Innovations

Bias Perceptions of Generative AI: A Comparison of Public Sentiment on Twitter and Workplace Perspectives

[1]Evans Uhunoma; [2]Maryann Asemota
[1]Independent Researcher, United Kingdom
[2]National Research University Higher School of Economics, Moscow, Russia

[1]ORCID ID: 0009-0008-0357-4276; [2]ORCID ID: 0000-0002-9618-4056

Abstract: *Generative Artificial Intelligence (Gen AI) is perhaps one of the most significant technological inventions in the last decade.It enhances content generation across various domains, from personal messages to work-related tasks, encompassing text, images, and videos. However, there have also been several debates surrounding its inadvertent risks for bias and perpetuating stereotypes (Ferrara, 2023; Xavier 2024) both from gender and racial perspectives (Nicoletti & Bass, 2023; Zhou, 2024; Sadeghiani, 2024).Today, Gen AI is also being used in the workplace, with many organisations adopting custom-built Gen AI tools as part of their working tools and systems. Consequently, the objective of this research was tofind out if employees perceive work-related outputs of Gen AI to be biased against women, people of colour or neurodiverse people, and how this perception compares to that of the public, who use Gen AI tools for both work and non-work-related purposes. A mixed-methods approach was employed in this research. From workplace perspectives, quantitative data were collected using a structured questionnaire from UK employees. From the public perspective, qualitative data were collected through text mining using specific keywords from Tweets on X platform. Findings showed that, while workplace respondents reported modest levels of perceived bias across all groups, public sentiment analysis and themes showed significant mistrust and negative perception of bias in generative AI outputs for women and people of colour. There was however positive perception as it relates to Neurodiverse people, with the public data showing positive sentiments for Gen AI outputs as it relates to Neurodiverse people, as the users view it as a tool for helping dyslexic users communicate better.*

Keywords: *Generative AI, Gen AI, AI Bias, Workplace Perception, Sentiment Analysis, Social Media*

Introduction

Generative Artificial Intelligence (Gen AI) is perhaps one of the most significant technological inventions in the last decade. It enormously benefits businesses and individuals, with small benefits like notetaking or email creation and more significant benefits like complex task completion or image/video generation (Baidoo & Ansah, 2023; Chan & Hu, 2023). However, Gen AI has also become a major talking point in the world today, with various debates surrounding its inadvertent consequences and potential risks (Epstein et al., 2023; Chan & Hu, 2023; Ferrara, 2023; Xavier 2024). Consequently, new laws are even being created to ensure that its use and application are regulated and there are stringent requirements before systems with Gen AI are deployed to end-users (Smith & Cankert, 2024). One of the many reasons for this is that the outputs or answers that Gen AI tools provide to users can unintentionally reveal the biases that are inherent in the data used in training those Gen AI tools (Hall, 2024). Ferrara (2023) and Smith and Canker (2024) also found out that Gen AI has the potential for algorithm bias and systemic risks.

Problem statement

Over the years, a few studies such as Brynjolfsson and Raymond, 2023; Zhou and Lee, 2024; and Klingbeil et al, 2024 have tried to analyse the pros and cons of Gen AI usage in the world today, in terms of potential for bias and discrimination. A few other studies have specifically focused their study on AI bias relating to Gender (Nadeem et al. 2020; Nicoletti and Bass, 2023;Sadeghiani, 2024), Race (Nicoletti and Bass, 2023; Zhou et al., 2024; Sadeghiani 2024).

However, there is still limited understanding in terms of the empirical evidence on whether Gen AI tools reinforce stereotypes and bias relating to women, people of colour and neurodiverse people – an area which has not been focused on so much. Also, some studies have taken time to understand if different users or demographic groups perceive bias in answers received from Gen AI tools differently (Allan et al., 2024; Sadeghiani 2024). Nevertheless, this study aims to go deeper, specifically looking at the perception of users, who use Gen AI mostly for work purposes only (including employees using enterprise/ custom-build Gen AI) and how it may differ from the public who use Gen AI for varying, and different purposes – work-related or personal.

Research Objectives

To find out if employees perceive work-related outputs of Gen AI to be biased against women, people of colour and neurodiverse people, and if this is the same perception from the public using Gen AI for both work and non-work-related purposes.

Significance of Study

By offering insights from two different perspectives—workplace users and the general public – this study advances our understanding of how bias is perceived in Gen AI outputs. Firstly, this study offers a detailed understanding of how employees perceive generative AI bias, especially with regard to outcomes pertaining to women, people of colour, and neurodiverse individuals. Secondly, this study captures the broader sentiment of generative AI from the public perspective through social media posts (organic opinions). Lastly, this study looks at whether public opinion and workplace views are in line or not, highlighting similarities and differences between work-related and non-work-related use of Gen AI and its outputs.

Literature Review

Introduction

This section discusses relevant generative AI bias literature, including previous research that has been carried out relating to the subject, the concept of perception from workplace and public perspectives, and focus areas of content generation for work and non-work-related usage. This section also considers training data for public and enterprise generative AI tools, as well as relevant theories.

Generative AI and Bias

Generative AI is a branch of Artificial intelligence. While Artificial intelligence has existed since the 60s, Gen AI (a subset of AI) has become a relatively new concept only in the last 3 to 5 years (Hall, 2024; Zhigalko, 2024). Gen AI, which has over the years now built on the advances made in the AI field in general, is known to be able to create novel outputs. Examples of these outputs such as textual contents, images and sounds are generated through large-scale machine learning models trained on vast datasets to handle user inputs and use patterns discovered in the training data to forecast and provide answers that are logical and pertinent to the context (Mhlanga, 2023; Google Cloud Skills Boost, 2024). However, using generative AI tools has its own set of operational, ethical, and legal issues that need to be carefully considered and handled.

One of the major issues is its tendency to be biased and perpetuate stereotypes, due to the dataset used in training the model; which is inherently biased, and then ultimately impacts the outputs or responses received (Zhou et al., 2024). Given, that many Gen AI systems heavily rely on algorithms to make judgments (Fu et al., 2020), bias can then occasionally arise from systematic and repeatable mistakes made by the system, which may lead to unfair results like giving one category or demographic preference over another in ways that deviate from the algorithm's original purpose (Baker & Hun, 2022). For example, in a recent text-to-image Gen AI research conducted by Nicoletti and Bass (2023), they found out that, due to the training data used to train a text-to-image system, the system generated more images that perpetuate racial and gender stereotypes. They consequently concluded that the impact training data could have on output seen in Gen AI tools "…doesn't just replicate stereotypes or disparities that exist in the real world — it amplifies them to alarming lengths".

Training Dataset for Gen AI Platforms

Large and varied datasets, like text, photos, or other media from sources like books, the internet, and social media, make up training data for large language models (LLMs) or generative AI systems as we popularly call them. Through training the model with these datasets, the model learns linguistic patterns, context, and associations from this input, which helps it produce outputs that are logical and pertinent to the context (Nah et al., 2023).

While most people use public AI models like ChatGPT, Gemini, Claude and various others for their day-to-day Gen AI content generation, many organisations have started to invest heavily in creating their own custom-built, organisation-specific and enterprise-inclined Gen AI tools to be used solely for work and company related purposes (Blessing; 2023; Cook et al., 2024; Lishchynska, 2024). These enterprise-specific Gen AI tools are trained on the unique data of the company, including confidential information, internal papers, and client interactions. This enables the model to produce relevant results in line with the specific requirements and preferences of the company (Davenport & Alavi, 2023). This is different from public AI models like ChatGPT, Gemini and co which are trained on vast amounts of publicly accessible text and code, which may result in outputs that are less suited for particular organisational contexts but also provide more general knowledge. Below is a table summarising the differences between public and enterprise AI models.

Table 1: Key Differences Between Public and Enterprise-Gen AI Models

Feature	**Enterprise AI**	**Public AI Models**
Training Data	Organisation-specific data	Publicly available data
Customization	High	Limited
Control	Greater	Less
Data Privacy & Security	Higher	Lower
Use Cases	Specific to organisation	General-purpose

Table 1: Adapted from - O'Keefe, 2023; Lin, 2024; and Voruganti, 2024)

Content Generation in Public AI Systems versus Enterprise/Private Systems

People who use enterprise-specific Gen AI tools use it quite differently in terms of focus areas of content when compared to the everyday usage of public Gen AI tools. For example, recent analyses in 2024 by McKinsey, Deloitte and EY show that most organisations who have adopted Gen AI tool as an in-house technology or planning to do so in the near future use it for things like: conducting material research and development, creating work-specific tasks, presentations, emails, and work-related activities, analysing customer feedback and sentiments, optimising current products based on user feedback, developing new product ideas and designs, creating or automating specific code tasks, detecting errors or mitigating financial risks, etc. (Khan & Chen 2024; Booth et al., 2024; Satish et al., 2024). Users however use public AI models for everyday tasks that they would typically "Google"; content creations for articles, social media, and birthday messages; personal productivity like scheduling appointments or brainstorming ideas; writing jokes, generating games ideas, and many more. Essentially, given the kind of data each system interacts with daily, this could vary the LLMs outputs (public Gen AI and enterprise Gen AI outputs) significantly.

Perception

People's interactions and assessments of generative AI systems are greatly influenced by their perceptions. It includes users' subjective perceptions and assessments shaped by their backgrounds, prejudices, and social influences. Perception in the context of generative AI is affected by elements including the fairness of outputs, the transparency of AI processes, and the degree to which results match user expectations (Baek et al., 2024). Studies have shown that a mix of human input, training data, and system design errors frequently contribute to the sense of bias in AI. Gaining user trust, addressing ethical issues, and promoting inclusion in AI applications all depend on an understanding of these perspectives (Brauner et al., 2023; Jones-Jang & Park, 2023). To close the gap between technical performance and societal expectations, this study combines different viewpoints to investigate how the public and workplace view Gen AI biases.

Relevant Theories

Algorithmic Bias Theory: This theory looks at how biased training data, poor design decisions, or unexpected outcomes of algorithmic processing can all introduce biases into AI algorithms. It highlights how crucial it is to identify and lessen these biases in order to guarantee fair and equal AI systems (Obermeyer et al., 2019).

Perception Bias Theory: When people unconsciously make judgments based on their expectations, it's known as perception bias, and it can affect how they interact with AI systems (Robins & John, 1997).

Human-Centred Design Theory: This theory focuses on developing AI systems with human values, needs, and biases at the forefront of the design process. It seeks to develop more inclusive and approachable AI interfaces by including a range of user viewpoints (Giacomin, 2014).

Fairness in Machine Learning: This theory discusses the moral and societal ramifications of artificial intelligence, highlighting the necessity of responsibility, transparency, and equity in machine learning algorithms. It looks at ways to identify and lessen biases in order to stop biased results (Pessach & Shmueli, 2022).

Previous Research

In the last few years, there have been several studies examining the advantages and drawbacks of Gen AI, as a result of its increasing use in personal and workplace contexts. Zhou and Lee (2024) and Brynjolfsson and Raymond (2023) offer a thorough examination of how AI increases productivity in workplace usage, while posing moral questions, especially in relation to decision-making fairness. This line of thought is expanded upon by Klingbeil et al. (2024), who examined the unforeseen repercussions of AI use in business settings, including over-reliance and its effect on critical thinking.

In the context of this research, more specific studies have also been done in this area – with a focus on how bias appears in generative AI outputs, highlighting the significance of algorithmic design and training data (Sadeghiani, 2024; Allan et al., 2024). Sadeghiani (2024) examined current developments in addressing gender-specific biases, while Nadeem et al. (2020) and Nicoletti and Bass (2023) focused their attention on systemic patterns where generative AI perpetuates gender stereotypes. Similarly, racial bias is discussed by Nicoletti and Bass (2023), Zhou (2024), and Sadeghiani (2024), who highlighted the need for more inclusive training datasets and discrepancies in AI-generated content. Several other studies such as Tambe et al., 2019; Buolamwini, 2019; West et al., 2019; Uhunoma, 2022, have also shown algorithms and AI in general are inherently biased and outputs may discriminate against women, people of colour and those with disability.

In addition, there have also been a few studies focusing on users' perceptions of generative AI outputs. The perception of generative AI in higher education was examined by Arowosegbe et al. (2024), who focused on ethical and inclusivity issues. Secondly, in the study by Baek et al. (2024) on college students' usage and attitudes towards Gen AI, they observe both positive sentiments and doubt regarding the future of generative AI. Lastly, Choi et al. (2023), also did a comparative study of different demographic groups' perceptions of Gen AI, where they highlighted the necessity of transparent AI procedures while observing significant variations in user views across the different demographic groups.

Methodology

Research Design

A mixed-methods approach was employed in this study to explore and understand perceptions of bias existence in outputs generated by Gen AI (in the form of textual contents, images and sounds, original material, etc.), from public sentiment view and workplace perspectives. Consequently, the study combined qualitative data collected through text mining using specific keywords from Tweets (on X platform, formerly called Twitter), and quantitative data collected through a structured questionnaire from employees who use an enterprise-based/custom-built or generic Gen AI tool for work purposes.

Quantitative Dataset

To find out how employees using enterprise/custom-built or even generic Gen AI tools feel about generative AI outputs in terms of bias perception, a structured questionnaire was deployed to understand their perspective. A 5-point Likert-scale question assessing bias perceptions in three categories—outputs being biased towards women, people of colour, and neurodiverse people—were included in the survey. To find possible predictors of bias perception, additional demographic information was also collected, including gender, ethnicity, neurodiverse or not, age, length of using Gen AI tools, etc. 'Scenario-based questions on a Gen AI output' and 'suggestions for improvement of Gen AI where bias is perceived' questions were also asked.

The target population included any working-class UK residents who have used a Gen AI tool for work purposes (either public Gen AI tools like Chat GPT, Gemini, Claude, etc. or enterprise Gen AI tools, built specifically for their organisations). Specifically, we were interested in working-class people who used these tools more in the context of tasks related to communications, people-related decision-making, or creative content generation and ideas to reflect human values and perspective. Consequently, our population involves people who work in Marketing or Communication teams, Human Resources, Public Relations, or Customer service or Sales fields in the UK. Convenience sampling was used. This technique was chosen because it saves time and money, even though it has been considered relatively biased (Acharya, et al., 2013; Etikan & Bala, 2017). Although a larger sample size would have been ideal for this study, a total of 141 responses were collected. This was deemed adequate, as it was a mixed-method study, and other datasets were collected through the qualitative route, explained in the next section.

The structured questionnaire was administered through an online survey, through Google form. Participants were recruited through LinkedIn, and personal network, and data collection occurred between September and December 2024.To make sure the survey items were reliable and clear, a pilot test with 12 initial participants was carried out. The final questionnaire was improved by taking into account feedback.

Qualitative Dataset

To find out how the public (not restricted to employees) feels about generative AI outputs in terms of bias perception, Tweets related to Generative AI bias were gathered using NodeXL platform. NodeXL is a tool that allows users to gather data in the form of Tweets using specific keywords. It also allows users to carry out sentiment analysis to gauge public opinion on specific topics or keywords (NodeXL, 2024). Keywords such as "Generative AI bias OR Generative AI fairness OR Generative AI Stereotypes OR Inclusive Generative AI OR Generative AI discrimination AND Keywords related to Chat GPT, Gemini, Claude, Women, People of colour, neurodiverse people, etc.," were used as search criteria to gather relevant tweets between when Chat GPT was publicly introduced on November 30th, 2022, till December 2024.

A total of 1,926 tweets fitting the relevant keywords on Gen AI perceptions were extracted. Irrelevant tweets were removed, after which Sentiment and Thematic Analyses were performed on the dataset (this is discussed in the next section).

Ethical Considerations

The suggested standards of Macfarlane (2010) were implemented in order to guarantee that this study is conducted in accordance with ethical principles. This included not only notifying questionnaire respondents of their right to anonymity and the ability to withdraw at any moment but also obtaining verbal or written consent. Additionally, Saunders et al. (2019) advise that respondents' answers be handled privately and confidentially. Therefore, neither the questionnaire nor the qualitative data from Twitter contained any individually identifying information that was taken or examined.

Data Analysis

Quantitative Analysis

SPSS V.29 was used to carry out quantitative analyses including reliability tests, descriptive statistics and analysis of variance (ANOVA).

The internal consistency of the questionnaire items was evaluated through a *reliability test*. To assess how well the items measured the same underlying construct, Cronbach's alpha coefficient was computed (Nunnally & Bernstein, 1994). This phase was crucial to ensuring the reliability of the questionnaire and the validity of the subsequent data analysis. The Cronbach's alpha coefficient was 0.92, suggesting that the items in the questionnaire were internally consistent. See the table below.

Table 2: Reliability Statistics

Cronbach's Alpha	N of Items
.924	17

The output from the questionnaire responses was analysed using *descriptive statistics* to give an overview of the survey results, with particular attention paid to frequency distributions, variability and central tendencies. This aids in summarising the participants' demographic traits and general patterns in how they answered survey questions.

To find out if there are statistically significant variations between the means of the different demographic groups, an *Analysis of Variance (ANOVA)* was utilised (Larson, 2008). Based on demographic factors like gender, race, neurodiverse status, etc., this study evaluated variations in bias perception. This makes it easier to spot trends or differences in how various employee groupings view generative AI outcomes.

Qualitative Analysis

Text data from Tweets on X platform were categorised as negative, or positive using *sentiment analysis*. This investigation gauges public opinion regarding generative AI bias using NodeXL natural language processing tool. Combining these opinions gives us a better understanding of how the general public views Gen AI outputs, which we can then use to compare to that of employees who use it for work purposes to see the similarities or differences in perceptions.

Finding recurrent themes or patterns in qualitative data, like Twitter comments, is known as *thematic analysis* (Braun & Clarke, 2012). This approach highlights complex viewpoints and contextual insights by methodically coding and classifying texts to identify important themes and attitudes regarding generative AI bias from a public opinion standpoint.

Results

In this section, the results from the quantitative and qualitative analyses are presented starting with the former.

1. Workplace: Bias Perception of Gen AI Outputs against Women, People of Colour and Neurodiverse People

The results from the quantitative data collected using the structured questionnaire are presented below, starting with the demographic distribution of respondents, results of the bias perception for women, people of colour and neurodiverse people, and how perceptions differ within each demographic group.

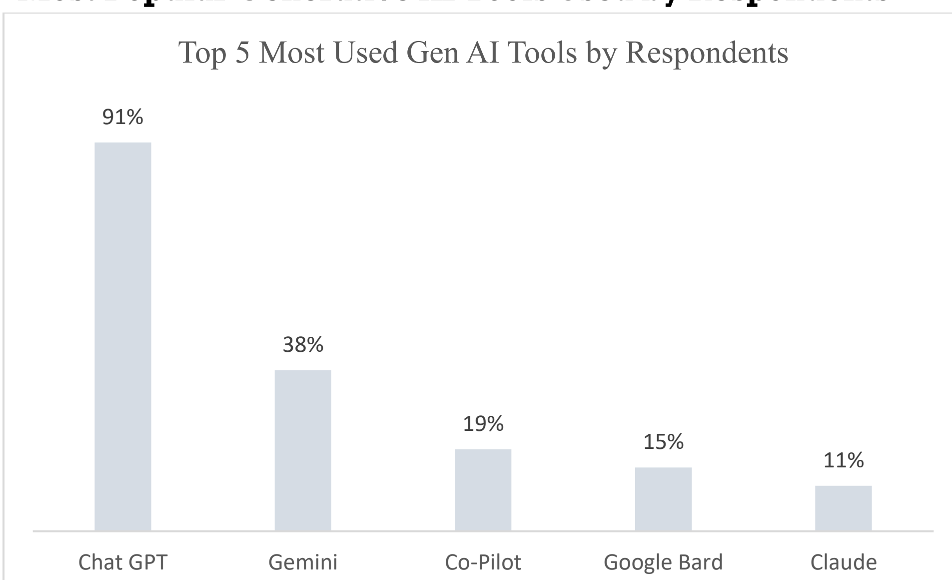
Demographic Distribution of Respondents

Table 3: Demographic Distribution of Respondents

Gender of Respondents	**Frequency**	**Percentage**
Male	72	51.1%
Female	69	48.9%
Others / Prefer not to say	0	0%
Ethnicity of Respondents	**Frequency**	**Percentage**
Black	126	89.4%
White	5	3.5%
Asian	7	5.0%
Hispanic / Latino	3	2.1%
Is Respondents Neurodiverse?	**Frequency**	**Percentage**
Yes	19	13.5%
No	122	86.5%
Age Group of Respondents	**Frequency**	**Percentage**
16 to 25 Years	29	20.6%
26 to 35 Years	91	64.5%
36 Years and Above	21	14.9%
Length of Time Respondent Has Used Gen AI	**Frequency**	**Percentage**
Less than 1 Month	8	5.7%
1 Month to 6 Months	23	16.3%
More than 6 Months	110	78.0%

A total of one hundred and forty-one people completed the questionnaire. The respondent's demographics collected vary across their gender, ethnicity, if they were neurodiverse, age group and how long they have been using at least one Gen AI tool. The respondents' distribution by gender was almost the same, with 51% identifying as Male and the remaining 49% identifying as Female. None of the respondents identified as any other gender or preferred not to say. In terms of Ethnicity, most of the respondents (around 89%) were of Black - British, Caribbean and African heritage. The remaining 10% of respondents were distributed across Asian (5%), White (3.5%) and Hispanic/ Latino (2.1%). In terms of whether the respondents identify as Neurodiverse, most (86.5%) did not identify as one, and the remaining 13.5% identify as Neurodiverse people. About two-thirds of the respondents were between the ages of 26 and 35 years old, 20% were between 16 and 25 years old and the remaining 15% were 26 years or above. Lastly, most of the respondents (78%) have used Generative AI tools for more than 6 months, 16% of them have used it for 1 to 6 months, while only 6% of them have used it for less than a month.

Figure 1: Most Popular Generative AI Tools Used by Respondents



In addition, data from the respondents showed that Chat GPT is the most used Gen AI tool for their work, with 91% of respondents using it for at least one work purpose or the other. The top 5 Generative AI Tools they use for work purposes are Chat GPT, Gemini, Co-Pilot, Google Bard, and Claude in that order. Some respondents mentioned other Gen AI tools, some of which are enterprise Gen AI tools, custom-built for their organisations.

Perception of Bias against Women, People of Colour and Neurodiverse People

In responding to the bias perception questions in the survey, participants chose their responses following the 5-point Linkert scale format which was adopted in the questionnaire. 5 represented Strongly Agree and 1 represented Strongly Disagree, with 3 being Neutral. The mean and the standard deviation of the result for the three groups; Women, People of Colour and Neurodiverse People can be seen below.

Table 4: Mean and Standard Deviation of Respondents Perception Score

Items	Mean Score	Standard Deviation
Bias perception of Gen AI outputs towards Women	2.09	1.01
Bias perception of Gen AI outputs towards People of Colour	2.29	1.22
Bias perception of Gen AI outputs towards Neurodiverse People	2.15	1.04

Following the above table, the Weighted Average Mean is calculated by summing the mean of the individual items and dividing it by the number of items.

That is *Weighted Average Mean* = Sum of Individual Item Means / Number of Items (i.e. 6.53/3 = *2.18)*

There are a couple of things to note from this. First, the weighted average mean of 2.18, which is closer to the "Disagree" category on the Likert scale (where 2 = "Disagree" and 3 = "Neutral"), suggests that, from the perspective of those who use Gen AI in the workplace, there is generally low perception of bias in the outputs they see on a day-to-day basis from Gen AI, toward women, people of colour, and neurodiverse individuals. This perhaps suggests that overall perceptions of bias in Gen AI output from Perspective of people in the workplace are leaning toward disagreement rather than agreement. Secondly, the individual mean score for all three groups (Women: 2.09, People of Colour, 2.29 and Neurodiverse People, 2.15), similarly suggest the same i.e. perceptions of bias in Gen AI outputs from workplace perspectives, are leaning toward disagreement for each item, rather than agreement. Nevertheless, there are however differences in the levels of perception in the three groups (women, people of colour and neurodiverse people) as seen in the table below, comparing the individual mean score to the weighted average mean score. That is,

- Where Individual Mean > Weighted Average Mean (2.18): High Perception
- Where Individual Mean ≤ Weighted Average Mean (2.18): Low Perception

Table 5: Decision Table of Respondents' Perception Score

Items	Mean Score	Standard Deviation	Decision
Bias perception of Gen AI outputs towards Women	2.09	1.01	Low Perception of Bias
Bias perception of Gen AI outputs towards People of Colour	2.29	1.22	High Perception of Bias
Bias perception of Gen AI outputs towards Neurodiverse People	2.15	1.04	Low Perception of Bias

Following the above table, with a mean score of 2.29, item 2 indicates a substantially greater perception of bias than the weighted average (2.18) and had a wider range of perception as suggested by the standard deviation (1.22). Practically speaking, the mean score (2.29) is still in the "low" range, but it may suggest that respondents believe bias is a greater problem in this area (bias perception on People of Colour) than in the other two areas. For the first and third items, compared to the overall average, these items' mean scores of 2.09 and 2.15, respectively, indicate comparatively lower perceptions of bias, which could indicate that respondents see even less bias in these two categories than they do for People of Colour category.

Variations of Opinions and Perceptions within Each Demographic Group

Perception of generative AI outputs biased on Women by Different Gender.

Relatively, men perceive generative AI outputs to be less biased than women do, as evidenced by the fact that the mean perception score for females is 2.2, higher than the mean score for males, which is 2.0. Perceptions within each group appear to be rather constant, with slightly greater variability among females, according to the standard deviation (0.9 for males and 1.0 for females). The relatively low F-value of 1.48 suggests that the difference in perception scores between males and females is not larger than the difference within each group. This implies that there might not be significant perceptional differences between them. The observed variations between their views are also not significant, according to a p-value of 0.27. Therefore, there is little data to imply that gender affects people's perceptions of whether generative AI outputs are biased towards women.

Perception of generative AI outputs biased on People of Colour by Different Race/Ethnicity.

For both Black and white people, the mean perception score was 2.3. For Asians, it was 2.4, while for Hispanics or Latinos, it was 1.5. This suggests that compared to other groups, Hispanic individuals perceive significantly less bias in generative AI results. The F-value was also low, suggesting that the variance in the groups' perception scores is not greater than the variance within each group; finally, the p-value of 0.75 further suggests that the observed differences between their perceptions are not significant. The standard deviations among all groups, however, are similar, suggesting relatively consistent perceptions within the groups. The lack of variation between the groups may be the result of most of the respondents being of the same ethnicity - Black (89%).

Perception of generative AI outputs Biased on Neurodiverse People by Neurodiverse and Non-Neurodiverse People.

The mean perception score for Neurodiverse people was 2.4 as opposed to that of people who are not Neurodiverse (2.1). This indicates that Neurodiverse people relatively perceive slightly higher bias in generative AI outputs for them, than people who are not Neurodiverse. However, the standard deviations between both groups were similar, suggesting relatively consistent perceptions within the groups; the F-value was also quite low, indicating that the variance of the group's perception scores is not greater than the variance within each group; and lastly, the p-value of 0.25 also indicate that the observed differences between their perceptions are not significant. The lack of significant difference may be due to the fact that most of the respondents were not neurodiverse (86%).

Perception of generative AI outputs bias based on the Age Group of Respondents

The major callout here was the mean perception score of different age groups as it relates to their perception of bias against People of Colour. The mean score for the age group 36 years and above was 2.7, compared to those 16 to 25 years (2.1) and 26 to 35 years people (2.2). This indicates that older people perceive slightly higher bias in generative AI outputs for People of Colour than younger people. However, a p-value of 0.1 indicates that the observed differences between their perceptions are not significant at 0.01 and 0.05 but significant at 0.1. The lack of significant difference at 0.01 confidence level may be due to the number of respondents (141). A larger number of respondents may yield a significant result. There were no significant differences in their perception towards the perspective of Women and Neurodiverse people.

Perception of generative AI outputs bias based on How Long Respondent Has Used Gen AI Tools

People who have been using Gen AI tools for less than 1 month, relatively perceived higher bias for People of Colour (2.8) with a 1.2 standard deviation, and for Neurodiverse people (2.9) with a 1.3 standard deviation. People who have used Gen AI for a longer period (more than 6 months) were more cautious and indicated that they have perceived less bias when compared to others in all three groups: 1.9 for

Bias on women; 2.3 for Bias on People of Colour; and 2.1 for Bias on Neurodiverse people.

2. Public Opinion: Bias Perception of Gen AI Outputs against Women, People of Colour and Neurodiverse People

The results from the qualitative data collected through text mining of tweets related to Generative AI keywords are presented below, starting with the results of the sentiment analysis (created using the NodeXL platform) and then the thematic analysis of the tweets.

Twitter Users' Sentiment on Generative AI Outputs: More Positive or Negative Sentiments?

Table 6: Sentiment Analysis of Tweets (via NodeXL)

Description	Count
Total number of tweets extracted	1,927 Tweets
Total number of words in the tweets	41,497 Words
Total number of words categorised as Positive Sentiments	1,210 Words
Total number of words categorised as Negative Sentiments	2,454 Words

Each piece of social media content is usually categorised into distinct emotional categories, such as positive, negative, or neutral, as the result of text analysis for semantic orientation (Zhang et al. 2012). The category type employed in this study is Positive vs. Negative Sentiment, as recommended by Bartov et al. (2018), where they describe this category type as consistent with the nature of "bullish" and "bearish," where positive sentiment denotes an optimistic outlook on the market and negative sentiment denotes a pessimistic outlook.

A total of 1, 927 tweets relating to Generative AI and Bias or Fairness keywords were extracted using NodeXL. These tweets had over 40,000 words and following the recommendation by Bartov et al (2018), the total number of words in the tweets were categorised into emotional sentiments using NodeXL. The result categorised the tweets into two: the tweets which had a more optimistic outlook about generative AI usage, output or general sentiments and those which had a more pessimistic or cautious view about generative AI. Irrelevant words (without a positive or negative connotation) in these tweets were not categorised, the full tweets were however used for thematic analysis, and the result of this is presented in the next section.

3,664 words in these tweets had a directional connotation or sentiments. About one-third of these categorised words had more positive sentiments towards generative AI, whereas the other two-thirds words geared towards more negative sentiments and standpoints about generative AI. Overall, the data above shows that the sentiments of users on X (formerly Twitter), lean more towards negative perception of generative AI usage and output than positive.
Some examples of each of these tweets can be seen below.

Examples of tweets with words leaning towards or showing positive sentiments:

> *"…ChatGPT was also found to have zero gender bias when diagnosing mental health issues."*
>
> *"Only 10% of educators surveyed used generative AI in their classrooms last year. From rostering to dyslexia screening, AI helps teachers spend more time with students…"*
>
> *"Chat GPT is fine if you know how to use it. Wasn't a political question at all so no bias. Just a technical one which it's good with…*
>
> *"…Recent studies found that ChatGPT prescribed therapy accurately 97% of the time when given scenarios about mental health issues."*

Examples of some tweets with words leaning towards or showing negative sentiments:

> *"Chat GPT refuses to make me a woman blacksmith, If I wanted poorly written stories with excessive liberal bias, I'd just use Chat GPT…"*
>
> *"Over months of reporting @ redacted and I looked at thousands of images from @redacted and found that text-to-image AI takes gender and racial stereotypes to extremes worse than in the real world.*
>
> *" …I discovered that nearly all image generative AI models are (or were at the time of our study) very sexist and gender biased. What a shock! #AI models have bias against #women and #girls"*
>
> *"…#AI undervalues women concerning their capabilities, occupations and characteristics.*
>
> *"…last week I raised concerns here on the diversity of the community especially regarding #AI making discrimination and bias of disabled and neurodiverse people worse. The diversity of voices matters"*

The table below shows the most frequently used hashtags, words and word pairs that users on X, used when making tweets related to Generative AI.

Table 7: Top 10 Hashtags, Words and Word Pairs that Appeared the Most in the 1,927 Tweets

#	Top 10 Hashtags Used in the Tweets	Top Words/Pairs that Appeared the Most in Tweets
1	#ai	Bias
2	#bias	Chat
3	#chatgpt	GPT
4	#ethicalai	AI
5	#openai	Chat, GPT
6	#inclusion	Generative, AI
7	#generativeai	GPT, Bias
8	#ml	Bias, Chat
9	#futurism	Bias, Discrimination
10	#intelligencefactory	Gen, AI

Recurring Themes among Twitter Users' Perceptions
Further analysis of the texts mined from X was carried out using thematic analysis. This entailed grouping the gathered textual data into relevant themes before presenting them in order to address the research objectives. Below are the five major themes gathered from the qualitative dataset.

The overarching perception is that of scepticism: A lot of users on X (formerly Twitter), had doubts about the inclusivity and dependability of generative AI results. They highlighted worries about biases, inconsistencies, and the ethical implications of AI technologies while still appreciating the technology's potential. An example of a tweet here is: *"It's crucial for tech companies like Google to address issues like this promptly. Balancing AI tools to prevent bias and discrimination is a significant challenge that requires constant attention. It's essential to ensure that these tools are fair and inclusive for everyone."*

More users perceive bias in outputs they receive from Gen AI relating to women and people of colour, but less for Neurodiverse people: The issues around perceived bias were more around race and gender. Many users believe that generative AI output perpetuates gender and racial stereotypes in texts and imagery outputs and are prejudiced, against women and people of colour. Many pointed to instances of stereotyped when generating photos of janitors (generating more Indian or Latino faces) or Leaders (generating more male-centred outcomes and emphasising the need for more inclusive training datasets). One respondent said: *"…Today I asked for 3 compliments for a woman, then I asked for a man. Its replies are different from each other. For example, it uses beauty for women, leadership for man. Then I asked it again, if women and man are equal why their compliments are different? And it corrects itself. I don't know if it learns from our warnings or not"*

However, some users were more neutral or positive with respect to outcomes for Neurodiverse people. Neurodiverse users feel that Gen AI tools such as ChatGPT and Microsoft Co-pilot are useful for helping them summarise and articulate their ideas, indicating a heavy reliance on technology for efficient communication. One user who uses Gen AI for work purposes said: *"I'm neuro-diverse, so sometimes I struggle with accurately describing or writing things. I can put prompts in Chat GPT, and it words them well. In emails or teams' messages, Microsoft copilot can summarise key points for me or give suggested responses. It is brilliant for me."*

Inherent bias from training data: Some of the users highlighted the biases present in training datasets, indicating that AI results are strongly impacted by the variation and diversity of data which is consistent with the algorithm bias theory. To lessen these problems, they emphasised the significance of incorporating narratives and experiences from a range of populations. One X user said: *"Be aware! AI systems can learn biases from training data, leading to unfair results. Regular evaluation, diverse datasets, and inclusive teams can help minimise biases and ensure safer and more ethical AI applications."* Another said: *"Training data bias can cause unfair outcomes! To use AI safely, ensure diverse datasets, ongoing monitoring, and human oversight. Let's build inclusive and ethical AI models for a better future! #AIethics #AIsafety"*

Role of users (using the right prompts): Some users emphasised people's input will usually have a great effect on the outputs or responses that they get from generative AI. They maintained that clear, objective cues could enhance outcomes by lowering the possibility of skewed, inconsistent or biased answers. One user said: *Outputs are defined by inputs; the output is dependent on the input - a generic input will produce a generic output that is not orientated towards minorities or people who want to be considered as "special".* Another used this prompt to get a more inclusive image: *"I asked AI to create a picture(s) for #IWD2024 using prompt: Imagine a gender equal world. A world free of bias and discrimination. A world that's diverse, equitable, and inclusive, where difference is valued and celebrated. Collectively we can all #InspireInclusion".*

Discussions

Different Perceptions from Workplace and Public Perspectives

While workplace respondents reported modest levels of perceived bias across all demographics, public sentiment analysis showed significant mistrust and negative perceptions of bias in generative AI outputs for women and persons of colour. This disparity is in line with Allan et al. (2024) and Sadeghiani (2024), who noted that the public was worried about generative AI reinforcing prejudices because of the constraints of training data. The Algorithmic Bias Theory, which highlights the importance of training data, can be used to explain the disparity; workplace tools might use more inclusive and structured datasets than publicly accessible AI systems. Additionally, as Baek et al. (2024) point out, public usage frequently occurs in settings where biases are more obvious, such as political discourse or text-to-image generation.

Relatively Lower Bias Perceptions from the Workplace for Women and Neurodiverse People

Perceptions of bias in the workplace, while overall low, were relatively greater for people of colour and lower for gender and neurodiverse individuals. The latter part of this statement is in contrast to Nicoletti and Bass (2023), who discovered significant biases associated with gender in the outputs of generative AI. The latter part of this is also in contrast with the findings of the public perspectives, where most data suggested significant bias for women and gender stereotypes. A possible reason for this may be explained by the theory of Fairness in Machine Learning. That is, given that in a workplace context, a high value on inclusion and fairness is placed in their policies, and data structuring, this may explain why workplace Gen AI systems produce less perceived biases. The first part of this workplace findings on greater perceptions of bias for people of colour is similar to the findings of Zhou and Lee (2024), who noted that the increased perception of racial bias points to continued difficulties in attaining racial parity.

Positive Sentiment for Neurodiverse People

The results from the public perspectives show more positive sentiments for Gen AI outputs, as it relates to Neurodiverse people, as the users view it positively as a tool for helping dyslexic users communicate better. This is consistent with the findings of Choi et al. (2023), who discovered that users from a variety of backgrounds appreciate AI's usefulness when it meets certain demands. This conclusion is supported by the Human-Centered Design Theory, which promotes AI systems that are customised to meet user needs and exhibit inclusivity through usefulness.

Impact of Prompts and Usage Duration

Workplace respondents who had less than a month of exposure to generative AI had more unfavourable opinions and negative sentiments, which is consistent with research by Baek et al. (2024), which found that perception is influenced by user familiarity. Public users who have positive sentiments about Gen AI output also opined that knowing the right prompt to use impacts the output users receive, and other users who may perceive bias are not very familiar or skilful in using the right prompts to get the best and unbiased outputs. This phenomenon is also explained by the Perception Bias Theory, which contends that when users gain more proficiency with AI systems, their initial mistrust fades.

Inherent Bias in Gen AI Models and Demand for More Diverse Training Data

In order to lessen biases in generative AI outputs, users and respondents from the public and workplace argued for the inclusion of more diverse datasets. This backs up Sadeghiani's (2024) findings, which highlighted the value of dataset diversity. The fairness and usefulness of AI systems are directly impacted by inclusive training data, demonstrating the Garbage-In, Garbage-Out Principle. Cleaner, more organised datasets may be advantageous for workplace tools, which would account for their reduced perceptions of bias.

Conclusion

This research paper highlights how different user demographics, as well as application settings (workplace or public) interact to create complex perceptions of generative AI bias in both public and workplace settings. In contrast to decreasing perceptions of bias in the workplace, public attitude shows more scepticism and perceived bias toward women and persons of colour. The Fairness in Machine Learning Theory promotes inclusive training datasets and equitable algorithmic outputs, highlighting the necessity of openness and equality in AI design. Future research should explore additional datasets and application contexts to provide a more comprehensive understanding of Gen AI biases.

Limitation

This research acknowledges that its demographic representation and dataset size are limited. First, the lack of racial diversity among survey participants may have affected workplace findings. Secondly, population representation might be improved by using more samples and a structured cluster sampling. In addition, Twitter was the only source of publicly available data; blogs and Reddit, which can provide more nuanced sentiment fluctuations, were not included. Future research will be more generalizable and provide insights into how people perceive generative AI if data sources, and participant demographics are expanded.

Data Avaialability
The dataset generated from the survey and the text-mined public data used in this research are available in Fig share repository.

Funding Declaration
This research received no external funding.

Acknowledgement
This article was prepared within the framework of the Basic Research Program of the HSE University. Maryann Asemota acknowledges support of the Basic Research Program of the HSE University.

References:

1. *Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it. Indian journal of medical specialties, 4(2), 330-333.*
2. *Allan, K., Azcona, J., Sripada, Y., Leontidis, G., Sutherland, C. A., Phillips, L., & Martin, D. Stereotypical Bias Amplification, and Reversal, in an Experimental Model of Human Interaction with Generative AI.*
3. *Arowosegbe, A., Alqahtani, J.S. and Oyelade, T., 2024, October. Perception of generative AI use in UK higher education. In Frontiers in Education (Vol. 9, p. 1463208). Frontiers Media SA.*
4. *Baek, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. Computers and Education: Artificial Intelligence, 7, 100294.*
5. *Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. SSRN Electronic Journal.*
6. *Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 1-41.*
7. *Bartov E, Faurel L, Mohanram PS (2018) Can Twitter help predict firm-level earnings and stock returns? TheAccounting Review, 93(3):25–57*
8. *Booth, B. Donohew, J. Wlezien, C.& Wu, W. (2024). Generative AI fuels creative physical product design but is no magic wand. Mckinsey. Retrieved October 5, 2024, from www.mckinsey.com*

9. *Braun, V., & Clarke, V. (2012). Thematic analysis. American Psychological Association.*
10. *Brauner, P., Hick, A., Philipsen, R., & Ziefle, M. (2023). What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. Frontiers in Computer Science, 5.*
11. *Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.*
12. *Buolamwini, J. (2019). Response: Racial and Gender bias in Amazon Rekognition—Commercial AI System for Analyzing Faces. Medium. Retrieved October 5, 2024 from medium.com*
13. *Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20(1).*
14. *Choi, S., Jang, Y., & Kim, H. (2023). Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. International Journal of Human–Computer Interaction, 39(4), 910-922.*
15. *Cook, S., Hagiu, A., & Wright, J. (2024). Turn generative AI from an existential threat into a competitive advantage. Harvard Business Review, 102(1), 118-125.*
16. *Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard business review, 96(1), 108-116.*
17. *Epstein, Z., & Hertzmann, A. (2023). Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld,*
18. *Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. 2023. Art and the science of generative AI. Science, 380(6650), 1110-1111*
19. *Etikan, I., & Bala, K. (2017). Sampling and sampling methods. Biometrics & Biostatistics International Journal, 5(6), 00149.*
20. *Ferrara, E. (2023). Fairness and Bias in Artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci, 6(1), 3.*
21. *Fu, R., Huang, Y., & Singh, P. V. (2020). Artificial Intelligence and Algorithmic Bias: Source, Detection, Mitigation, and Implications. In Pushing the Boundaries: Frontiers in Impactful OR/OM Research (pp. 39–63). INFORMS.*
22. *Giacomin, J. (2014). What is human centred design? The Design Journal, 17(4), 606–623.*
23. *Google (2024). Google cloud skills boots. Retrieved October 6, 2024, from www.cloudskillsboost.google*
24. *Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. Journal of Computer-Mediated Communication, 28(1).*
25. *Khan, K. & Chen, D. (2024). How AI in customer service can turn pain points into wow moments. Retrieved October 5, 2024,from www.ey.com*

26. *Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent andcosts of overreliance on AI. Computers in Human Behavior, 160, 108352.*
27. *Lalchand Satish; Srinivas, Val; Maggiore, Brendan; Henderson, Joshua (2024). Generative AI is expected to magnify the risk of deepfakes and other fraud in banking. Deloitte. Retrieved October 5, 2024, from www2.deloitte.com*
28. *Larson, M. G. (2008). Analysis of variance. Circulation, 117(1), 115-121.*
29. *Lishchynska, D. (2024). Top Custom Generative AI Development Companies in 2024. Retrieved October 15, 2024, from botscrew.com*
30. *Macfarlane, B. (2010). Researching with Integrity: The Ethics of Academic Enquiry.*
31. *Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. In FinTech and artificial intelligence for sustainable development: The role of smart technologies in achieving development goals (pp. 387-409). Cham: Springer Nature Switzerland.*
32. *Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Gender bias in AI-based decision-making systems: a systematic literature review. Australasian Journal of Information Systems, 26.*
33. *Nah, F. F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304.*
34. *Nicoletti, L & Bass, D. (2023). Humans are biased. Generative AI is even worse. Bloomberg. Retrieved October 11, 2024, from www.bloomberg.com*
35. *NodeXL (2024). GTransform your Marketing Strategy with NodeXL Pro + Insights. NodeXL. Retrieved December 14, 2024, from nodexl.com*
36. *Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory. McGraw-Hill.*
37. *Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.*
38. *Ok, Emmanuel & Blessing, Moses. (2023). TOPIC: Building Custom Applications: A Step-by-Step Guide to Integrating Generative AI in Your Workflow AUTHOR: Moses Blessing.*
39. *Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. ACM Computing Surveys (CSUR), 55(3), 1-44.*
40. *Peter MacDonald Hall (2024). Impact of generative AI on DEI. Fairerconsulting. Retrieved October 9, 2024,from www.fairerconsulting.com*
41. *Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In Handbook of personality psychology (pp. 649-679). Academic Press.*

42. *Sadeghiani, A. (2024). Generative AI Carries Non-Democratic Biases and Stereotypes: Representation of Women, Black Individuals, Age Groups, and People with Disability in AI-Generated Images across Occupations. arXiv preprint arXiv:2409.13869.*
43. *Saunders, M., Lewis, P., & Thornhill, A. (2019). Research methods for business students, 8th ed. In Pearson eBooks. dspace.uniten.edu.my*
44. *Smith & Cankert (2024). EU Act to enter into force. Deloitte. Retrieved October 11, 2024 from www.deloitte.com*
45. *Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. California Management Review, 61(4), 15-42.*
46. *Uhunoma, E (20222). Facial Recognition Technology and Racial Discrimination a Study on the Design of On-Demand-Asynchronous Video Interview Solutions Used in Recruitment. International Journal of MultidisciplinaryResearch, 6(7) 3240-3258.*
47. *West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating systems. AI Now. Retrieved October 5, 2024 from cdn.vox-cdn.com*
48. *Xavier, B. (2024). Biases within AI: challenging the illusion of neutrality. AI & SOCIETY, 1-2.*
49. *Zhang, Z., Li, X., & Chen, Y. (2012). Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews. ACM Transactions on Management Information Systems (TMIS), 3(1), 1-23.*
50. *Zhigalko B. (2024). The impact of generative AI on DEI. Faireconsulting. Retrieved October 9, 2024, from www.fairerconsulting.com*
51. *Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. PNAS Nexus, 3(3).*
52. *Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). Bias in generative ai. arXiv preprint arXiv:2403.02726.*