risk due to their ability to negatively impact wellbeing. Take, for example, the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) risk assessment algorithm, which is utilized to evaluate parole eligibility. Research has shown that the algorithm disproportionately predicts higher recidivism rates among women and Black defendants than what is accurate (Hamilton, 2019; Larson et al., 2016). This form of bias poses a greater risk to individuals than an algorithm designed for individual use, such as image search underrepresenting women in certain professions (Metaxa et al., 2021).

At the same time, the opaque nature of algorithms used by organizations may limit individual's feelings of efficacy in dealing with this bias than with an algorithm designed for individual use. Using the example of a biased image search algorithm once more, when an individual faces this issue, they can adopt strategies to bypass the bias, like typing in different, more specific search terms. Conversely, when encountering a biased algorithm used by a corporation, the individual subjected to it finds themselves without any means of addressing or rectifying the situation. To our knowledge, the literature has yet to develop a thorough understanding of how risk and efficacy perceptions differ between individual and organizational use of algorithms and their associated biases. Based on our arguments and evidence above, we aim to fill that gap by proposing the following hypotheses:

H1. Participants will report higher perceived risk with organizational-based algorithmic biases compared to individual-based algorithmic biases.

H2. Participants will report lower self-efficacy in mitigating algorithmic bias for organizational-based algorithms compared to algorithms designed for individual use.

1.4. Racial/ethnic and gender differences in risk and efficacy perceptions

Next, we seek to understand how such risk and efficacy perceptions related to algorithmic bias vary across different demographic groups. Historically, algorithmic bias has disproportionately exhibited three types of demographic biases: gender bias, racial/ethnic bias, and skin tone bias. Starting as early as the 1970s, there were instances of algorithms favoring men over women in hiring decisions (Schwartz, 2019). More recently, gender bias has been found in a plethora of algorithms. For instance, search engines, are more likely to both perpetuate potentially harmful stereotypes against women and underrepresent them in results (Kay et al., 2015; Otterbacher et al., 2017), and algorithms for credit scoring and loan approvals persistently display gender-based bias against women (Das et al., 2023; Hassani, 2021).

Racial and ethnic bias, where an algorithm advantages White individuals over those from other racial and ethnic groups, is also prevalent. For example the COMPAS algorithm for determining criminal recidivism over-identifies Black inmates as being likely to reoffend (Laqueur & Copus, 2022). Furthermore, algorithms often exhibit reduced functionality for individuals with darker skin. Automated water faucets (Qi Ren et al., 2022), facial recognition algorithms (Buolamwini & Gebru, 2018) and smartwatches (Boonya-Ananta et al., 2021; Ray et al., 2021) have been shown to perform significantly worse for those with dark as opposed to light skin.

Instances of algorithmic bias against specific groups, such as women and people of color (Buolamwini & Gebru, 2018) may shape these groups' risk and efficacy perceptions in responding to algorithmic bias. In addition, research has demonstrated that algorithm knowledge varies by socioeconomic status (Cotter & Reisdorf, 2020; Gran et al., 2021; Gruber & Hargittai, 2023), such that socioeconomically disadvantaged individuals may have less knowledge to help develop clear risk perceptions and efficacy beliefs. While these demographic differences likely manifest in algorithmic bias, to our knowledge, there lacks an understanding of how individuals' risk and efficacy perceptions of such bias may be shaped by their sociodemographic identities. Individuals identifying with groups that are more frequently subjected to algorithmic bias may conceptualize and respond to this issue in a manner distinct from those in groups less prone to encountering such bias. Given that racial/ethnic and gender biases are most prevalent, we pose the following research question:

RQ2: How do individuals' racial/ethnic and gender identities influence their perceptions of the risks and efficacy associated with algorithmic bias?

2. Methods

This study used a concurrent triangulation mixed methods design (Creswell & Poth, 2018, p. 392) that combines and compares quantitative and qualitative data collected simultaneously to explore individuals' understanding and risk and efficacy perceptions of algorithmic bias across individual and organizational uses, as well as how demographic characteristics shape these perceptions. Toward this end, we administered an online survey consisting of both closed and open-ended questions.

Using both quantitative and qualitative data provides a unique perspective on this new research area. While quantitative data can provide a general understanding of the differences in individuals' perceptions of risk and efficacy between the types of algorithms, qualitative data gives participants the opportunity to elaborate on their thoughts, feelings, and perceptions surrounding their understanding of algorithmic bias (e.g., their knowledge, their understanding, their perceived susceptibility, their perceived severity, and the protective actions taken towards algorithmic bias).

2.1. Procedure

Following Creswell and Poth (2018) concurrent triangulation design, we collected quantitative and qualitative data simultaneously to provide complementary perspectives on algorithmic bias perceptions. Our triangulation approach involved three phases: (1) independent analysis of each dataset, (2) systematic comparison of findings, and (3) integration to identify convergent and divergent patterns.

Participants completed an online survey via Qualtrics. To test H1 and H2, we first measured their risk and efficacy perceptions of biased algorithms designed for organizational and individual use adapted from existing measures (see below). Then we asked a series of open-ended questions to answer RQ1 and RQ2. Example questions include: "What is your understanding about algorithmic bias? Have you heard about it before, what do you think it means for an algorithm to be biased?" and "Do you think that you can find a workaround to mitigate any negative effects of bias?" For the quantitative data, we conducted t-tests and MANOVA analyses to test our hypotheses regarding differences in risk and efficacy perceptions between algorithm types and across demographic groups.

In addressing RQ1 and RQ2 for the qualitative data, we used the Taguette software (Rampin & Rampin, 2021) to organize the qualitative data and employed thematic analysis to identify, analyze, and report patterns (i.e., themes) within the open-ended responses, following Braun and Clarke's (2006) six-phase process. This approach was chosen for its flexibility in interpretation, which aligns well with our exploratory aims (Braun & Clarke, 2006). Following initial data familiarization, two authors independently conducted descriptive coding, with particular attention to narrative recurrence and repetition (Owen, 1984). The authors then met to compare codes, resolving any discrepancies to finalize a comprehensive coding list. Subsequently, we collaboratively organized these codes into potential themes, examining their relevance to the research questions on perceived risks and efficacy beliefs related to algorithmic bias. Finally, we independently reviewed and refined the themes to ensure they captured the essential elements of the data.

In the integration phase, we systematically compared findings from both datasets to address our research questions. For RQ1 (perceptions of risks and efficacy), we examined whether qualitative themes aligned with quantitative patterns. For RQ2 (demographic influences), we compared statistical differences with qualitative narratives about identity-based experiences. We specifically looked for convergence (i.e., where both datasets supported similar conclusions), complementarity (i.e., where datasets provided different but mutually enriching perspectives), and divergence (i.e., where findings appeared contradictory, warranting further examination). This approach allowed us to leverage the breadth of quantitative findings with the depth of qualitative insights to provide a comprehensive understanding of algorithmic bias perceptions.

2.2. Participants

To determine the number of participants necessary for the survey, we used G*Power (Faul et al., 2009) to conduct an a priori power analysis for a test comparing two dependent means to address H1 and H2. With a small effect size ($dz = 0.2$), power of 0.80, and alpha level of 0.05, the analysis indicated that 199 participants would be required. We then oversampled to 388 participants in order to capture adequate representation across different demographic groups for our secondary analyses. All participants were recruited from an undergraduate subject pool. Undergraduates were selected for this study because young, educated individuals are more likely to be algorithmically aware (Oeldorf-Hirsch & Neubaum, 2023). Because our goal was to understand risk and efficacy perceptions vis-à-vis specific types of algorithms (i.e., individual and organizational), undergraduates who have relatively high algorithmic awareness and previous user experience are well-suited to provide rich insights. Of the participants, 74% ($n = 254$) identified as female, 25% ($n = 87$) as male, and 1% ($n = 3$) as nonbinary; the remaining 11% ($n = 44$) declined to state their gender. The participants had an average age of 19.90 (SD = 1.87). 22% of participants self-identified as East or South Asian ($n = 76$, with 42 East Asian, 21 South Asian, and 16 mixed Asian), 3% as Black/African American ($n = 10$), 26% as Latino/Latine(x)/Hispanic ($n = 89$), 5% as Middle Eastern ($n = 16$), 38% as White/Caucasian ($n = 130$), and 6% identified with multiple racial categories ($n = 20$). The remaining participants ($n = 47$, 12%) either selected

"Other" ($n = 3$) or declined to state their race. Participants or their parents had a median income of $60,000–$89,999.

2.3. Measures

Appendix A provides a full list of items used in the survey.

2.3.1. Risk perception: Organizational algorithms

We used five items adapted from Wilson et al. (2019) to assess risk perceptions of algorithmic bias in algorithms used by organizations. Example items include "How concerned are you (if at all) about algorithmic bias in algorithms that are used by organizations?" and "Considering any potential effects that algorithmic bias in algorithms that are used by organizations might have on you personally, how concerned are you about algorithmic bias?" Items were measured on a 5-point scale from (1) *not at all* to (5) *extremely* ($\alpha = 0.91$, $M = 2.6$, SD = 0.87).

2.3.2. Risk perception: Individual use algorithms

We used five items adapted from Wilson et al. (2019) to assess risk perceptions of algorithmic bias in algorithms used by individuals. Example items include "How concerned are you (if at all) about algorithmic bias in algorithms that are used by individuals?" and "Considering any potential effects that algorithmic bias in algorithms that are used by individuals might have on you personally, how concerned are you about algorithmic bias?" Items were measured on a 5-point scale from (1) *not at all* to (5) *extremely* ($\alpha = 0.93$, $M = 2.37$, SD = 0.90).

2.3.3. Efficacy beliefs: Organizational algorithms

We used 10 items adapted from Schwarzer and Jerusalem (1995) to assess efficacy beliefs relating to bias found in algorithms that are used by organizations. Example items include "I am confident that I could deal efficiently with unexpected events related to algorithmic bias in algorithms that are used by organizations" and "When I am confronted with algorithmic bias in algorithms that are used by organizations, I can usually find several solutions." Items were measured on a 5-point scale from (1) *not at all true* to (5) *exactly true.* ($\alpha = 0.92$, $M = 2.74$, SD = 0.74).

2.3.4. Efficacy beliefs: Individual use algorithms

We used 10 items adapted from Schwarzer and Jerusalem (1995) to assess efficacy beliefs relating to algorithmic bias found in algorithms that are used by individuals. Example items include "I am confident that I could deal efficiently with unexpected events related to algorithmic bias in algorithms that are used by individuals" and "When I am confronted with algorithmic bias in algorithms that are used by individuals, I can usually find several solutions." Items were measured on a 5-point scale from (1) *not at all true* to (5) *exactly true* ($\alpha = 0.95$, $M = 2.9$, SD = 0.81).

3. Results

3.1. Quantitative data

Consistent with H1 and H2, t-tests revealed individuals viewed algorithms that are used by organizations ($M = 2.6$) as riskier than those used by individuals ($M = 2.37$), $t(362) = 5.35$, $p < 0.001$, and participants felt lower efficacy when dealing with algorithms used by organizations ($M = 2.74$) than those used by individuals ($M = 2.90$), $t(351) = -4.53$, $p < 0.001$.

RQ2 asked how perceptions of algorithmic risk vary across demographic dimensions, specifically race, ethnicity and gender. We conducted a MANOVA to assess how risk perceptions of algorithmic bias differ, if at all, by gender and race/ethnicity. A power analysis using G*Power indicated that a total sample of $N = 132$ was required to detect a medium-sized multivariate effect ($f^2(V) = 0.0625$) with power = 0.95 and $\alpha = 0.05$ in a MANOVA with four dependent variables and a 2 (gender) × 3 (race/ethnicity) between-subjects design. This six-group design implies a minimum of approximately 22

Table 1. Descriptive statistics by group.

Gender	Race/ethnicity	Efficacy(Indv)	Efficacy(Org)	Risk(Indv)	Risk(Org)	*n*
Female	Asian	2.82 (0.89)	2.57 (0.74)	2.47 (1.02)	2.60 (0.92)	41
	LatinX	2.76 (0.75)	2.62 (0.76)	2.25 (0.95)	2.42 (0.84)	55
	White	2.99 (0.83)	2.76 (0.74)	2.25 (0.83)	2.47 (0.79)	101
**	Total (Female)**	**2.89 (0.82)**	**2.68 (0.74)**	**2.30 (0.90)**	**2.48 (0.82)**	**197**
Male	Asian	3.06 (0.78)	2.93 (0.65)	2.37 (0.81)	2.70 (0.79)	21
	LatinX	3.05 (0.87)	3.04 (0.69)	2.45 (0.95)	2.65 (0.80)	12
	White	3.05 (0.85)	2.81 (0.89)	2.61 (0.97)	2.89 (0.98)	29
**	Total (Male)**	**3.05 (0.81)**	**2.90 (0.77)**	**2.50 (0.90)**	**2.78 (0.91)**	**62**

Note: The bold values are significant of $p < .05$. Indv = Individual-use algorithms; Org = Organizational-use algorithms. Values represent *M* (SD). "Total" rows collapse across race within each gender group.

participants per group to ensure sufficient power and stable estimates. To meet this requirement, we restricted the analysis to the three largest racial/ethnic groups in the dataset: White ($n = 130$), Asian (East or South Asian; $n = 62$), and LatinX ($n = 67$). Participants identifying as Black ($n = 10$), Middle Eastern ($n = 16$), multiracial ($n = 20$), or "Other"/unspecified ($n = 47$) were excluded due to small subgroup sizes. Similarly, only participants who identified as male ($n = 62$) or female ($n = 197$) were included; participants identifying as nonbinary ($n = 3$) or declining to report gender were excluded. The final analytic sample consisted of $N = 259$ participants. Participants who did not complete the risk and efficacy measures were excluded from analysis.

We conducted a MANOVA with four dependent variables: perceived risk and efficacy in addressing algorithmic bias in individual-use and organizational-based contexts. The results revealed a significant multivariate effect of gender, Pillai's trace = 0.043, $F(4, 240) = 2.67$, $p = 0.033$, and no significant multivariate effect of race/ethnicity, Pillai's trace = 0.020, $F(8, 482) = 0.59$, $p = 0.784$. Follow-up univariate ANOVAs revealed that gender significantly predicted both organizational efficacy, $F(1, 244) = 4.62$, $p = 0.033$, and organizational risk perception, $F(1, 244) = 5.94$, $p = 0.016$. In both cases, male participants reported higher efficacy ($M = 2.90$, SD = 0.77) and greater perceived risk (M = 2.78, SD = 0.91) than female participants ($M = 2.68$, SD = 0.74 for efficacy; $M = 2.48$, SD = 0.82 for risk). Gender differences for individual-use efficacy ($p = 0.174$) and individual-use risk ($p = 0.164$) were nonsignificant. Post hoc comparisons using Tukey-adjusted estimated marginal means showed no significant differences between racial/ethnic groups on any of the four dependent variables (all $p > 0.28$). See Table 1 for all means and standard deviations by group.

3.2. Qualitative data

3.2.1. Perceived risks

In response to RQ1, we identified several key dimensions in individuals' discussions about their perceptions of the risks associated with algorithmic bias across both individual and organizational uses. While most participants showed basic algorithmic awareness regarding algorithmic effects (e.g., curated content, personalized feeds based on search history), many acknowledged having limited understanding of algorithmic operations or the specific mechanisms underlying algorithmic bias. A few participants, therefore, expressed appreciation for this research's role in raising their awareness on this issue.

Although a noticeable portion of participants did not view algorithmic biases as problematic and considered their own risks minimal given their benefits, the majority expressed significant concerns. These participants frequently perceived biased algorithms as malfunctioning systems that violate users' best interests and undermine social justice with distorted information and unfair outcomes, often attributing these biases to intentional developer decisions and flawed designs, and sometimes to limitations of training data, or inherent biases in human behavior. In reviewing the data, we observed that "bias" was occasionally interpreted as personal biases in processing online information. Although this understanding diverged from our defined concept of algorithmic bias, it nonetheless highlighted a broader recognition among participants: algorithmic content creation is shaped by both platform algorithms and individual user influences. Taking an inductive approach, we further examined how participants conceptualized and perceived algorithmic bias. Four pathways of risks emerged from the data: mental health, privacy and exploitation, fairness and discrimination, and segmentation and polarization.

3.2.1.1. Mental health risks. Participants expressed concerns about potential self-esteem harm due to algorithmic bias, particularly for individuals from underserved groups exposed to content that negatively targets them. They noted that, over time, such exposure could damage self-perceptions and deteriorate mental well-being. For example, some participants worried that curated content, often designed to provoke stronger reactions, could negatively impact emotions due to the profit-driven nature of algorithmic platforms. Continuous exposure to filtered beauty content, for instance, can propel body anxiety and harm self-esteem. In another vein, some participants highlighted the time-consuming nature of social media and expressed frustration with increasingly trivial algorithmic content. For instance, one participant shared, "I am no longer being mentally stimulated, before my TikTok was cute cooking videos and now I know all about Jojo Siwa's life." They expressed that both excessive and insufficient mental stimulation caused by algorithmic curation could have detrimental effects on users' mental health. Moreover, participants expressed concerns that mentally unwell individuals might be exposed to increasingly addictive and harmful content, as algorithms could learn their preferences and perpetuate vicious cycles.

3.2.1.2. Privacy and exploitation risks. Concerns about data privacy and security were prominent, with participants repetitively expressing uncertainty about "who is behind it and what they are doing with the information they have access to." The ongoing datafication intensified participants' negative feelings about being monitored and tracked without explicit consent, from facial recognition to phone monitoring. Many felt insecure about the safety of their data and expressed low trust in corporations' integrity and intentions, worrying about potential underhanded tactics used to profit from users' private information. Additionally, issues like data breaches, leaks, and hacking raised significant concerns about cyber security and the risks of identity theft and fraud, leading participants to question the adequacy of platforms' data storage and management practices.

These privacy concerns were closely linked to the risk of algorithmic bias. Given the opaque nature of data collection and algorithmic decision-making, users remained unclear about how biases in the data might shape the content they see or withhold critical information. This lack of transparency not only fueled mistrust but also placed vulnerable communities at a higher risk of exposure to harmful or exclusionary content. Moreover, participants' worries about privacy and corporate exploitation intensified their concerns over algorithmic bias, as they viewed algorithms as potentially exploiting and profiling them based on biased data inputs.

3.2.1.3. Fairness and discrimination risks. At the societal level, biases embedded in organizational algorithms were perceived as a systemic threat to "individual rights, societal norms, and even demographic processes." As digitalization increasingly penetrates all aspects of life, essential social sectors—such as employment, civic participation, law enforcement, financing, and healthcare—are becoming heavily reliant on algorithmic decisions. These decisions were perceived to be largely based on algorithmic categorization and determination of personal traits, such as sex, race/ethnicity, and sexual orientation. When their categorization mismatches with the prioritized groups in training data, individuals may face negative consequences, such as higher rates of misdiagnosis in medical contexts.

Furthermore, the robotic nature of algorithms raised concerns, as one participant noted, "they could … not take into account the backstory of a patient and their emotions since it is a strictly computed entity." Over time, participants feared that algorithmic bias could exacerbate social inequity and further disadvantage marginalized groups, as their individual values and behaviors are quantified and evaluated by algorithms biased against them. As one example, the facial recognition system has been repeatedly brought up as a potential threat to criminal justice, as it is recognized as biased against minorities due to white-dominated training processes:

> Facial recognition technology used by police can suffer from racial biases, often resulting in higher misidentification rates for people of color compared to white individuals. This can lead to wrongful arrests, invasions of privacy, and a disproportionate surveillance burden on certain communities. Such biases erode trust in law enforcement and perpetuate racial discrimination.

As a result of social grouping, disadvantaged groups may face even more restricted access to financial, healthcare, and legal resources if decisions are influenced by algorithmic bias. The consequences are serious, as these groups are further deprived of essential life opportunities. As one participant noted, "biased algorithms could result in discriminatory hiring practices, denial of loans to qualified individuals, or unfair targeting by law enforcement, all of which can have serious negative consequences for individuals and society as a whole." Participants were hence highly concerned about equity, inclusion, and fairness issues in their discussion. Additionally, certain vulnerabilities may also be exploited more by algorithms (e.g., targeted exposure to financial aid advertisements, predatory scams) for financial gains, which further exacerbates these challenges facing disadvantaged groups.

3.2.1.4. Segmentation and polarization risks. According to participants, curated feeds based on personal identity can lead to differentiated content views among groups, limiting exposure to diverse perspectives. Living within these ingroup echo chambers, shaped with sociodemographic categories and ideological views, in turn, raises the risk of perpetuating intergroup stereotypes and biases in assessing others, thereby exacerbating issues of fairness. This has important societal implications if content curation is intended by platforms for political purposes—a manifestation of organizational use of biased algorithms. Organizational algorithm use, in turn, can create a lack of individual awareness around certain societal issues, with users experiencing different media agendas that weaken the shared foundation for social discourse. As one participant noted, "it could cause more division within society as people become more reinforced in their ways and less open-minded." By segmenting information feeds according to social groups, algorithmic bias is perceived to potentially deepen ideological polarizations and increase societal division.

3.2.2. Efficacy beliefs

We asked participants what they viewed as potential actions they could take in confrontation to algorithmic bias to further explore RQ1 in terms of individuals' efficacy beliefs. Although most participants expressed low self-confidence in addressing potential risks, they proposed various strategies as potential individual and collective solutions.

3.2.2.1. Feelings of powerlessness as an individual. A noticeable portion of participants expressed difficulty in imagining potential solutions to counteract the negative effects of algorithmic bias. These participants frequently used phrases like "I believe," "in my personal opinion," and "I am not sure if I could," indicating their perceptions of limited individual agency over issues that were not under their control. Consequently, some participants adopted an out-of-sight, out-of-mind approach, stating, "I don't know what I could possibly do about it so I don't really worry about it." This shows that a lack of self-efficacy can impede individuals' motivations to engage in further actions to combat algorithmic biases; rather, people may choose to accept the inevitable risks. Indeed, in certain instances, this sense of powerlessness in confronting algorithmic bias escalated to a broader pessimism regarding individuals' influence against systemic forces, as one participant reflected: "sometimes its not worth fighting the system if you want to get far." Participants felt there was no true escape from algorithmic bias in the digital era unless they completely abstained from digital media use. As one participant remarked, "the only way I think you could avoid it for sure is by covering your camera." Even those who desire to address these challenges expressed the need for guidance and resources.

3.2.2.2. Strategic consumption as proactive actions. In contrast, some participants believed they could take personal actions to combat algorithmic bias by strategizing their content consumption and planning their user behavior. First, several participants would turn to more diversified input by increasing the number of media outlets they consume, hoping this could entail a more comprehensive understanding of the world by hearing different voices. One participant suggested to "try not to immediately believe what you see because it may be tailored to your opinion," which implies the importance of fact-checking. They additionally advocated for leveraging the learning nature of algorithms to train the algorithms by actively self-selecting into more diverse content (i.e., "be more calculated with searches").

One participant innovatively suggested the idea of "start acting like I'm someone I'm not for them to give me things I normally wouldn't see," while another mitigated bias by "observing and analyzing both sides of the propositions that I would be voting on" during the 2024 presidential election. As another example, one participant mentioned their experience of deliberately not engaging with targeted advertisements, and another mentioned hitting "uninterested" to reduce their exposure to certain content.

They also emphasized privacy preservation in response to their privacy concerns and called for more caution in management of online information and algorithmic data collection. This may include reduced use as a possible remedy. Some participants believed that users can periodically exploit the forgetfulness of digital media, stating "you could probably become inactive to erase the biases," so they deleted social media applications for a few months or restarted with new accounts. In general, most participants believed more mindfulness in examination and consumption of algorithmic feed can reduce the negative impacts of algorithmic bias. The ability to engage in these creative solutions, of course, requires basic algorithmic awareness and understanding.

3.2.2.3. Collective responsibility for mitigating bias. To address algorithmic bias from corporate perspectives, participants suggested solutions in the development phase by reducing bias in training data (e.g., "I would start by making sure the information they learn from includes lots of different kinds of people") and emphasizing "data preprocessing, algorithmic design, and post-processing techniques." Several suggested introducing incentives, regulations, and open audits for organizations and institutions to promote algorithmic fairness. Additionally, they believed collaboration among stakeholders could help mitigate negative effects (e.g., "incorporating insights from customers, product engineers, marketing teams, and external consultants to help mitigate the bias"). For instance, one participant mentioned external tools designed to eliminate such biases. This theme highlights a focus on customer rights. At the collective level, organizations can mitigate bias by fostering inclusive data practice and enhancing transparency.

3.2.3. Social identity

RQ2 explores how gender and racial/ethnic identity influence individuals' perceptions of the risks and efficacy associated with algorithmic bias. We found three themes associated with one's social identity in shaping their perceptions, namely, (1) psychological ease for dominant groups, (2) systematic disadvantages against minorities, and (3) cultural practices and raised awareness. We additionally identified sexuality as another salient identity factor, which has been less addressed in the literature.

3.2.3.1. Psychological ease for dominant groups. Participants imagined that minorities are treated unfairly as a ramification of algorithmic bias, which aligns well with the fairness and discrimination risks outlined above. By contrast, dominant social groups may instead benefit from algorithmic bias. One participant imagined the systemic inequality would manifest in the manner that "white cishet (particularly men) may view algorithmic bias as less of a risk because they are the dominant group/identity and the systems usually work in their favor." This view was corroborated by our data that White male participants also perceived their privilege in this process: " … I am also a white male, so I likely would not be the target of an algorithmic bias." Dominant social groups might experience greater psychological safety, as reflected in several blunt remarks, such as, "I do not see them as a risk since they are programmed in my favor." Reasonably, many participants expressed concerns that algorithmic bias could perpetuate racism within dominant groups, as echo chambers among these groups are more likely to present racist or sexist behaviors as acceptable, distorting people's perceptions.

3.2.3.2. Systematic disadvantages against minorities. In contrast, participants who are female, queer, and/or individuals of color, especially Hispanic or Black, are confronted with disproportional harms of algorithmic bias. One significant harm is stereotyping based on social categories, associating dominating groups with positive connotations and minorities with negative ones. Often going unnoticed, the bias has been shaping our social perception and cultural context, leading to unfair treatment and

discrimination (cf., Noble, 2018). Segmentation risks can also emerge, as participants were exposed to more in-group members and retrospectively felt "at risk of thinking the world is safer than it actually is." At the practical level, algorithmic bias can induce extra negativity to minorities in a world governed by robotic quantification and computing. Participants recognized the algorithmic harm on multiple aspects of social equity, such as assignment of credit scores, loan distribution, job applications, healthcare and facial recognition. A sample comment goes as follows: "I feel like certain minority groups (class, gender, race, etc.) may be unfairly profiled, discriminated against, denying opportunities, and perpetuating stereotypes." Algorithmic bias was considered an amplifier of existing systematic barriers facing minorities.

Based on personal experience and social identity, many participants expressed higher concerns about algorithmic biases because they were, for example, a young woman, or a non-binary queer Latinx. They expressed fear for the disproportionate mental health risks tied to their everyday consumption and information-seeking of online content. Moreover, they suffered from algorithmic unfairness in real life scenarios. Because the medical information presented online is created with stronger male-orientation, for instance, female participants noted their difficulties in finding relevant information and had to conduct more extensive research with modified keywords. As another example, one participant was frustrated by their difficulties to get loans because their parents did not have social security. These concerns about disproportionate algorithmic harm and lived experiences of unfair treatment highlight the unique challenges faced by disadvantaged groups in the era of algorithmic dominance.

3.2.3.3. Cultural practices and raised awareness. In response to systematic disadvantages, marginalized communities engage in cultural practices and develop unique discourses. This, in turn, fosters heightened cultural sensitivity and general distrust toward digital technology, as they feel targeted by algorithms: "Probably, because as a Latina I feel like my mom would be like 'Oh my god that computer is spying on us' and get a little paranoid like a lot of other Hispanic people in my family do;" "I'm Jewish and have learned to be skeptic of anything that has to do with tech and like individual data collection." The truth is that such paranoia is not baseless and is often ingrained through acculturation. More specifically, individuals from minority groups may also experience greater perceptions of concrete algorithmic risks. For instance, their personal experience of being stereotyped and discriminated sensitized their awareness of potential risks. Positively, this may cultivate "heightened concerns about fairness, equity, and justice in algorithmic decision-making processes" and an understanding of "algorithmic bias as a systemic issue that reflects broader social inequalities and power dynamics." Consequently, minorities may be more likely to engage in active efforts to combat these biases.

3.3. Data integration

The results from both the quantitative and qualitative analyses consistently show differences in perceived risks and efficacy associated with algorithms used by organizations versus those used by individuals. Quantitatively, participants viewed organizational algorithms as significantly riskier than individual ones and felt a lower sense of efficacy when dealing with organizational algorithms. Qualitatively, this perception was echoed in participants' statements, such as one respondent who mentioned: "When it comes to algorithms used by an individual, I am less confident in dealing with algorithmic bias as opposed to algorithms used by organizations." This convergence of data from both methods underscores the heightened concern and reduced confidence individuals have regarding organizational algorithms, reflecting a broader skepticism toward the institutional deployment of these technologies.

Conversely, while our quantitative analysis did not reveal any significant differences in perceptions of risk and efficacy based on race or ethnicity, the qualitative data provided significant insights into this dimension. For example, a Jewish participant highlighted concerns about how algorithmic decisions could amplify stereotypes or lead to unequal treatment, particularly on online platforms and surveillance systems: "antisemitic content might affect what gets recommended or filtered, or in surveillance systems where biases could impact how individuals are profiled or monitored." Moreover, another

participant echoed this sentiment, stating: "Being Hispanic and a woman, I can see algorithmic bias leading to some potential opportunities being more difficult to get."

Beyond race and ethnicity, gender differences in our quantitative findings present another intriguing pattern that requires deeper examination. The quantitative finding that male participants reported both higher perceived risk and greater efficacy regarding organizational algorithmic bias initially appears counterintuitive, given that algorithmic bias typically disadvantages women and racial minorities rather than White males. However, this pattern may reflect different conceptualizations of "risk." While female and marginalized participants in the qualitative data focused on personal vulnerability and direct harm, male participants may have been assessing risk in more abstract, societal terms. The qualitative data suggest that male participants, particularly White males, often acknowledged their relative immunity from personal targeting: "I am also a white male, so I likely would not be the target of an algorithmic bias," yet expressed concerns about broader systemic implications for society. For example, one White male participant noted "algorithm bias could certainly cause damage to politics and society by influencing what and how people think and vote, but I do not view it as a potential harm to me." This detached perspective may enable them to recognize organizational algorithmic bias as a significant societal risk while simultaneously feeling confident in their ability to address it.

4. Discussion

This study contributes to understanding of algorithmic bias by examining individuals' perceptions through the lens of the Risk Perception Attitude (RPA) Framework (Rimal & Real, 2003). Utilizing a concurrent triangulation mixed methods design, we underscore significant differences in individuals' perceived risk and efficacy, particularly between organizational and individual-use algorithms, while illuminating the complex ways that race, ethnicity and gender shape these perceptions. By extending the RPA framework to algorithmic contexts, we demonstrate how risk perceptions and efficacy beliefs interact in technological settings. Furthermore, while previous research has extensively documented the existence and impacts of algorithmic bias, our study provides new insights by empirically investigating how different demographic groups perceive and respond to these technological risks—a perspective that has been largely absent from the algorithmic bias literature.

4.1. Organizational versus individual algorithms

Quantitative data indicated that individuals perceive algorithms used by organizations as significantly riskier than those used by individuals. This perception is likely due to the higher stakes involved with organizational algorithms, which can impact aspects of life such as healthcare, criminal justice, and financial stability. For example, biased healthcare algorithms can lead to disparities in medical treatment, while biased risk assessment tools in the criminal justice system can disproportionately affect minority groups. The lower efficacy scores associated with organizational algorithms further suggest that individuals feel less capable of mitigating the risks posed by these algorithms.

Within the RPA Framework, individuals can fall into one of four categories based on their risk perceptions and efficacy beliefs: *indifferent*, *avoidant*, *responsive*, and *proactive attitudes*. Our findings suggest that many individuals fall into the avoidant attitude category when it comes to organizational algorithms. According to the RPA, individuals with avoidant attitudes are less likely to engage in self-protective behaviors, such as attempting to remove themselves from online databases or actively developing strategies to mitigate algorithmic bias. This is corroborated by qualitative finding that those with high-risk perceptions but low efficacy beliefs were more likely to passively accept algorithmic bias and rationalize their inaction, a pattern that was particularly salient among marginalized racial, ethnic, and gender groups. Our data suggest that intersecting identities create particularly complex patterns of risk perception and efficacy beliefs. Participants with multiple marginalized identities (e.g., being both a racial/ethnic minority and a woman) described experiencing compounded vulnerabilities to algorithmic bias while feeling even more powerless to address them. This intersectional effect may intensify avoidant attitudes, as these individuals perceived multiple vectors of potential discrimination yet felt overwhelmed by the complexity of addressing intersecting algorithmic biases.

On the other hand, our qualitative data indicate that some participants, particularly those who are White and male, perceived algorithmic biases as posing little risk to them, regardless of algorithm type. It's important to note that this represents a minority perspective within our sample, as our quantitative data demonstrate that participants overall perceived organizational algorithms as significantly riskier than individual-use algorithms ($M = 2.6$ vs. $M = 2.37$). For this subset of participants who perceived low personal risk from organizational algorithms, despite the general trend of higher risk perception, they also reported feeling little efficacy in addressing organizational algorithmic biases. According to the RPA, these individuals would exhibit an indifferent attitude, making them the least motivated to find workarounds for algorithmic bias compared to the other groups (cf., Overbye-Thompson & Rice, 2025 for more information on workarounds to algorithmic bias). However, our quantitative data show that both male and female participants feel more efficacious in dealing with individual algorithms than organizational ones, while male participants perceive there being a greater risk for algorithmic bias in organizational algorithms than female participants. Thus, for individual-use algorithms, these participants would likely exhibit a proactive attitude, meaning they are more inclined to engage in self-protective behaviors, according to the RPA. This proactive attitude was reflected in the qualitative data, where many participants mentioned their ability to find workarounds for individual-use algorithms. These findings together elaborate the RPA in the novel context of algorithmic biases.

4.2. Strengths and limitations

This study's qualitative data provided insights into people's risk and efficacy perceptions when dealing with algorithmic bias in algorithms used by organizations (e.g., facial recognition algorithms used by police, hiring algorithms) and by individuals (e.g., Google search, facial recognition on phones). Although our qualitative data suggested racial/ethnic differences in risk and efficacy perceptions, quantitatively we did not find the same results in our sample. Future research should include a sample with more participant diversity to determine if these results replicate. This discrepancy may reflect several methodological considerations. First, standardized Likert scales may not adequately capture the culturally-informed ways different communities conceptualize risk. Our qualitative findings suggest that marginalized groups may not necessarily report *higher* numerical risk perceptions, but rather demonstrate *qualitatively different* frameworks for understanding algorithmic risks. Second, social desirability bias may have influenced participants' responses to standardized risk perception measures, with some marginalized participants potentially underreporting concerns about algorithmic bias to avoid reinforcing negative stereotypes about being "overly sensitive" to discrimination, due to learned strategies for navigating academic spaces, or as a result of internalized responses to having their discrimination concerns previously dismissed or minimized.

Third, our study is limited by its sample and cannot be generalized to the broader population. We selected a student sample because younger, educated individuals are more likely to be aware of algorithms and can therefore provide insights into their risk perceptions of algorithmic bias. However, this choice creates an important paradox in algorithmic bias research: those most vulnerable to algorithmic bias often have the least algorithmic literacy and are therefore underrepresented in studies like ours. This sampling limitation affects the broader generalizability of our findings. Our participants' high educational levels likely enabled them to articulate sophisticated algorithmic concerns and propose strategic responses that may not be accessible to vulnerable populations such as older adults, individuals with lower educational attainment, or economically disadvantaged groups. These populations may experience algorithmic bias differently and lack the digital literacy skills needed for the strategic consumption behaviors our participants described. Future research should prioritize inclusive methodologies that capture the experiences of those who are simultaneously most affected by algorithmic bias and least equipped to address it. Fourth, while our qualitative data revealed varying levels of algorithmic awareness among participants, we lacked standardized measures to control for these differences in our quantitative analyses. Future research should thus build upon this finding and quantitatively measure algorithmic literacy or prior knowledge of algorithmic bias as predictors of risk and efficacy perceptions vis-à-vis algorithmic bias.

Despite the limitations of our study, it has several strengths. Our concurrent triangulation mixed methods design allowed us to capture both the breadth of perceptions across our sample through quantitative measures and the nuanced, contextual understanding of how individuals conceptualize algorithmic bias through qualitative analysis. This approach was particularly valuable given the nascent state of research on algorithmic bias perceptions, as it enabled us to identify patterns while also discovering unexpected themes, such as the four risk dimensions, that emerged from participants' own experiences and language. Furthermore, while research has been conducted on algorithm awareness and algorithmic fairness, *perceptions* of algorithmic bias remain largely understudied. This is despite the fact that many algorithms exhibit bias, often disadvantaging the already-marginalized groups. Our findings indicate that people feel more at risk from bias in algorithms used by organizations than those used by individuals, yet they also feel less capable of addressing such biases. This suggests that individuals may be less likely to take protective measures against organizational algorithmic bias, even though it could be beneficial for them. Therefore, future research or interventions aimed at combating algorithmic bias should focus on increasing individuals' efficacy.

4.3. Practical implications

The findings of this study offer practical insights for individuals navigating algorithmic systems in their daily lives. For individual-use algorithms like search engines and social media, users can actively protect themselves by diversifying their information sources, fact-checking content, and strategically managing their digital footprint through practices like selective engagement with content. When confronting organizational algorithms in healthcare, employment, or financial contexts where users reported feeling less efficacious, individuals can benefit from documenting their experiences with algorithmic systems and requesting explanations for algorithmic decisions that affect them. Furthermore, since users perceive organizational algorithms as higher risk yet feel less able to address their biases, organizations should prioritize making their algorithmic systems more transparent and accountable. Educational initiatives should reach out to populations that are vulnerable to algorithmic bias, particularly given that marginalized groups may adopt an avoidant attitude that may lead them to miss out on important services and opportunities.

4.4. Policy implications

Our findings also point to several important policy-level recommendations that could address algorithmic bias at a systemic level. Given that participants perceived organizational algorithms as significantly riskier yet felt less efficacious in addressing their biases, regulatory frameworks are needed to bridge this gap between individual vulnerability and limited personal agency. For example, our results support the need for mandatory algorithmic auditing requirements, particularly for organizational algorithms used in high-stakes domains such as healthcare, criminal justice, employment, and financial services. Since participants expressed concerns about the opacity of organizational algorithmic systems and their limited ability to address biases in these contexts, policymakers should consider implementing regular bias audits conducted by independent third parties. These audits should specifically examine the four risk dimensions identified in our study: mental health impacts, privacy and exploitation risks, fairness and discrimination outcomes, and segmentation and polarization effects.

In addition, our findings on inequalities in efficacy perceptions suggest that algorithmic literacy initiatives should be incorporated into broader digital rights education policies. Such programs should be designed with particular attention to addressing the differential impacts of algorithmic bias on various demographic groups, helping to build the collective efficacy needed to advocate for fair algorithmic systems.

5. Conclusion

This study contributes to the new body of literature on algorithmic bias by highlighting how perceptions of risk and efficacy vary based on the type of algorithm and demographic factors. The integration

of both quantitative and qualitative data offers a comprehensive understanding of these perceptions, providing valuable insights into the perceived risk of algorithmic bias. Given the lack of empirical literature focusing on better understanding algorithmic bias and its risks, adopting a mixed methods research design was of key importance to holistically understand individuals' perceptions towards algorithmic biases. By combining both forms of data, our results invite further study in this new body of literature by giving other researchers a better depth of understanding. By leveraging the RPA Framework, this research underscores the importance of considering both risk perceptions and efficacy beliefs when developing future studies aimed at combating the negative impacts of algorithmic bias, which can lead academic researchers, organizations and individuals to develop better strategies to reduce algorithmic bias in hopes of making algorithms more equitable for everyone.

Ethics approvals

This study has obtained ethics approval from the UCSB Human Subjects Committee, protocol # 9-24-0246.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

CRediT: **Hannah Overbye-Thompson**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing; **Erick A. Garcia**: Formal analysis, Writing – original draft; **Xinyi Zhang**: Formal analysis, Writing – review & editing; **Laurent H. Wang**: Writing – review & editing.

ORCID

Hannah Overbye-Thompson http://orcid.org/0000-0001-5235-9486
Erick A. Garcia http://orcid.org/0009-0008-2275-9014
Xinyi Zhang http://orcid.org/0000-0001-5216-2093
Laurent H. Wang http://orcid.org/0000-0002-6795-3039

Data availability statement

The data that support the findings of this study are available upon reasonable request from the corresponding author. The data are not publicly available due to the sensitive nature of the topic discussed by the article.

References

Balog-Way, D., McComas, K., & Besley, J. (2020). The evolving field of risk communication. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *40*(S1), 2240–2262. https://doi.org/10.1111/risa.13615

Bhuiyan, J. (2024, December 14). *She didn't get an apartment because of an AI-generated score – and sued to help others avoid the same fate*. The Guardian. https://www.theguardian.com/technology/2024/dec/14/saferent-ai-tenant-screening-lawsuit

Boonya-Ananta, T., Rodriguez, A. J., Du Le, V. N., Ramella-Roman, J. C., & Ajmal. (2021). Monte Carlo analysis of optical heart rate sensors in commercial wearables: The effect of skin tone and obesity on the photoplethysmography (PPG) signal. *Biomedical Optics Express*, *12*(12), 7445–7457. https://doi.org/10.1364/BOE.439893

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, *81*, 15. https://proceedings.mlr.press/v81/buolamwini18a.html

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512. https://doi.org/10.1177/2053951715622512

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, *10*(1), 1–12. https://doi.org/10.1057/s41599-023-02079-x

Clearview AI. (2025). *Clearview AI | Facial Recognition*. https://www.clearview.ai
Cotter, K., & Reisdorf, B. C. (2020). Algorithmic knowledge gaps: A new dimension of (digital) inequality. *International Journal of Communication*, (19328036), 14.
Creswell, J. W., & Poth, C. N. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 4691–4697). https://doi.org/10.24963/ijcai.2017/654
Das, S., Donini, M., Gelman, J., Haas, K., Hardt, M., Katzman, J., Kenthapadi, K., Larroy, P., Yilmaz, P., & Zafar, M. B. (2021). Fairness measures for machine learning in finance. *Journal of Financial Data Science*, *3*(4), 33–64. https://doi.org/10.3905/jfds.2021.1.075
Das, S., Stanton, R., & Wallace, N. (2023). Algorithmic fairness. *Annual Review of Financial Economics, 15*, 565–593. https://doi.org/10.1146/annurev-financial-110921-125930
Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561
Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, *24*(12), 1779–1796. https://doi.org/10.1080/1369118X.2020.1736124
Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test part 3: Demographic effects. (NIST IR 8280; p. NIST IR 8280)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280
Gruber, J., & Hargittai, E. (2023). The importance of algorithm skills for informed Internet use. *Big Data & Society*, *10*(1), 20539517231168100. https://doi.org/10.1177/20539517231168100
Hamilton, M. (2019). The sexist algorithm. *Behavioral Sciences & the Law*, *37*(2), 145–157. https://doi.org/10.1002/bsl.2406
Hassani, B. K. (2021). Societal bias reinforcement through machine learning: A credit scoring perspective. *AI and Ethics*, *1*(3), 239–247. https://doi.org/10.1007/s43681-020-00026-z
Johnson, K. (2022). *How wrongful arrests based on AI derailed 3 men's lives*. Wired. https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/
Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3819–3828). https://doi.org/10.1145/2702123.2702520
Laqueur, H. S., & Copus, R. W. (2022). An algorithmic assessment of parole decisions. *Journal of Quantitative Criminology*, *40*(1), 151–188. https://doi.org/10.1007/s10940-022-09563-8
Larson, J., Angwin, J., Kirchner, L., Mattu, S. (2016). *How we analyzed the compas recidivism algorithm.* ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
Metaxa, D., Gan, M. A., Goh, S., Hancock, J., & Landay, J. A. (2021). An image of society: Gender and racial representation and impact in image search results for occupations. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–23. https://doi.org/10.1145/3449100
Mueller, J. (2023). *How to delete yourself from the clearview AI database: 9 steps*. https://www.wikihow.com/Delete-Yourself-from-the-Clearview-AI-Database
Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
Obermeyer, Z., Brian, P., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.)*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
Oeldorf-Hirsch, A., & Neubaum, G. (2023). Attitudinal and behavioral correlates of algorithmic awareness among German and U.S. social media users. *Journal of Computer-Mediated Communication*, *28*(5), zmad035. https://doi.org/10.1093/jcmc/zmad035
Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, 13. https://doi.org/10.3389/fdata.2019.00013
Otterbacher, J., Bates, J., & Clough, P. (2017). Competent men and warm women: Gender stereotypes and backlash in image search results. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6620–6631). https://doi.org/10.1145/3025453.3025727
Overbye-Thompson, H., & Rice, R. E. (2025). Understanding how users may work around algorithmic bias. *AI & Society*, 1–15. https://doi.org/10.1007/s00146-025-02498-1
Overbye-Thompson, H., Hamilton, K. A., & Mastro, D. (2024). Reinvention mediates impacts of skin tone bias in algorithms: Implications for technology diffusion. *Journal of Computer-Mediated Communication*, *29*(5), zmae016. https://doi.org/10.1093/jcmc/zmae016
Owen, W. F. (1984). Interpretive themes in relational communication. *Quarterly Journal of Speech*, *70*(3), 274–287. https://doi.org/10.1080/00335638409383697

Paek, H.-J., & Hove, T. (2017). Risk perceptions and risk characteristics. In H.-J. Paek & T. Hove (Eds.), *Oxford research encyclopedia of communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.283

Qi Ren, X., Heacock, H., BCIT School of Health Sciences, Environmental Health. (2022). Sensitivity of infrared sensor faucet on different skin colours and how it can potentially effect equity in public health. *BCIT Environmental Public Health Journal*. https://doi.org/10.47339/ephj.2022.216

Rampin, R., & Rampin, V. (2021). Taguette: Open-source qualitative data analysis. *Journal of Open Source Software*, *6*(68), 3522. https://doi.org/10.21105/joss.03522

Ray, I., Liaqat, D., Gabel, M., & de Lara, E. (2021). Skin tone, confidence, and data quality of heart rate sensing in WearOS smartwatches. *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops)* (pp. 213–219). https://doi.org/10.1109/PerComWorkshops51409.2021.9431120

Resende, G. H., Nery, L. F., Benevenuto, F., Zannettou, S., & Figueiredo, F. (2024). A Comprehensive *view of the biases of toxicity and sentiment analysis methods towards utterances with* African American *English expressions*. arXiv:2401.12720. arXiv. https://doi.org/10.48550/arXiv.2401.12720

Rimal, R. N., & Real, K. (2003). Perceived risk and efficacy beliefs as motivators of change. *Human Communication Research*, *29*(3), 370–399. https://doi.org/10.1111/j.1468-2958.2003.tb00844.x

Schwartz, O. (2019). *Untold history of AI: Algorithmic bias was born in the 1980s – IEEE Spectrum*. https://spectrum.ieee.org/untold-history-of-ai-the-birth-of-machine-bias

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. *Causal and Control Beliefs*, *35*(37), 82–003. https://doi.org/10.1037/t00393-000

Shi, J., & Kim, H. K. (2020). Integrating risk perception attitude framework and the theory of planned behavior to predict mental health promotion behaviors among young adults. *Health Communication*, *35*(5), 597–606. https://doi.org/10.1080/10410236.2019.1573298

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society*, *7*(2), 20563051211008828. https://doi.org/10.1177/20563051211008828

U.S. EEOC. (2024). *What is the EEOC's role in AI?* https://www.eeoc.gov/sites/default/files/2024-04/20240429_What%20is%20the%20EEOCs%20role%20in%20AI.pdf

Wilson, B., Hoffman, J., Morgenstern, J. (2019). *Predictive inequity in object detection* (arXiv:1902.11097). arXiv. http://arxiv.org/abs/1902.11097

Wilson, R. S., Zwickle, A., & Walpole, H. (2019). Developing a broadly applicable measure of risk perception. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *39*(4), 777–791. https://doi.org/10.1111/risa.13207

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*, *43*(5), 807–824. https://doi.org/10.1177/0163443720972314

About the authors

Hannah Overbye-Thompson is a doctoral candidate in the Department of Communication at the University of California, Santa Barbara. Her work examines how algorithmic and AI bias shape human behavior, perception, and technology adoption.

Erick A. Garcia is a PhD student in the Department of Communication at the University of California, Santa Barbara. His research draws on social psychological theories to examine how digital media affects intergroup relations. He employs both quantitative and qualitative methodologies to investigate online hate and approaches to curtail it.

Xinyi Zhang is an Assistant Professor in Department of Human Communication Studies at California State University, Fullerton. Her research investigates negative experiences with digital media, such as stress and inequality. She uses quantitative and qualitative methods to examine how emerging technologies shape well-being, human–machine relationships, and structural disparities.

Laurent H. Wang is a doctoral candidate in Department of Communication at University of California, Santa Barbara. His research studies communication technology, privacy, and social inequality. He employs quantitative and qualitative methods to examine how people (1) negotiate privacy decisions in various technological contexts and (2) navigate digital inequality challenges.

Appendix A

Measures

Risk perception

Response options (for all items):
Please tell us how you feel about the following:
1 = Not at all, 2 = Slightly, 3 = Moderately, 4 = Very, 5 = Extremely

Organizational Algorithms

- How concerned are you (if at all) about algorithm bias in algorithms that are used by organizations?
- When you think about algorithm bias in algorithms that are used by organizations for a moment, to what extent do you feel fearful?
- When you think about algorithm bias in algorithms that are used by organizations for a moment, to what extent do you feel anxious?
- When you think about algorithm bias in algorithms that are used by organizations for a moment, to what extent to you feel worried?
- Considering any potential effects that algorithm bias in algorithms that are used by organizations might have on you personally, how concerned are you about algorithm bias?

Individual Use Algorithms

- How concerned are you (if at all) about algorithm bias in algorithms that are used by individuals?
- When you think about algorithm bias in algorithms that are used by individuals, to what extent do you feel fearful?
- When you think about algorithm bias in algorithms that are used by individuals for a moment, to what extent do you feel anxious?
- When you think about algorithm bias in algorithms that are used by individuals for a moment, to what extent to you feel worried?
- Considering any potential effects that algorithm bias in algorithms that are used by individuals might have on you personally, how concerned are you about algorithm bias?

Efficacy beliefs

Response options (for all items):
Please tell us how true the following statements are to you:
1 = Not True at All, 2 = Slightly True, 3 = Moderately True, 4 = Very True, 5 = Exactly True

Organizational Algorithms

- I could always manage to solve difficult problems relating to algorithm bias in algorithms that are used by organizations if I try hard enough
- If confronted with algorithm bias in algorithms that are used by organizations, I can find the means to get what I want
- It is easy for me to stick to my aims and accomplish my goals if confronted with algorithm bias in algorithms that are used by the organizations
- I am confident that I could deal efficiently with unexpected events related to algorithm bias in algorithms that are used by organizations
- Thanks to my resourcefulness, I know how to handle unforeseen situations relating to algorithm bias in algorithms that are used by organizations
- I can solve most problems relating to algorithm bias in algorithms that are used by organizations if I invest the necessary effort
- I can remain calm when facing algorithm bias in algorithms that are used by organizations because I can rely on my coping abilities
- When I am confronted with algorithm bias in algorithms that are used by organizations, I can usually find several solutions
- If I am in trouble relating to algorithm bias in algorithms that are used by organizations I can usually think of a solution
- I can usually handle algorithm bias in algorithms that are used by organizations

Individual Use Algorithms

- I can usually handle algorithm bias in algorithms that are used by organizations
- If confronted with algorithm bias in algorithms that are used by individuals, I can find the means to get what I want
- It is easy for me to stick to my aims and accomplish my goals if confronted with algorithm bias in algorithms that are used by individuals
- I am confident that I could deal efficiently with unexpected events related to algorithm bias in algorithms that are used by individuals
- Thanks to my resourcefulness, I know how to handle unforeseen situations relating to algorithm bias in algorithms that are used by individuals
- I can solve most problems relating to algorithm bias in algorithms that are used by individuals if I invest the necessary effort
- I can remain calm when facing algorithm bias in algorithms that are used by individuals because I can rely on my coping abilities
- When I am confronted with algorithm bias in algorithms that are used by individuals, I can usually find several solutions
- If I am in trouble relating to algorithm bias in algorithms that are used by individuals I can usually think of a solution
- I can usually handle algorithm bias in algorithms that are used by individuals