SSRG International Journal of Humanities and Social Science
ISSN: 2394-2703/ https://doi.org/10.14445/23942703/IJHSS-V12I5P110

Volume 12 Issue 5, 97-113, Sep-Oct 2025


*Original Article*

# Investigating Awareness and Perception of Bias in AI-Driven Platforms: A Survey-Based Study

Aishaani Agarwal

*Harvest International School, Karnataka, India.*

*Corresponding Author : aishaani@gmail.com*

Received: 25 August 2025 Revised: 29 September 2025 Accepted: 16 October 2025 Published: 30 October 2025

***Abstract*** *- Artificial Intelligence (AI) increasingly influences decisions in everyday life, from education/health care/and homework to employment/social media. While AI offers multiple advantages of efficiency and innovation, anxiety about fairness, accountability, and transparency in AI systems remains at the forefront of conversation both in the news and public debate. Researchers have raised these issues, pointing to the fact that bias can be introduced at multiple levels in AI—from the training data the AI is trained on to the design choices the AI developer made to the way the user interacts with the AI decision-making system. Researchers have posited important questions related to how AI systems can affect society and people's responses to these AI systems. As such, this study examines how participants perceive and experience bias in AI systems (i.e., decision-making systems) with a focus on trust, accountability, and demographic characteristics such as age, gender, and education. Based on survey data from the youth participants, this research examines where participants perceive bias in AI systems, how participants feel about AI systems making day-to-day decisions in their daily lives, and who participants think should be liable when there is bias in AI systems. The research not only contributes to the ongoing dialogue about fairness and trust in AI but also has implications for design, governance, and public engagement.*

***Keywords*** *- Accountability, AI bias, Artificial Intelligence, Human-Computer Interaction, Perception Bias.*

## 1. Introduction

In the last few decades, Artificial Intelligence (AI) has seen remarkable growth driven by machine learning algorithms, increased availability of data, and enhanced computational capacity. Today, it is deeply embedded in everyday human life. As a transformative technology, AI has enabled automation across industries and households with enhanced decision-making capabilities. Its applications have expanded to domains such as personal assistance, navigation, and education (R. S. T. Lee, 2020).

AI models such as ChatGPT and Meta AI excel in natural language processing. They are used in chatbots, translation, and interactive conversations. These tools are not just widely used but have also become increasingly popular in consequential domains such as healthcare, the labour market, and education, raising novel questions about accountability and fairness (Madsen & Toston, 2025).

Additionally, in medicine, large language models are being explored for tasks like disease diagnostics, writing radiology reports, scientific research writing, and medical teaching aids (Xiao et al., 2024). AI has very rapidly changed how knowledge is disseminated and perception influences decisions. By filtering information flows, AI-powered news aggregators affect perception and decision-making, thereby influencing public discourse. The sociotechnical systems approach views society as deeply interconnected and reveals that any technology, like AI, does not exist in isolation but is always embedded in social and cultural contexts. Hence, from a sociotechnical systems perspective, such platforms are not neutral. They possess biases shaped by societal norms and biases. AI has a large influence on information gathering and steering perception (Liu & Xu, 2024). Artificial intelligence (AI) systems now influence business decisions, government policies, and conversations in everyday lives, generating outcomes that extend far beyond individuals (Ntoutsi et al., 2020b). These systems provide solutions that can enhance efficiency by producing data-driven solutions rapidly, yet there is a very real risk to individuals, specifically regarding employment opportunities or access to medical care, when using flawed algorithms (Datta, Tschantz, and Datta, 2015).

Some stark examples of this risk include the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) tool, which is a risk assessment algorithm used in the U.S criminal justice system. It is used to predict how likely someone is to commit another crime if released. This tool was erroneously assigning a higher recidivism risk score to Black defendants (Angwin, Larson, Mattu, & Kirchner, 2016) (racial bias) and Google's ad targeting system serving fewer ads for high-paying jobs to women, potentially indicating gender bias (Datta, Tschantz, & Datta, 2015). These examples amplify concerns surrounding discriminatory effects inherently present in the design and deployment of AI systems. Recent literature identifies an urgent priority to "find, fix, and avoid" the type of bias in AI at every aspect of design and development so as not to make existing inequalities worse (Nazer et al., 2023).

Bias in AI arises from unbalanced data, flawed models, and human interactions replicating structural inequalities such as race and caste (Rasali et al., 2024). These datasets reflect the enduring impacts of racism and bias, perpetuating global disparities in health, income, and opportunity (Rasali et al., 2024). As AI continues to influence the fields of health, education, and finance, these biases may become perpetuated and amplified if allowed to remain unchallenged (Bonarini, 2022). Hence, it is critical for AI developers, designers, and stakeholders to take responsibility by raising awareness amongst the users and collaborating to ensure that AI systems are fair to prevent bias in AI-powered systems (Bonarini 2022).

AI systems often reflect the biases of the sociocultural context from which the data-driven organization emerges (Ali et al., 2021). This is troubling because it has a minimum chance of harming some AI users from all demographics (Leavy, 2018). For example, when AI models are mostly trained on Western or English-heavy content, they end up drowning out voices from non-Western or local communities (Binns, 2018). This kind of bias can create harm in areas like hiring, where it strengthens existing prejudices and leads to unfair treatment (Ali et al., 2021).

AI bias can also develop and build upon existing models from the way people use it, as these models learn from user interactions and behaviour. These models often reflect social hierarchies and repeat harmful stereotypes (Leavy, 2018). AI algorithms can reinforce bias if users repeatedly click on biased content by suggesting similar content. This can be a cause for loops within responses that normalize discrimination, like both gender bias and racial bias. Additionally, people who belong to lower-income or marginal caste groups might be underrepresented in data when they are collected or misclassified, so discriminatory treatment results (class bias, caste bias) (Binns, 2018). AI systems will have the potential for normalization and for routinization unless active action recognizes and rectifies these biases. To resolve these concerns, experts need to ethically intervene, sociologically intervene, and create policy at all stages of AI design (Ali et al., 2021). Most of the work on AI bias has focused on tech fixes like making better algorithms, fixing data sets, or using clear AI to cut down unfair results (Hou et al., 2024; Gorska & Jemielniak, 2023). However, existing literature shows that people often perceive AI systems to be fair and objective, overlooking how bias can be embedded through data selection, design choices, and information sorting by the algorithm (Brauner et al., 2023; Peralta et al., 2021). While some studies have examined the intersection of the public's hopes and fears regarding AI, very few integrate views in areas like jobs, media, or daily choices (Brauner et al., 2023). This gap highlights the need for a framework that can combine technical, ethical, and social perspectives.

Additionally, Prior research shows that an individual's reaction to bias varies based on their understanding of AI, degree of human-likeness in the AI, or whether they are primed to look for errors (Hou et al., 2024). However, these evaluations are often limited to experimental settings and specific use-cases, lacking broader investigation into the perception of AI in everyday tools such as recommendation systems, chat assistants, or translators (Gorska & Jemielniak, 2023). Furthermore, experts have looked at varied algorithms in online networks (Peralta et al., 2021); only a few studies investigate how things like age, gender, or tech know-how affect views on AI being fair and honest (Brauner et al., 2023). Addressing this gap is essential to align developers' intentions with user expectations and thus, fostering socially responsible AI systems.

This research seeks to investigate and understand how individuals perceive and experience bias in AI systems. In addition to examining demographic factors (e.g., gender and age), the research investigates areas such as work, daily life, and politics in terms of perceived bias. Another research objective is to examine how trust in AI systems or AI-generated information could shape a response to bias. Prior research suggests that individuals with higher levels of education and confidence working with technology demonstrate more awareness of bias in AI. Additionally, demographic factors (e.g., gender and/or age) shape perceptions of bias, and higher levels of trust in AI are associated with less critical perceptions of the existence of bias.

Despite the growing focus on AI ethics and fairness, we still have a considerable research gap about how regular users-not developers or experts- perceive, interpret, and experience bias in their day-to-day interactions with AI systems.

To build AI systems and models that can be trusted and are fair, there is a need to understand how people perceive AI bias (Peters, 2022; Brauner et al., 2023b). Addressing this gap is important because the current literature is largely technical or institutional and lacks an in-depth exploration of embodied users' perspectives of AI bias. The findings could help AI developers make systems that are easier to understand and more transparent and give policymakers better guidance on how to create rules that are clear and responsible (Bildirici, 2024). Additionally, teachers could use these insights to improve AI literacy, helping people learn to question and think more critically about the tools they use (Ifenthaler et al., 2024).

## 2. Literature Review

Bias in artificial intelligence has been assessed in numerous areas, including employment, healthcare, media, and everyday digital communication. Researchers have repeatedly highlighted that bias can be infused into AI systems at all stages, either due to biased training datasets, design decisions, or due to user interaction that served to reinforce a certain stereotype (Ntoutsi et al., 2020; Hou, Tseng, & Yuan, 2024). For example, image generation systems frequently depict fewer images of women in a professional context, thus solidifying a cultural stereotype (Gorska & Jemielniak, 2023). Similarly, biased employment algorithms have perpetuated inequities, exemplified by Amazon's hiring algorithm, which reversed gender equity for women (Hou et al., 2024). These examples emphasize that bias is not a part of "the technology" or a bug but, in fact, a systemic representation of the values that are built into data and design and that survey and enforce existing inequities surrounding data and intelligent systems.

In further developing this argument, Binns (2018) specifies that algorithmic fairness cannot simply be framed as a technical optimization problem. Based on theories of justice in political philosophy, he notes that fairness involves normative considerations about whose interests are given priority and whose values shape system development. Technical concepts of fairness, for example, equalized odds or demographic parity, fail to account for context and structure-based inequities that produce algorithmic outcomes. The process of reducing bias should then be taken to encompass both computational and ethical consequences and involve stakeholder engagement in participatory decision making and adjustment of the moral and ethical frameworks that inform AI development (Binns, 2018).

Trust has become an important topic for understanding how people think about biased AI. There is a general assumption that AI is neutral and objective, but public perceptions are much more complicated. More precisely, studies have found that people sometimes trust AI more than they would human beings because it is perceived as being agnostic and sometimes even justified to have a human doubting AI's accuracy when contradictory evidence is presented (Gerlich, 2024; Lee, 2018). People also weigh trust based on whether they are directly harmed or not when making such decisions. For instance, where individuals have been biased (for example, women applying for jobs that utilize algorithm-based systems), individuals would resist AI-provided recommendations or feedback, whereas others might, even the group in question, trust the recommendation (Hou et al., 2024). These findings align with your research in which participants demonstrated criteria of trust and a preference for human feedback compared to AI-based feedback for more consequential decisions.

Social media outlets complicate AI bias further by amplifying and normalizing prejudices via algorithmic filtering. Evidence suggests recommendation systems and content moderation are responsible for creating echo chambers, polarization, and reinforcing discriminatory narratives (Peralta, Kertész, & Iñiguez, 2021). The algorithmic approach privileges some perspectives and diminishes others, meaning algorithms do not simply reflect established biases; they also shape opinion formation at a system, or social, scale. Your survey responses reflected participants' concerns with regional, linguistic, and class-based biases, which often extend the structures of algorithmic curation in social networks. This implies bias is both a technical and social phenomenon that is interconnected to the platforms where people engage with AI daily.

Finally, explainability and transparency are generally considered essential strategies to mitigate AI bias. The literature indicates that when AI systems include interpretable explanations, users can better identify bias or unfairness and contest AI-based outcomes (Hou et al., 2024; Ghasemaghaei & Kordzadeh, 2024). In contrast, if an AI tool functions as a "black box," individuals may be prone to accept output without question and assume that the AI remains objective or unbiased. Your participants overwhelmingly supported transparency initiatives: most endorsed that AI should explain why it arrived at a certain decision. The consensus of findings reinforces the need for both technical responses, which involve approaches such as increasing dataset diversity and fairness metrics, in addition to sociotechnical approaches (e.g., regulation, education of users) for accountability and trust in AI systems.

## 3. Materials and Methods

### *3.1. Participants*

This study used the convenience sampling approach, which is suitable given the exploratory nature of this study, but carries limitations in terms of generalizability. A total of 66 individuals finished the survey, with a uniformity of 33 male and 33 female respondents. The participants' ages ranged from 13 years to 18 years and 19 years to 27 years, where 35 respondents were grouped in the 13 to 18 age category, and the other 31 respondents were grouped in the 19 to 27 age category. While the survey form offered additional gender options ("Other" and "Prefer not to say"), none of the respondents selected these categories.

### *3.2. Materials*

An online questionnaire, designed and hosted on Google Forms, was distributed to participants to be filled out. The survey was divided into four sections: (a) Demographics, (b) AI usage, (c) Perceptions of AI bias, and (d) Accountability in AI. The demographic section gathered information on gender, age, and educational attainment, which were then used as nominal variables in the statistical analysis.

The AI usage section explored the frequency of many AI systems (Siri, Alexa, ChatGPT, Gemini, Copilot, etc.) being utilized. The perceptions section inspected participants' awareness and recognition of AI bias, privacy concerns, and attitude towards bias in AI outcomes. Lastly, the accountability section explored participants' opinions on the responsibility and accountability of AI bias. The survey combined multiple-choice questions and open-ended questions.

### *3.3. Procedure*

The survey link was distributed mainly through social networks, peer groups, and direct sharing with friends via group chats. This method generated fast responses that remained consistent with the convenience sampling strategy. Participants accessed the Google Form from their own personal devices (mobile phones, computers, and tablets) and completed the survey at their own pace. No personal information was collected, and participants were free to withdraw from submitting the survey without any penalty or repercussions. Data was collected from 19th July 2025 to 31st July 2025.

### *3.4. Data Analysis*

Both descriptive and inferential statistical approaches were employed to analyse the collected responses. Data gathered via the Google Forms was structured, after which it was exported to DATAtab, an online platform designed for statistical analysis. The descriptive analysis included frequency distributions for all nominal variables; the findings were then visualized using the same software.

In terms of inferential analysis, a Chi-Square ($\chi^2$) test was performed to investigate the associations among several key nominal variables, including age group, gender, and educational attainment. A 95% confidence interval was used for the test ($\alpha = 0.05$) to determine if statistically significant links were present between the demographic variables and the respondents' perceptions of AI-related bias.

In the analysis, hypothesis testing was used to determine whether the differences between groups were statistically significant or random. A Chi-square process was selected because the nature of the independent variables was categorical, as reflected in the types of frequency distributions extracted.

Qualitative coding was done for open-ended responses to identify meaningful themes that grouped related answers together. Although this was not the main analytic focus, these codes yielded further insights and helped contextualize and understand the quantitative findings.

Overall, the methodology used a process of systematic data collection via an online survey and appropriate statistical tests (in this case, Chi-square analysis) along with data visualization.

## 4. Results and Discussion

The aim of this study was to examine how people with different demographic characteristics perceive bias in AI systems and whether demographic factors such as age, gender, and education are related to their perceptions.

Descriptive and inferential statistics were applied to the survey responses from 66 subjects about their attitudes towards AI bias, accountability, and transparency. The findings were then organized thematically by demographic characteristics (i.e., respondents were less than or greater than 30 years old, and/or female, male, or high school education)

### *4.1. Demographics*

A total of 66 participants filled in the survey, with an equal split between males (n = 33, 50.0%) and females (n = 33, 50.0%). The ages of the participants spanned 13 to 27 years, with 35 participants (53.0%) aged 13 to 18 years and 31 participants (47.0%) aged 19 to 27 years.

Regarding education, 28 participants (42.4%) were in secondary school, 26 participants (39.4%) were pursuing undergraduate studies, and 12 participants (18.2%) had completed or were pursuing postgraduate education. In general, the demographic distribution yielded a well-balanced data set that permitted hypothesis testing with acceptable precision and comparability between groups.

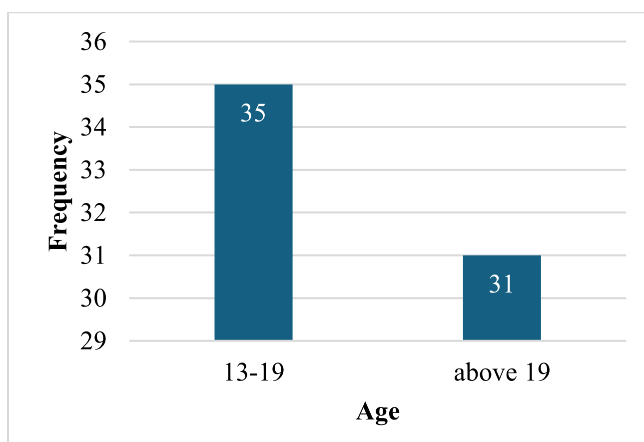


**Fig. 1 Frequency Distribution by Age**

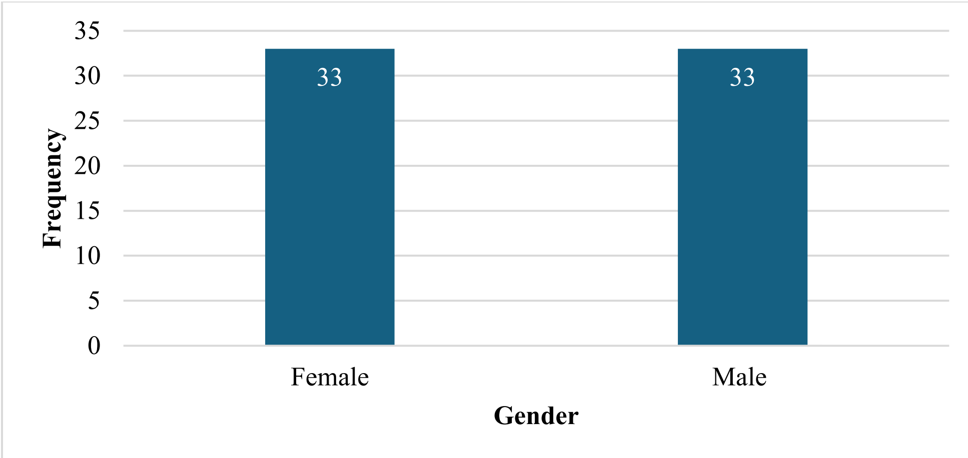


**Fig. 2 Frequency Distribution by Gender**

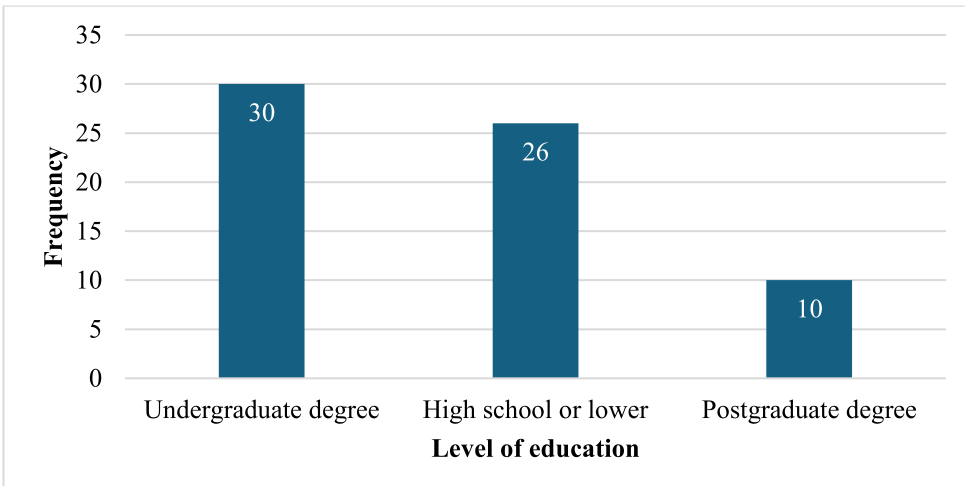


**Fig. 3 Frequency Distribution by Level of Education**

### *4.2. AI Usage Patterns*

The results indicate that 38 respondents used the AI tools frequently, suggesting a strong dependence on AI tools in their daily lives. A smaller group of respondents (9) reported using AI tools sometimes, while a very small number of respondents reported using AI tools rarely and very rarely.

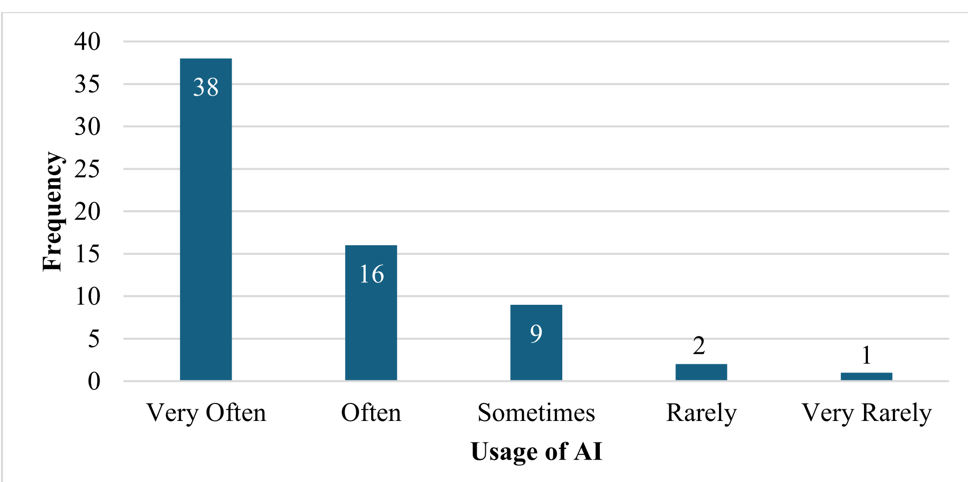


**Fig. 4 Distribution of usage frequency of AI**

In the context of platforms and tools, ChatGPT was the most popular with 58 respondents using it, followed by Gemini (27), Siri (20), Alexa (20), MetaAI (16), and Claude (16). There are some platforms mentioned as used but with lower reports (Perplexity, Google Assistant, and Copilot).

Overall, both general findings and findings related specifically to ChatGPT indicate a high rate of use of AI tools. Overall, the data indicates a high adoption rate, especially with ChatGPT, which reflects the predominance of this AI tool among the participants surveyed.

### *4.3. Awareness and Experience of AI Bias Usage Patterns*

The data shows a clear level of awareness of AI bias, with 31 participants reporting having experienced or witnessed biased or unfair behaviour demonstrated by AI. There were 19 respondents who were not sure, while only 13 respondents did not experience or witness biased behaviours. Furthermore, most participants were uncomfortable with AI outputs, with 28 of them reporting privacy concerns, 22 of them concerned about inaccuracy, while only 3 said they fully trust AI. Most participants either agreed (35) or strongly agreed (22) that AI systems can be biased.

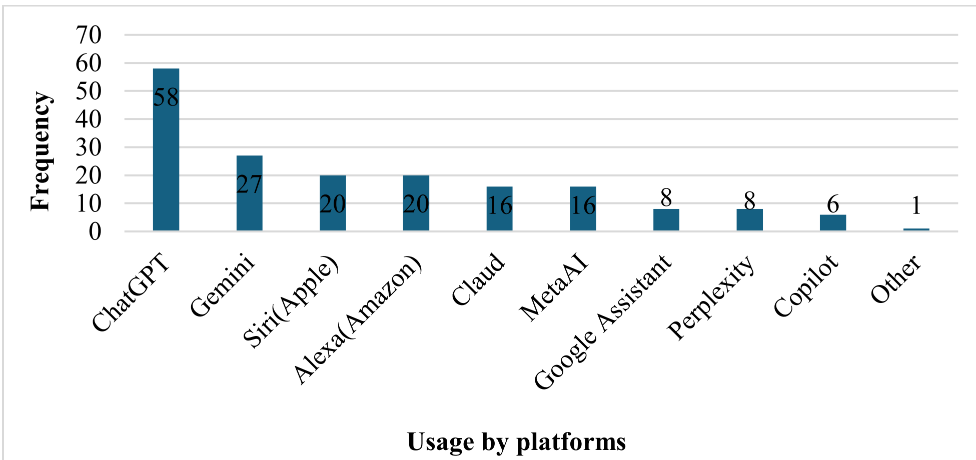


**Fig. 5 Frequency distribution of usage by platforms**

In terms of specific types of bias, 43respondents cited gender bias, followed by racial/ ethnic (35), regional (30), income/class (27), linguistic (23), socioeconomic (21), and caste bias (14). Most respondents agreed that the AI bias contributes to discrimination "to some extent" (36), with the most shared attitude toward bias found in AI systems being concerned, as 39 respondents were somewhat concerned, and another nine were very concerned. Overall, the data suggests that respondents are aware of AI bias and multiple kinds of bias and exhibit a clear level of concern for its effects.

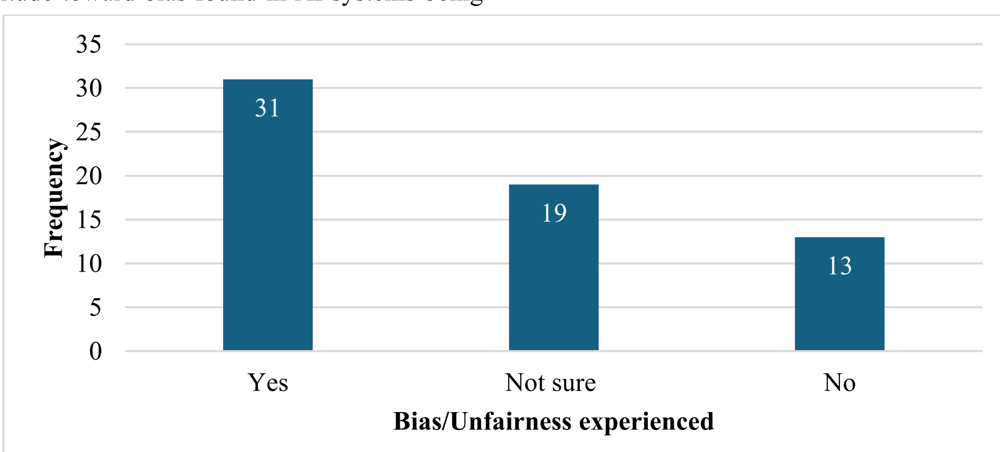


**Fig. 6 Respondents' Experience of Bias in AI**

### *4.4. Accountability and Responsibility*

The results suggest that most participants believe developers and programmers (29) are accountable for bias in AI, followed by companies that deploy the AI (14). Members noted government/regulators (9), users (9), or autonomous committees (2) as other possible stakeholders in accountability. In asking, however, about ways to address bias in AI, members were asked about solutions to identify the more important strategies, as the process may involve more than one strategy to address their concerns about bias. In terms of ways to address bias, participants overwhelmingly believed that increasing transparency and using diverse, representative training data were most important, followed by audits, education, ethical guidelines, and diverse teams, to ensure diminutions in bias in AI.

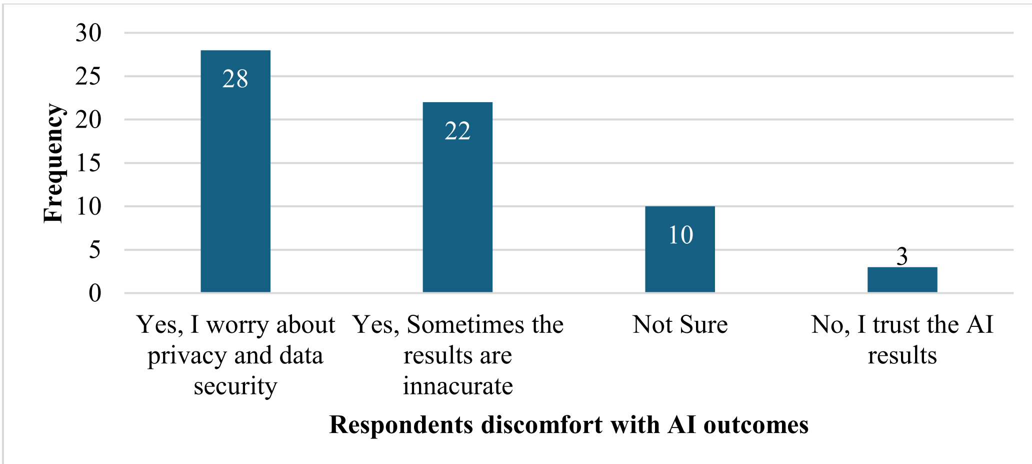


**Fig. 7 Respondents' Discomfort with AI Outcomes**

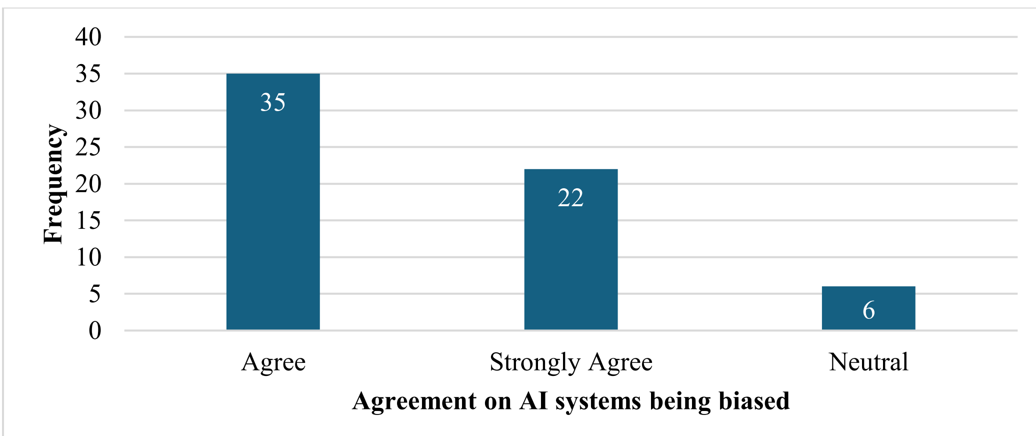


**Fig. 8 Level of Agreement on AI Systems Being Biased**

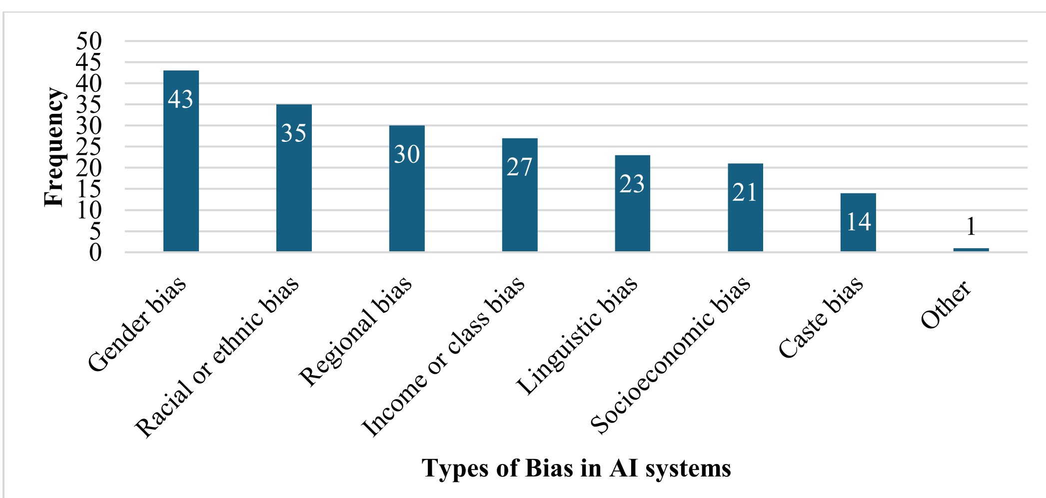


**Fig. 9 Types of Biases Identified in AI Systems**

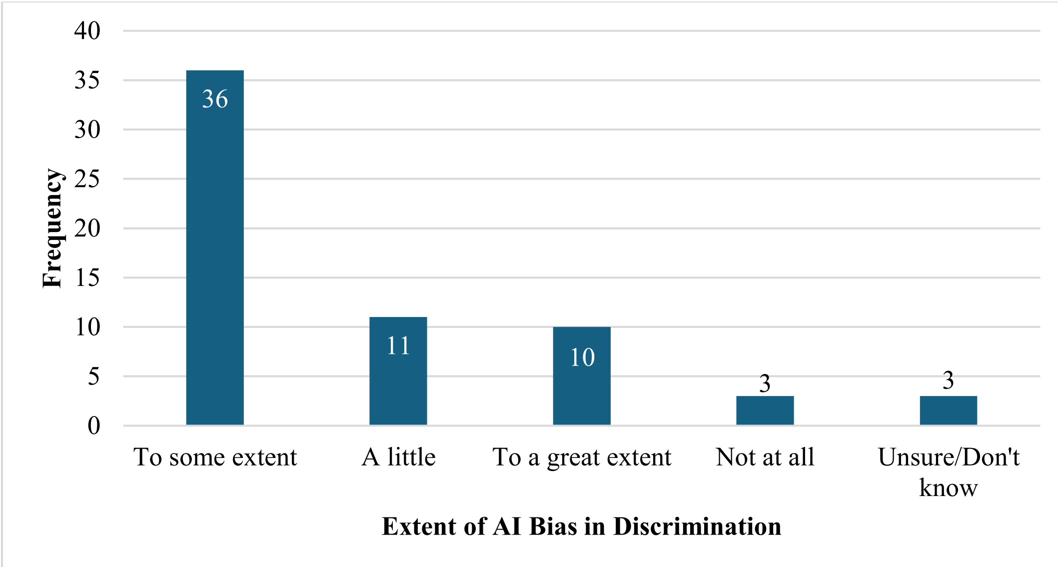


**Fig. 10 Perceived Impact of AI Biases on Discrimination and Stereotyping**

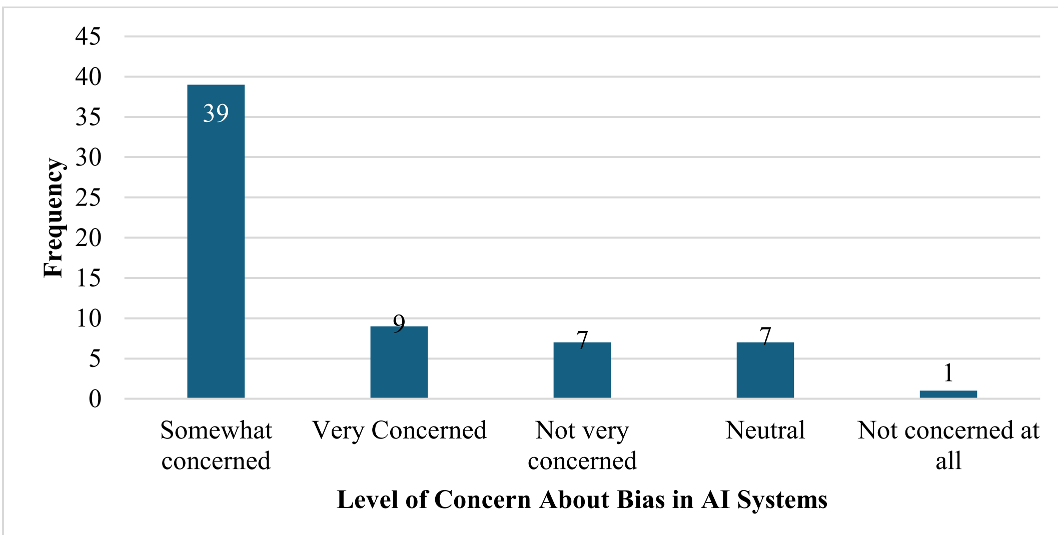


**Fig. 11 Level of Concern About Bias in AI Systems**

Participants also had a strong consensus that organizations should be accountable for biased AI decisions, with 58 respondents agreeing or strongly agreeing and only 8 expressing neutrality or disagreement.

Trust in decision-making was more contextual, as most (n = 52) claimed it "depends", while some responded in favour of humans (n = 9) and others responded in favour of the AI systems (n = 5).

In conclusion, the respondents expect accountability to fall on the developers and organizations deploying the system (after the fact), and respondents see transparency and diverse data as means to reduce biases, and there is situational trust between human and AI-based decision-making systems.

### *4.5. Age*

Participants' perceptions of AI bias were also affected by age. There was a significant relationship between age and how strongly participants believed AI bias could cause discrimination or stereotyping ($\chi^2$, p = .044); the younger and older groups differed in strength of belief. Age was also significantly related to agreement that organizations, whether private, government, or non-profit, developing or using AI, should be held accountable if the organizations' AI systems perpetrate bias ($\chi^2$, p = .018). Age was not significantly related to whether AI systems can be biased (p = .074), whether AI systems should explain how their decision-making processes resulted in perceived bias (p = .072), or who should address AI bias (p = .336). The data suggests that awareness of accountability draws differences between groups, while general awareness of AI bias remains constant across age groups.

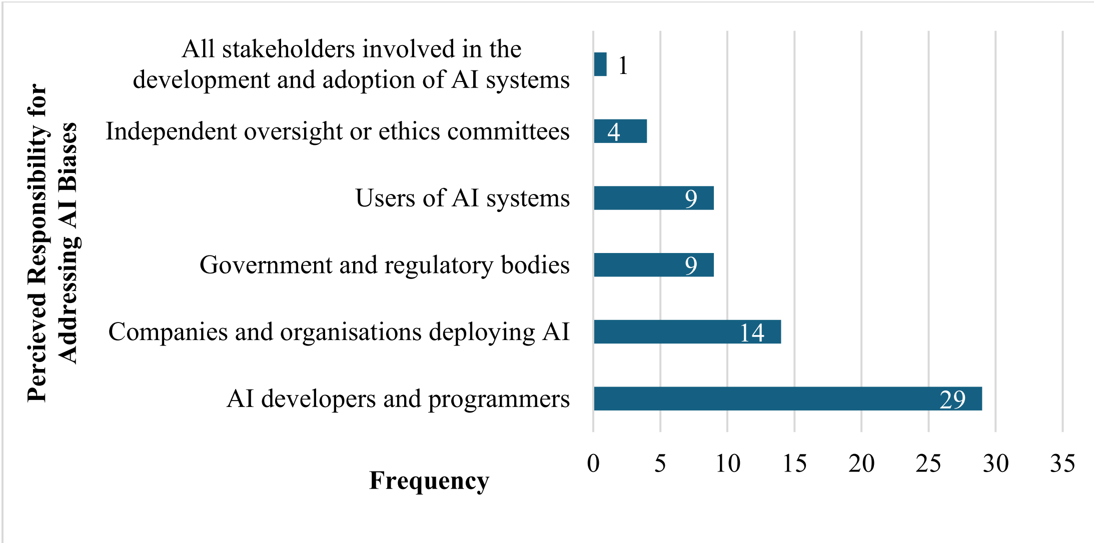


**Fig. 12 Perceived Responsibility for Addressing AI Biases**

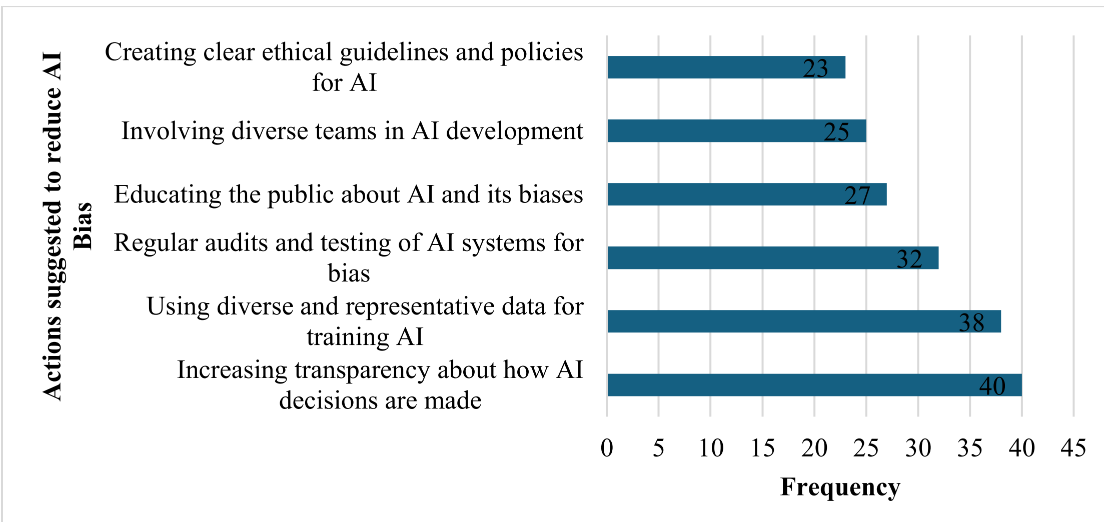


**Fig. 13 Actions Suggested to Reduce AI Bias**

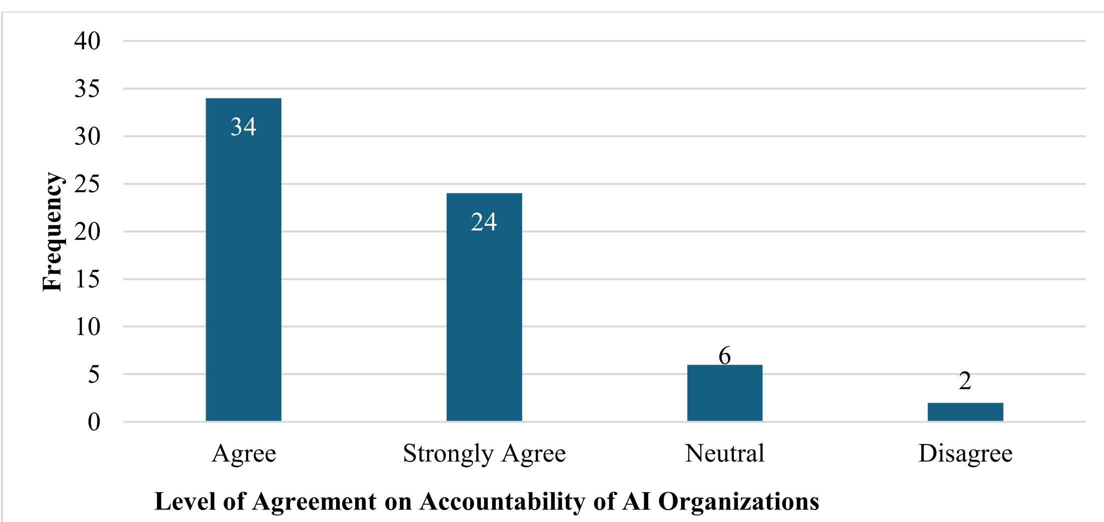


**Fig. 14 Level of Agreement on Accountability of AI Organizations**

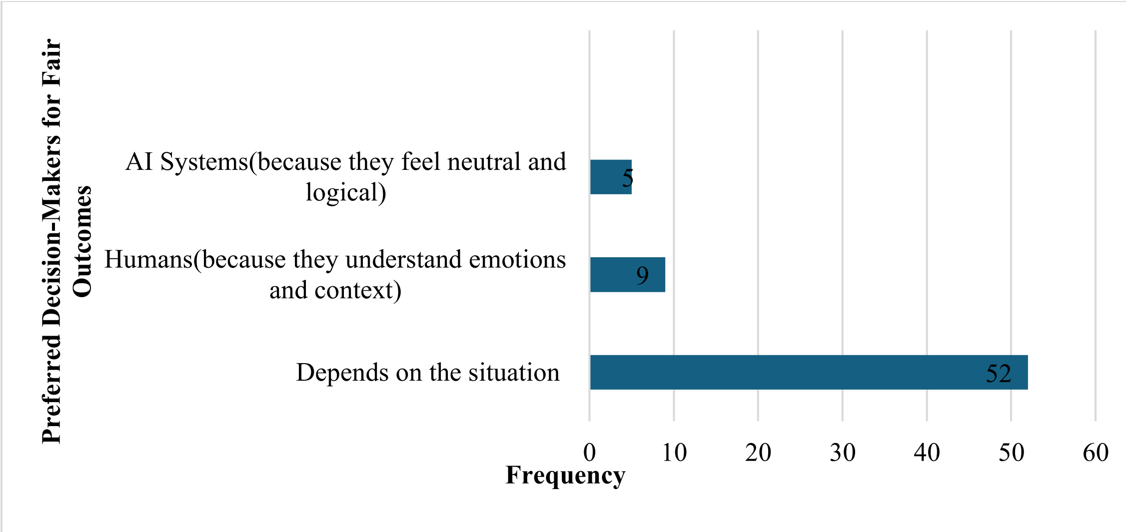


**Fig. 15  Preferred Decision-Makers for Fair Outcomes**

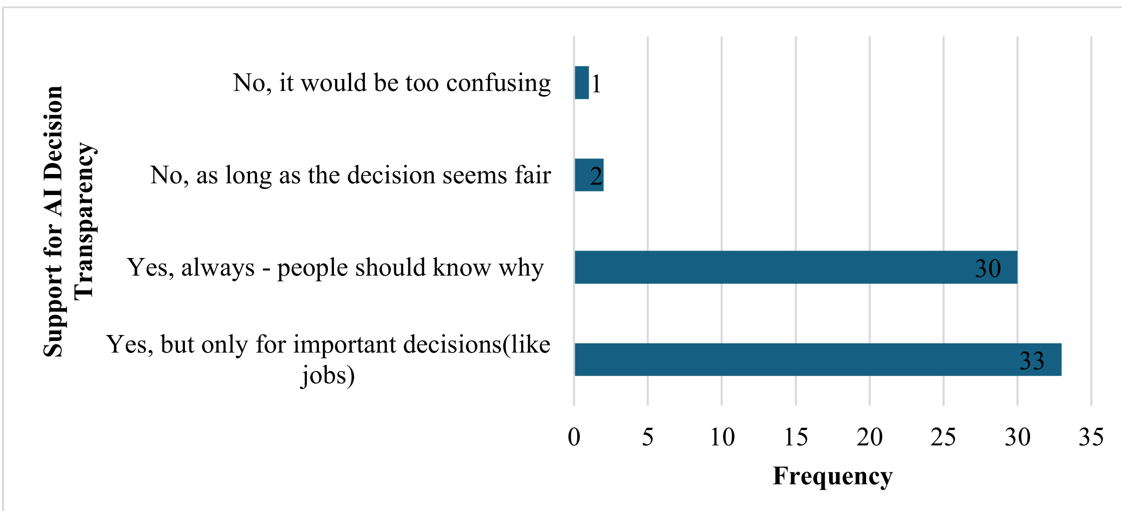


**Fig. 16  Support for AI Decision Transparency**

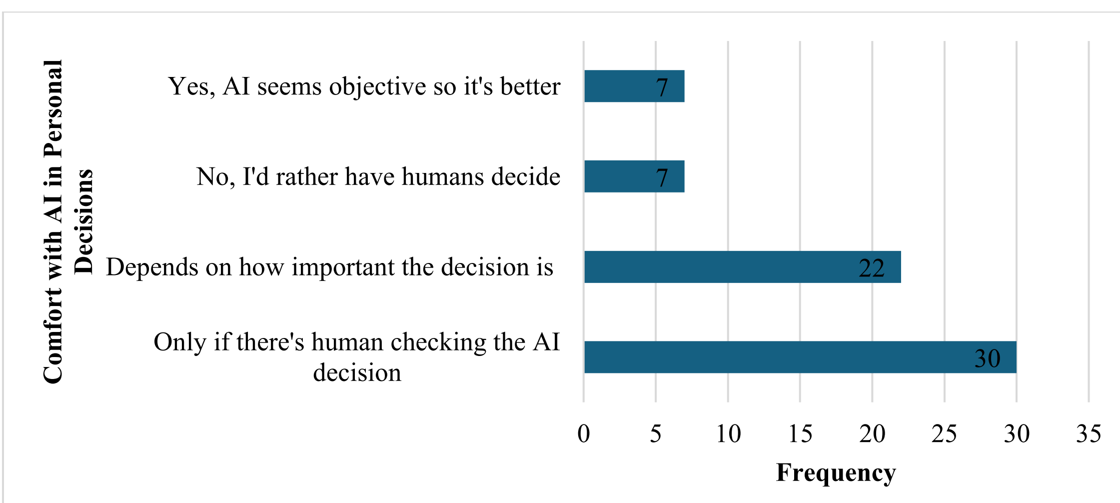


**Fig. 17 Comfort level with AI Making Personal Decisions**

### *4.6. Gender*

Some gender differences were found as well. There were major disagreements between the participants' perspectives on accountability, in which female and male participants varied in their belief that organizations should be criticized due to biased outcomes from AI ($\chi^2$, p = 006). Gender was also statistically significantly associated with attitudes about who should be responsible for correcting AI bias (p = 044) and if AI systems should explain their choices ($\chi^2$, p = 032). AI systems should explain their choices ($\chi^2$, p = .032). On the other hand, gender was not associated with the perception that any AI system can be biased (p = .118), and all other associations were not significant (p > .058), including the potential for AI biases to shape discriminatory or stereotypical behavior (p = .100). These results illustrate that both men and women were equally likely to recognize and disagree with AI bias, but that transparency and accountability expectations differed.

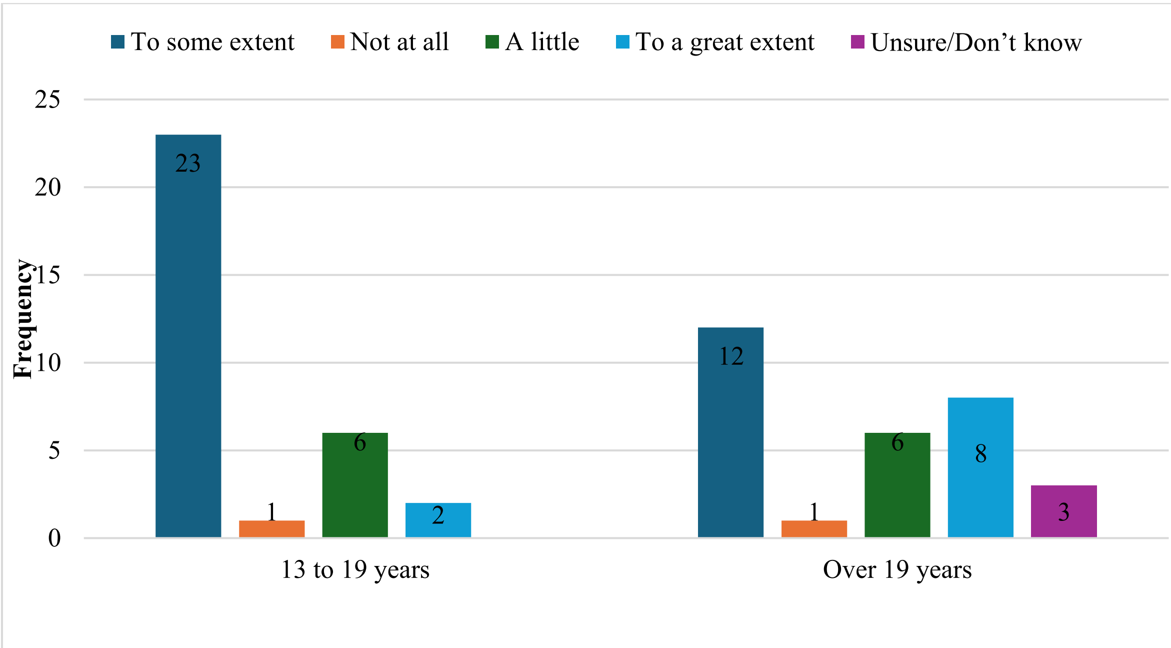


**Fig. 18 Perceived Contribution of AI Biases to Discrimination or Stereotyping by Age Group**

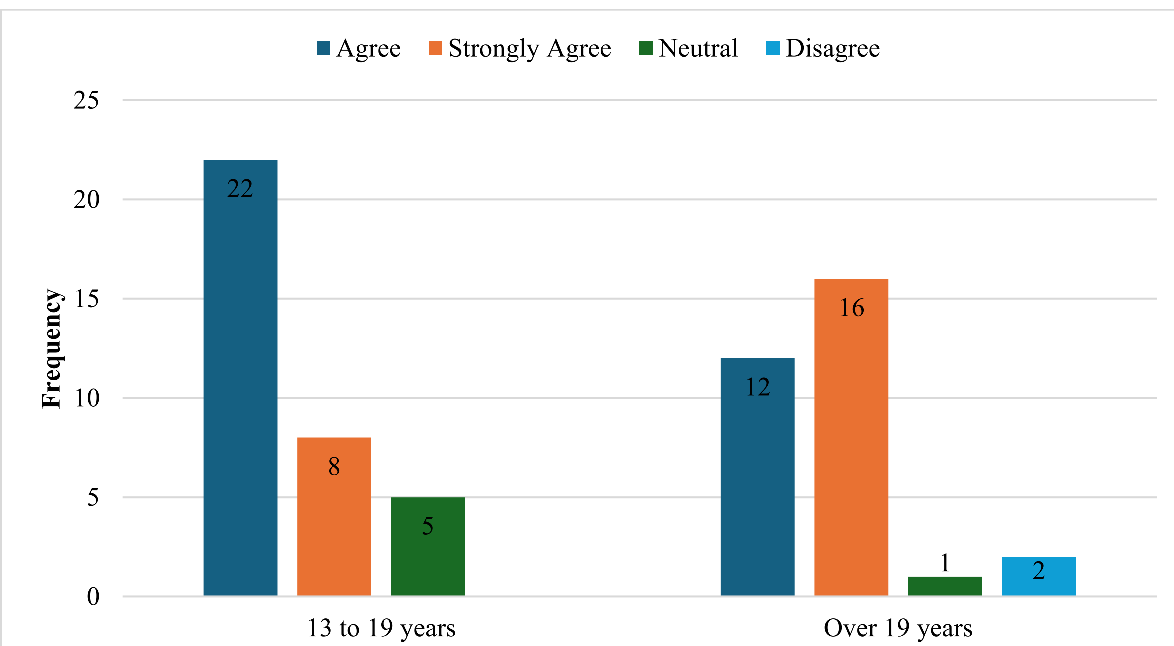


**Fig. 19 Agreement on Organizational Accountability for AI Biases by Age Group**

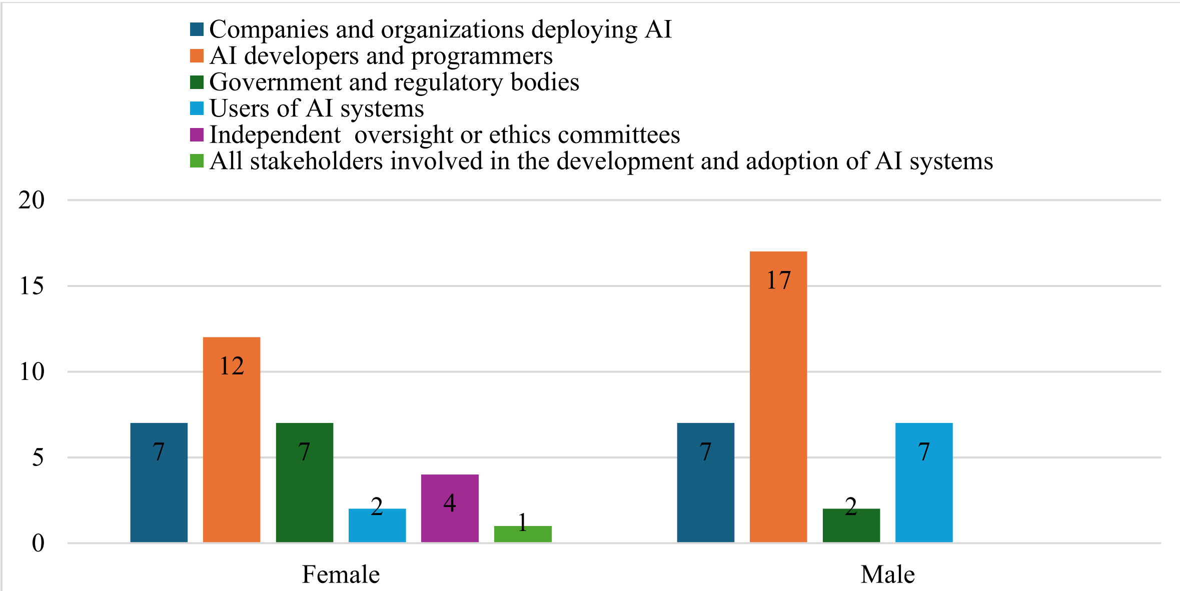


**Fig. 20 Perceptions of Responsibility for AI Deployment by Gender**

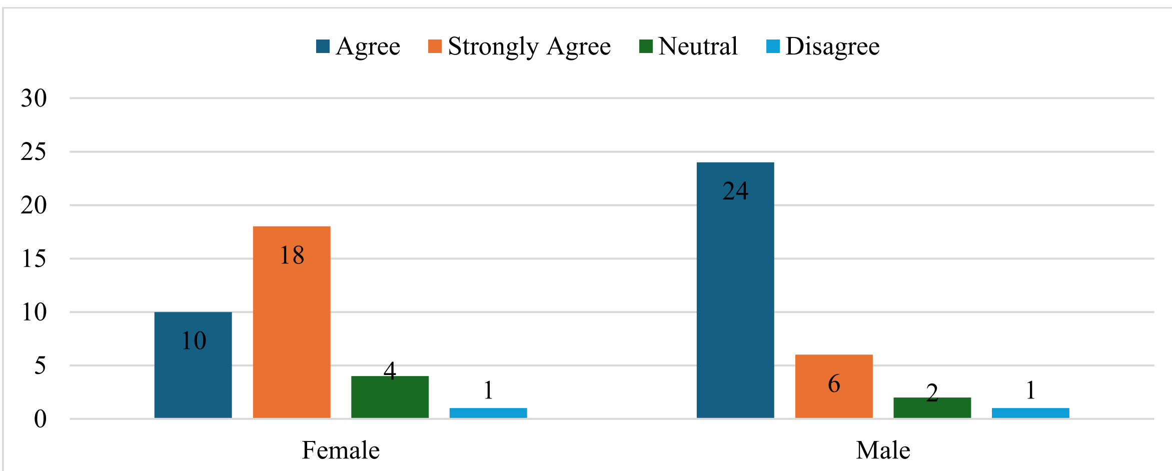


**Fig. 21 Gender Differences in Agreement on AI Transparency**

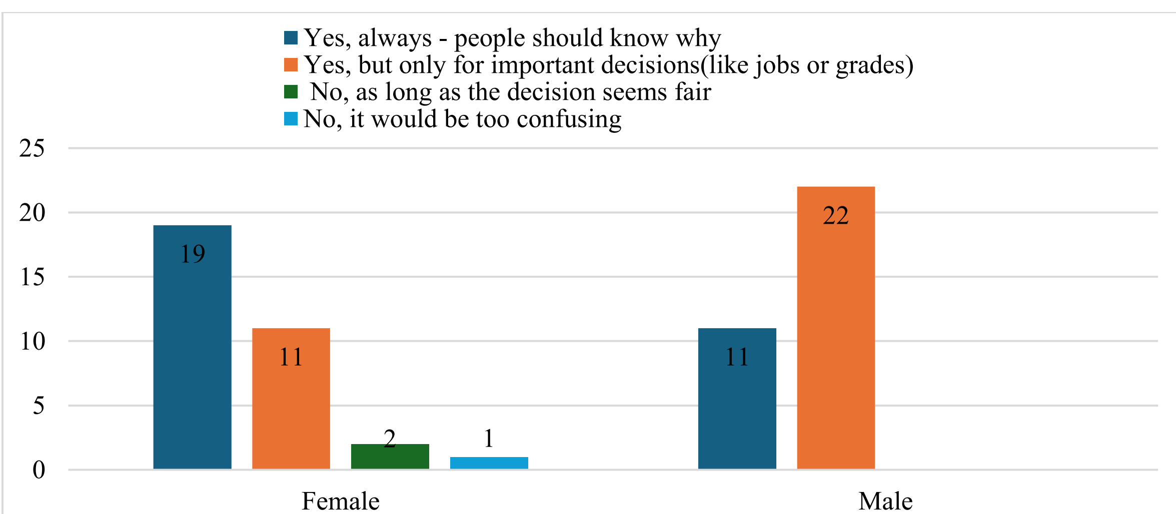


**Fig. 22 Gender-Based Preferences for AI Decision Explanations**

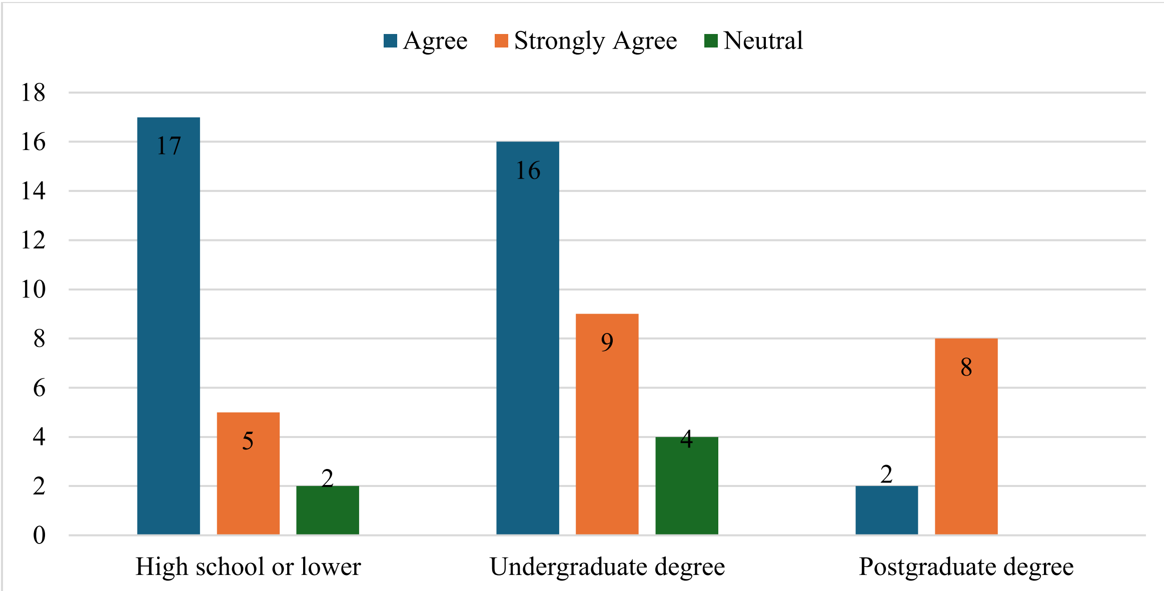


**Fig. 23 Educational Background and Agreement on AI Transparency**

### *4.7. Level of Education*

Educational attainment had less of an effect on participants' perceptions. While there was a relationship between educational attainment and believing the statement "AI systems can be biased" ($\chi^2$, p = .016), together respondents with higher education attainment tended to be more in agreement with the statement. However, educational attainment was not related to whether AI bias results in discrimination (p = .728), accountability of organizations (p = .260), who is responsible for AI bias (p = .317), or whether an AI system should explain its decision-making (p = .638).

Thus, while education appears to have some influence in developing awareness of bias, it does not greatly determine face-value opinions about responsibility or transparency. It should also be noted that the postgraduate group (n = 10) was relatively small compared to the undergraduate (n = 30) and high school or lower (n = 26) groups, which may limit the strength of comparisons across education levels.

### *4.8. Discussion*

The primary objective of this study was to understand how people conceptualize and operationalize bias in AI systems with a focus on accountability, transparency, and trust. The findings show a general acknowledgment that AI can indeed generate bias, and that participants were generally uncomfortable with AI-based bias. Participants mostly agreed that companies either developing or using AI should be held responsible, and they strongly emphasized accountability for transparency and explainability of AI. This led credence to citizens' expectations that both AI systems must be trustworthy ethically, in addition to being trustworthy functionally.

This study's findings are consistent with recent literature suggesting that the public's engagement with AI is not purely based on its technical performance, but rather that they engage with AI in terms of perceptions of fairness and trustworthiness (Brauner et al., 2023; Liu & Xu, 2024). Participants demonstrated strong ethical expectations. Further, people expect organizations to manage harms caused by AI responsibly. The population sampled indicated broadly that they recognized that AI systems can be biased (Figure 8). For the most part, participants strongly agree or agree with the statement "AI systems can be biased," a finding that fits with studies that show AI users do not see AI as neutral and see it as having gone through human values (Ntoutsi et al., 2020; Rasali et al., 2024). Many participants have experienced bias or observed bias (Table 6), which, again, supports findings of growing concern in AI being used to discriminate in real-world situations, such as hiring practices, healthcare practices, and criminal justice, determining guilt and/or innocence (Simonite, 2019; Mattu, 2023).

Participants experienced a range of concerns about an AI-biased output in general, with discomfort coming more from consequences in part, in relation to serious or high-stakes scenarios, like higher levels of discomfort were related to engaging an AI system to help decide impact jobs, education, and financial matters. More than 60% indicated they were uncomfortable with AI deciding something of interest could possibly impact your life (Figure 17), matching findings by HR (Callahan, 2023; Madsen & Toston, 2025). The discomfort can be amplified because it relates to the place of AI in possibly "overthinking" or being human-like, such as a healthcare algorithm that continually deprioritizes the needs of a Black patient (Obermeyer et al., 2019).

With respect to trust in AI, participants largely perceived AI to be a tool for everyday practicalities, with trust not being offered for significant decision-making, suggesting they understand AI to be an assistant, not a replacement (Hou et al., 2024). Organizational accountability was a major theme, with more than 70% of participants agreeing that developers and users have a shared responsibility for bias and harm (Figure 14). This echoes the literature that suggests accountability as a major standard in AI ethics (Ethics of (AI and Robotics, 2020; Pratt, 2020).

Transparency and explainability were also important. Almost all the participants agreed that AI could and should be able to generate explanations for its decisions (Figure 16). The group had three suggestions for how to reduce bias: improving data quality, confirming diversity in training sets with representation of many social categories, and adopting a multi-disciplinary approach to bias reduction, again reflecting calls for both technical and sociotechnical approaches (Bonarini, 2022; Gorska & Jemielniak, 2023). Demographic differences informed some aspects of AI perceptions. Age differences were significant for perceptions of AI-induced discrimination or stereotyping ($p = .044$) and expectations of organizational accountability ($p = .018$). Older participants (19–27 years) were more likely to demand accountability than younger participants (13–18 years), consistent with research suggesting older adults demonstrate greater ethical concerns regarding technology governance (Wang et al., 2022; Brauner et al., 2023).

Gender was positively associated with attitudes toward accountability ($\chi^2$, $p = .006$), responsibility for addressing bias ($p = .044$), and expectations of explainability ($p = .032$). Although awareness of bias was equal between male and female participants, females scored higher on demanding accountability and transparency, consistent with literature noting that women often exhibit greater sensitivity to discrimination due to personal experiences and socialized ethical responsibility (Gender Bias in AI, 2020).

Education influenced perceptions of bias awareness but not normative expectations. Higher education was associated with stronger agreement that AI systems can be biased ($\chi^2$, $p = .016$), but it did not significantly affect views on discrimination, accountability, or transparency. This suggests that education may improve technical understanding of bias but does not automatically translate into ethical expectations, highlighting the importance of educational initiatives that integrate both technical literacy and ethical reasoning (Bildirici, 2024).

In conclusion, there was widespread awareness about bias in AI; however, there was variability between demographic groups in their expectations of accountability and transparency. Younger students were aware of bias but did not place many expectations on organizations being responsible for it, unlike older participants and female participants, who emphasized accountability. More educated participants were more aware of bias but did not have any strong expectations as it related to ethics as well.

These findings have a significant set of recommendations, first they show the necessity of inclusive frameworks of AI governance that consider a range of user perspectives across the AI supply chain (Nazer et al., 2023), designers for instance should build AI systems with an emphasis on fairness and transparency, instead of just addressing compliance with regulation, in order to foster user trust. Secondly, initiatives specifically meant to educate policymakers on AI literacy should place an emphasis on how to develop a technology literacy mixed with technical, social, and ethical attributes to support critical examination of AI-derived decisions while mindful of algorithmic opacity. Third, solutions that consider hybrid models of decision making are practical, where AI produces more efficient decision options for the human user, but human users still support fairer and ethical reasoning (Manyika et al., 2019).

This research is novel in its integrative consideration of how demographic variables such as age, gender, and education shape ethical expectations of AI. In connecting perceptions of bias with accountability and transparency concerns, this study offers a more nuanced understanding of how trust is socially and ethically constructed in relation to AI.

This research builds upon existing research, and unlike previous research that focused mostly on general feelings towards AI ethics, this work will examine more nuanced and comparative patterns across demographic groups. It expands upon previous research in that it connects data based on perceptions as well as a quantitative analysis to provide a better understanding of how ethical awareness emerges due to age and gender. It also provides much-needed participant data from younger participants, a group that has been inadequately explored in AI bias research and is a key gap in the literature. Overall, it should be noted that addressing AI bias does not only mean a better technical solution, but also improving datasets or algorithms. Rejecting bias in AI systems will also require multi-dimensional approaches that include technologists, ethicists, policymakers, and users (Ali et al., 2021; Okidegbe, 2023). Biased AI systems are likely to replicate and amplify existing social inequities at scale without a combination of this approach.

## 5. Limitations of the Study

Although this study provides interesting insights about perspectives regarding bias in AI, it has several limitations. Convenience sampling enables the collection of data but limits the generalizability of the findings. Likewise, the sample size was small, which limited statistical power and the ability to observe nuanced effects in the data.

Another limitation was the disproportion of younger participants in the sample. This sample with a large proportion of younger participants may have shaped the study findings, as younger individuals tend to engage more often with AI (e.g., social media, personal assistants), compared to older adults, who may be engaged with AI more often in their health care experiences or in financial services. This disproportion reduces the study's ability to capture perspectives of all ages.

## 6. Suggestions for Future Studies

To address these constraints, future research should focus on participants who are noticeably different in age, cultural contexts, or socioeconomic status to further develop one's knowledge of AI bias specific to different groups. An example of this difference would likely be seen between younger participants who might engage with an AI system via an ongoing daily communication versus older persons who engage with AI systems as part of their healthcare/finance routine. More broadly, diverse demographic groups will strengthen the findings and their use of the findings.

The use of longitudinal studies would also be an important advancement. Individuals' viewpoints of AI are not fixed and develop as technologies evolve and as they operate within AI system environments. Following participants over time would suggest the dynamics of trust or skepticism developing, increasing, or changing in response to their exposure to AI, regulation, or specific news about bias in an AI system.

The consideration of cultural and socioeconomic contexts would also be warranted. Cultural contexts shape how individuals define fairness and trust to understand AI bias perception better, and socioeconomic status determines which AI systems individuals may engage with and how much they may be affected by a biased outcome. A better understanding would inform the design of equitable and cognitive AI system solutions.

Overall, engaging in interdisciplinary work could advance this line of research. Future research may go beyond documenting perceptions, based on insights gained from sociology, psychology, and computer science, by designing a test of an intervention that addresses bias and builds trust. This research engages literature on both empirical academic research and practice to inform the design of AI systems to be fairer, more transparent, and more inclusive.

## 7. Conclusion

This study shows that although younger users are aware of AI bias and its discriminatory potential, age and gender differences can reveal mixed expectations about accountability and trust. Age and gender were significant influencers on perceptions of accountability, with older participants more likely to desire organizations to be held accountable and women more likely to hold transparency in a higher regard. Nevertheless, participants almost universally agreed that the AI bias contributes to discrimination and that the developer and organization have responsibility for their actions. Awareness itself is not enough, and what is warranted is a system to hold developers and organizations accountable.

The broader implication is that AI bias is not only a technical challenge but also a social and ethical one. Building fair and trustworthy AI requires a multidimensional approach: improving datasets and algorithms, designing transparent systems, and fostering AI literacy among users. By situating individual perceptions within broader debates on fairness, trust, and accountability, this study contributes to the ongoing call for inclusive governance frameworks. Only when technological innovation is coupled with ethical responsibility can AI systems avoid reproducing structural inequalities and instead support equitable social progress. Nevertheless, only if technical innovation were paired with some sense of ethical responsibility will AI systems be kept from themselves perpetuating structural inequality and working toward equitable social progress.

## References


[1] Raymond S. T. Lee, *Artificial Intelligence in Daily Life*, Springe, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Hanguang Xiao et al., “A Comprehensive Survey of Large Language Models and Multimodallarge Language Models in Medicine,” *Information Fusion*, vol. 117, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Chenjun Liu, and Yan Xu, “A Model for Evaluating the Effectiveness of News Dissemination under the Combination of Big Data and Artificial Intelligence,” 3rd International Conference on Data Analytics, Computing and Artificial Intelligence, Zakopane, Poland, pp. 606-611, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Lama H. Nazer et al., “Bias in Artificial Intelligence Algorithms and Recommendations for Mitigation,” *PLOS Digital Health*, vol. 2, no. 6, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Drona P. Rasali et al., "Cross-Disciplinary Rapid Scoping Review of Structural Racial and Caste Discrimination Associated with Population Health Disparities in the 21st Century," *Societies*, vol. 14, no. 9, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Vittoria Scatiggio, "*Tackling the Issue of Bias in Artificial Intelligence to Design AI-Driven Fair and Inclusive Service Systems. How Human Biases Are Breaching Into AI Algorithms, With Severe Impacts On Individuals And Societies, And What Designers Can Do to Face this Phenomenon and Change for the Better*," Thesis, pp. 1-84, 2022. [Google Scholar] [Publisher Link]

[7] Haroon Sheikh, Corien Prins, and Erik Schrijvers, "AI as a System Technology," *Mission* AI, pp. 85-134, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kwadwo Asante, David Sarpong, and Derrick Boakye, "On the Consequences of AI Bias: When Moral Values Supersede Algorithm Bias," *Journal of Managerial Psychology*, vol. 40, no. 5, pp. 493-516, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Xiaoyu Zhu et al., "Algorithm and Analytical Verification of Roller Straightening Process Model Considering Stress Inheritance Behavior," *AIP Advances*, vol. 15, no. 3, pp. 1-10, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Priyansh, and Amrit Kaur Saggu, "Building Trust in AI Systems: A Study on User Perception and Transparent Interactions," *International Journal on Science and Technology*, vol. 16, no. 1, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Fatih Bildirici, "Open-Source AI: An Approach to Responsible Artificial Intelligence Development," *Reflektif Journal of Social Sciences*, vol. 5, no. 1, pp. 73-81, 2024. [Google Scholar] [Publisher Link]

[12] Philipp Brauner et al., "What does the Public Think about Artificial Intelligence?—A Criticality Map to Understand Bias in the Public Perception of AI," *Frontiers in Computer Science*, vol. 5, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Dirk Ifenthaler et al., "Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners," *Technology, Knowledge and Learning*, vol. 29, pp. 1693-1710, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Shengnan Han et al., "Aligning Artificial Intelligence with Human Values: Reflections from a Phenomenological Perspective," *AI & Society*, vol. 37, pp. 1383-1395, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Andrea Papenmeier, Gwenn Englebienne, and Christin Seifert, "How Model Accuracy and Explanation Fidelity Influence user Trust," *arXiv preprint*, pp. 1-7, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Anna Fine, Emily R. Berthelot, and Shawn Marsh, "Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice," *Behavioral Sciences*, vol. 15, no. 4, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Teresa Sandoval-Martin, and Ester Martínez-Sanzo, "Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL·E and Bing Image Creator," *Social Sciences*, vol. 13, no. 5, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Andrea Ferrario, and Michele Loi, "How Explainability Contributes to Trust in AI," *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Seoul Republic of Korea, pp. 1457-1466, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Florian Pethig, and Julia Kroenung, "Biased humans, (Un)Biased Algorithms?," *Journal of Business Ethics*, vol. 183, pp. 637-652, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Philipp Brauner et al., "Misalignments in AI Perception: Quantitative Findings and Visual Mapping of How Experts and the Public Differ in Expectations and Risks, Benefits, and Value Judgments," *arXiv preprint*, pp. 1-27, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Kasper Trolle Elmholdt et al., "The hopes and Fears of Artificial Intelligence: A Comparative Computational Discourse Analysis," *AI & Society*, vol. 40, pp. 4765-4782, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] M. Callahan, Algorithms were Supposed to Reduce Bias in Criminal Justice—Do they?, Boston University Today, 2023. [Google Scholar]

[23] Ethics of Artificial Intelligence and Robotics, Stanford Encyclopedia of Philosophy, 2020. [Online]. Available: https://plato.stanford.edu/entrieS/ethics-ai/

[24] Ayesha Nadeem, Babak Abedin, and Olivera Marjanovic, "Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies," *Proceedings of the Australasian Conference on Information Systems*, pp. 1-12, 2020. [Google Scholar] [Publisher Link]

[25] James Manyika, Jake Silberg, and Brittany Presten, "*What do we do about the Biases in* AI?," Harvard Business Review, pp. 1-5, 2019. [Google Scholar] [Publisher Link]

[26] Jeff Larson et al., How we Analyzed the COMPAS Recidivism Algorithm, ProPublica, 2016. [Online]. Available: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[27] Eirini Ntoutsi et al., "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, vol. 366, no. 6464, pp. 447-453, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Mary K. Pratt, 5 Ways AI Bias Hurts Your Business, TechTarget Enterprise AI, 2021. [Online]. Available: https://www.techtarget.com/searchenterpriseai/feature/5-ways-AI-bias-hurts-your-business

[30] Maximilian Kasy, "*IZA DP No. 16944: Algorithmic Bias and Racial Inequality: A Critical Review*," IZA Discussion Papers, Institute of Labor Economics (IZA), pp. 1-27, 2024. [Google Scholar] [Publisher Link]