

Developing Racially Ambiguous Job Candidate Avatars with Strong versus Ambiguous Job Candidate Backgrounds in Preparation for a Field Experiment for Enhancing Inclusive Hiring and Reducing Implicit Bias

Alexa S. Trifilo
Cisco Systems Inc.

Gary Blau
Temple University

This study explores the utilization of generative Artificial Intelligence (AI) to create racially ambiguous avatars for job candidate interviews, aiming to enhance inclusive hiring practices and reduce implicit bias. With diversity recognized for boosting organizational performance, addressing hiring discrimination becomes crucial. This research investigates how avatars can anonymize candidates' racial identities, potentially transforming "blind" hiring to minimize subconscious biases. Through an iterative pilot process engaging 30 female experienced-in-hiring participants with information technology backgrounds, the avatars representing White, Asian, and race-neutral candidate females, and corresponding candidate job descriptions, were refined, setting the foundation for a follow-up field experiment. This experiment will examine avatars' effects on hiring recommendations and decisions by managers, manipulating strong versus ambiguous candidate qualifications, contributing insights into overcoming implicit biases. The initial pilot findings here suggest promising directions for technology's role in fostering equitable recruitment, providing a groundwork for future studies on technology-enhanced inclusive hiring strategies.

The General Research Question

Amid the backdrop of rapidly evolving AI technologies and their potential impact on the workforce, this study seeks to address a pivotal query: Can generative AI be used to create racially ambiguous avatars for future research, aiming to foster inclusive hiring practices and minimize implicit bias? Grounded in the notion that diverse workforces enrich organizations and broader society (Hoogendoorn, et al., 2013; Kaczmarek & Nyuur, 2022), this question not only delves into avenues for fostering inclusivity but also promises profound insights into human behavior—a topic of keen interest to scholars in psychology and sociology. Given the commercial novelty of generative AI and the limited rigorous studies thus far on avatar utilization in professional settings, this investigation holds promising significance for a broad audience, encompassing managers, educators, human resource practitioners, and other organizational stakeholders.

Literature Review

Workforce Diversity

Workforce diversity, defined as differences among people within a work group, is crucial for organizational success, enhancing performance through more heterogeneous teams (Hoogendoorn, et al, 2013; Roberson, 2019). However, achieving diversity is challenging due to hiring discrimination, which can occur at various stages of the hiring process, from resume screening to the interview and

final selection, often manifesting as implicit racial in-group bias (Dovidio & Gaertner, 2000). Despite widespread support for equal treatment across races, a significant racial discrepancy in hiring persists, largely attributed to implicit in-group bias, where individuals favor those perceived as belonging to their own group (Çoksan & Çingöz-Ulu, 2022; Quillian, 2006; Quillian & Midtboen, 2021; Tajfel, 1979). This bias can lead to decreased organizational diversity, undermining efforts to build inclusive workplaces (Young, 2022).

In Group Bias

Expanding on the topic of in-group and out-group formation, in light of Dovidio and Gaertner's (2000) insights, could significantly enhance the understanding of social categorization theory's implications in intergroup relations. The work of Çoksan and Çingöz-Ulu (2022) provides a compelling examination of how social categorization group norms influence the relationship between group identification and in-group bias, particularly among ethnic groups with a history of conflict. Their findings, rooted in social identity theory, suggest that in-group bias becomes more pronounced with stronger group identification when favoritism is the prevailing norm. This underscores the critical role of social categorization in shaping individual and collective behaviors, offering a valuable perspective on the mechanisms driving the formation of in-groups and out-groups.

Physical Identity and Generative AI Image Generators

Could removing physical identity from interviews, through "blind" hiring practices, reduce implicit racial in-group bias? Quillian & Midtbøen (2021) explored this in the context of resume screening, suggesting less bias at this initial stage. The proposal here extends to employing generative AI to create standardized avatars for interviews, potentially masking racial cues and minimizing biases. Generative Adversarial Networks (GANs), as introduced by Goodfellow et al. (2020), and enhanced through StyleGAN by Karras et al. (2019), represent significant advances in AI-driven image creation, offering a foundation for this approach. In this research, tools like Bing Image Creator (Microsoft, n.d.) and Artguru (ArtGuru, n.d.; AI4.Tools, n.d.) will be utilized. These platforms demonstrate the capability of generative AI to produce art based on textual prompts or user inputs, serving as models for generating avatars that obscure job candidates' racial identities.

Job Candidate Qualifications

Csonka-Peeren and Cozzarin (2021) presented a general conceptual model for studying the impact of ambiguity for entrepreneurship-related experiments. With regard to hiring, prior experimental research (Dovidio & Gaertner, 2000), using 194 white male or female psychology undergraduates across two separate academic time periods, i.e., 1988-1989, $n = 112$ and 1998-1999, $n = 84$, found that while self-reported expressions of overt prejudice for hiring (against blacks) declined over time, more subtle covert discrimination, i.e., implicit bias persisted. This bias was particularly manifest in ambiguous simulated hiring situations, i.e., where the match between a black candidate's qualifications and position criteria was unclear. With ambiguous qualifications for a position, lower hiring recommendations were made for black versus white job candidates. However, generalizability using samples of undergraduate students is low. This indicates the need in further investigation for creating an ambiguous situation to test the impact of avatars in mitigating hiring bias based on race.

Implicit Bias Assessment

Implicit bias is based on the concept that despite their overt denial, and even with organizational training to reduce bias, people may have automatic biases or preferences that operate outside their conscious awareness or control (Greenwald et al., 1998). Such bias can affect adversely hiring decisions (e.g., Derous & Ryan, 2018; Lira et al., 2017; Rooth, 2010), prejudice and discrimination (Quillian, 2006), and intergroup attitudes (Greenwald et al., 1998). For example, Rooth (2010) showed that Swedish recruiters were less likely to invite Arab/Muslim minority applicants to interview for positions when the Arab/Muslim IAT identified that these recruiters had more negative, implicit attitudes towards Arab/Muslims.

The Implicit Association Test (IAT) is a widely used psychological tool designed to measure implicit biases or unconscious associations between different concepts or

social groups (Greenwald et al., 1998). It has been applied in research studies examining implicit biases in hiring decisions (e.g., Derous & Ryan, 2018; Rooth, 2010), prejudice and discrimination (Quillian, 2006), and intergroup attitudes (Greenwald et al., 1998).

Specific Research Question Addressed

Based on this literature review, the more specific research question addressed in this study is: Can racially ambiguous job candidate avatars with strong versus ambiguous job candidate qualifications be developed in preparation for a field experiment for enhancing inclusive hiring and reducing implicit bias?

Method

Participants

The participants for this study included thirty females selected from a large technology company. These participants self-identified as White/Caucasian or Asian, currently reside in the United States, and have either completed interviewer training or are experienced interviewers. Given the nature of their employment, all participants were familiar with advanced communications technologies, including video and web conferencing. This ensured that all individuals involved had the necessary experience and skills pertinent to the study's requirements.

Participants in the study volunteered for a variety of reasons, reflecting a mix of personal interests, community involvement, and aspirations for professional development. A significant motivation was the study's AI image generation nature, attracting individuals keen on exploring new technology alongside the potential for personal growth. The desire for community connection and support was another major factor; participants were eager to use the study as a means to connect with others, especially focusing on supporting women's goals and ambitions. This demonstrated the importance they placed on forming meaningful connections and contributing to a shared cause. Altruism also played a crucial role, with several participants expressing a willingness to help and support others, showing a commitment to positively contributing to their communities.

Data Collection

The research methodology employed in this study was inherently iterative, designed to refine and enhance the survey instrument through continuous feedback loops. The methodology was meticulously structured around thirty participants, divided into three groups of ten. Each group sequentially completed the survey via the Qualtrics platform and then each subject participated in a thirty-minute in-depth interview session within 24 hours of survey completion. This cycle of survey and interviews spanned approximately one week for each group, followed by a week dedicated to analyzing the collected responses and feedback before progressing to interviews with the subsequent group. Data collection occurred from December 2023 into January 2024. Based on the insights gained from this analysis, modifications and updates were

be made to the survey, job description, avatar-based video vignettes, etc. to better capture the nuances of the participants' experiences and perspectives. This iterative cycle ensured that the survey evolved in response to the real-world applicability and accuracy of the questions, leading to more reliable and valid data collection across successive groups. University IRB approval was given on 11/8/23 (#31085).

Measures

The individual item variable measures used were: Gender (all female); Race (White or Asian); Completion of interviewer training (as hiring manager or individual contributor); Number of hiring interviews involved in; Years worked current company; Role within the organization (sales or other); Generational identification (Baby Boomer, Generation X, Generation Y, Generation Z); Self-assessed racial bias (where 1 = strongly disagree to 7 = strongly agree) ; IAT self-assessment (where 1 = Strong Automatic association of Euro Am with American and Asian Am with Foreign to 7 = Strong Automatic association of Asian Am with American and Euro Am with Foreign); Qualification manipulation of candidate (ambiguous versus strong); Perceived assessment of how well the candidate's skill set met the job requirements defined in the job description (where 1 = extremely ambiguous to 9 = extremely strong); Does the company currently use avatars in hiring (yes, no); and Participant comfort level with using avatars in hiring processes (where 1 = extremely uncomfortable to 9 = extremely comfortable).

Data Analysis

NVivo (2020) was employed to organize and scrutinize interview data, providing a structured platform for in-depth analysis. The software facilitated the segmentation of text into meaningful themes and categories, allowing for a granular examination of interview transcripts. With NVivo's advanced query capabilities and visualization tools, it was possible to trace recurring patterns and construct detailed narratives. This data served as a basis for enhancing various aspects of the planned-for field experimental design, including the formulation of survey inquiries and the creation of video vignettes.

Results

Sample Demographic

Table 1 outlines a comprehensive demographic and background profile of the participants involved in the study. All 30 participants are female, with a majority identifying as White (83%) and a smaller portion as Asian (17%). Their roles in the organization are predominantly within sales (90%), reflecting a strong business orientation. With respect to generational demographics, the sample is diverse, spanning from Baby Boomers to Generation Z, with Generation X being the most represented at 50%. Table 1 further details the participants' experience level, with an average of 49.3 hiring interviews conducted and an average tenure at their company of 11.8 years. The participants' self-assessment on racial bias reveals a spec-

trum of viewpoints, and their Implicit Association Test (IAT) results provide an insight into subconscious associations between ethnicity and national identity. Lastly, Table 1 also explores the participants' comfort levels with interacting with avatars in hiring, a novel aspect of the study that was introduced after initial pre-testing with 10 respondents.

Avatar Creation

This study's success hinges on the avatars being credible, competent, and racially neutral. Avatars, by definition, are digital entities characterized by anthropomorphic appearances, and they possess the capability to engage in interactions, whether controlled by a human or software. (Miao et al., 2022). Anthropomorphism refers to the act of ascribing human characteristics, attributes, or emotions to entities, objects, or animals that are not inherently human. Studies indicate that the perceived level of anthropomorphism in an avatar directly correlates with its perceived credibility and competence (Westerman et al., 2015).

People often categorize racially ambiguous figures as 'other', an outcome the study aimed to avoid (Albuja et al., 2023). To accomplish this, an iterative process was conducted. In step one, three avatar images were created using Bing image creator and were presented to a focus group of students in a Doctorate in Business Administration (DBA) program, encompassing diverse demographics (Microsoft Corporation, 2023). Neutral facial features, medium skin tones, and non-racially specific hair and eye colors were used. In assessing relatability and racial ambiguity, Avatar C ranked highest, with Avatar B close behind.

Step two used ArtGuru's AI Art Generator to then refine the avatars, enhancing their features to avoid stereotypes and promote anthropomorphism (ArtGuru, n.d.). Refinement was complemented with feedback from both AI assessments and DBA student evaluations, creating a feedback loop that grounded the design process in both objective data and subjective human perception. Six variations of each image were generated and evaluated, resulting in Avatar 1 and Avatar 4 being identified as female, least racially identifiable, and the most relatable.

For step three, eight female professionals, matching the criteria for participation in the study, were then engaged to evaluate the avatars. They reviewed sample videos with Avatar A and B, an Asian Avatar, and a White Avatar, from the stock videos available from D-ID, and noted their observations on the avatars' racial characteristics and relatability (D-ID, n.d.). The feedback provided a consensus that race was the most noticeable difference among the avatars, with varied individual responses to racial identification, especially for Avatar A. This group's preference for Avatar B's relatability, as shown in Table 2, was significant in the final selection process.

These images were then run through two AI platforms, based on language models that utilize deep learning, specifically transformer architectures, to generate human-like text based on the input it receives (Microsoft Corporation,

Table 1*Sample Demographic and Background*

Variable	(n = 30)
Gender	
Female	n = 30 (100%)
Race	
White	n = 25 (83%)
Asian	n = 5 (17%)
Completed Interviewer Training as	
Hiring Manager	n = 14 (47%)
Individual Contributor	n = 16 (53%)
Number of Hiring Interviews Involved in	Mean = 49.3 SD = 39.16; Range: 1 to 124
Years Worked at Company	Mean = 11.80 SD = 6.39
Role in Organization's Business	
Sales	n = 27 (90%)
Other (e.g., sales support, operations)	n = 3 (10%)
What Generation Do You Identify As	
Baby Boomer (1946-1964)	n = 3 (10%)
Generation X (1965 to 1980)	n = 15 (50%)
Generation Y (1981 to 1996)	n = 11 (37%)
Generation Z (born after 1996)	n = 1 (3%)
Consider Yourself Racially biased in any Way	
Strongly Disagree	n = 5 (17%)
Disagree	n = 13 (43%)
Somewhat Disagree	n = 2 (7%)
Neither Agree nor Disagree	n = 5 (17%)
Somewhat Agree	n = 3 (10%)
Agree	n = 2 (7%)
Strongly Agree	n = 0
IAT Results	
1 - Strong Automatic association of Euro Am with American and Asian Am with Foreign	n = 7 (23%)
2 - Moderate Automatic association of Euro Am with American and Asian Am with Foreign	n = 7 (23%)
3 - Slight Automatic association of Euro Am with American and Asian Am with Foreign	n = 0
4 - Little to no automatic preference between ethnicity and American or Foreign	n = 6 (20%)
5 - Slight Automatic association of Asian Am with American and Euro Am with Foreign	n = 6 (20%)
6 - Moderate Automatic association of Asian Am with American and Euro Am with Foreign	n = 4 (13%)
7 - Strong Automatic association of Asian Am with American and Euro Am with Foreign	n = 0
Qualification Manipulation Breakdown of Candidate	
Ambiguous	n = 15
Strong	n = 15
How well do you think the candidate's skill set met the job requirements defined in the job Description, where 1 = extremely ambiguous to 9 = extremely strong	Mean = 6.43 SD = 1.99
Does Company Currently Use Avatars in Hiring (not initially asked while pre-testing avatars)	
No	n = 20 (100%)
How comfortable were you interacting with the Avatar (n = 20)	
Extremely Uncomfortable	n = 2 (10%)
Quite Uncomfortable	n = 1 (5%)
Somewhat Uncomfortable	n = 1 (5%)
Slightly Uncomfortable	n = 5 (25%)
Neither Uncomfortable or Comfortable	n = 3 (15%)
Slightly Comfortable	n = 4 (20%)
Somewhat Comfortable	n = 2 (10%)
Quite Comfortable	n = 2 (10%)
Extremely Comfortable	n = 0

2023; OpenAI, n.d.). These platforms were asked to identify the race of the image. Examples of participant responses include:

(Avatar A) "The image appears to depict a humanoid robot or android. Robots and androids do not have a race as they are machines and not biological beings."

(Avatar B) "The image appears to depict an animated or

digital character. It's important to note that animated or digital characters don't have real-world racial identities in the same way humans do. Their features and designs are created by artists, and they might be influenced by various stylistic or cultural elements, but they don't belong to any specific "race" in the traditional sense. If you're asking about the intended ethnic or cultural inspiration for the

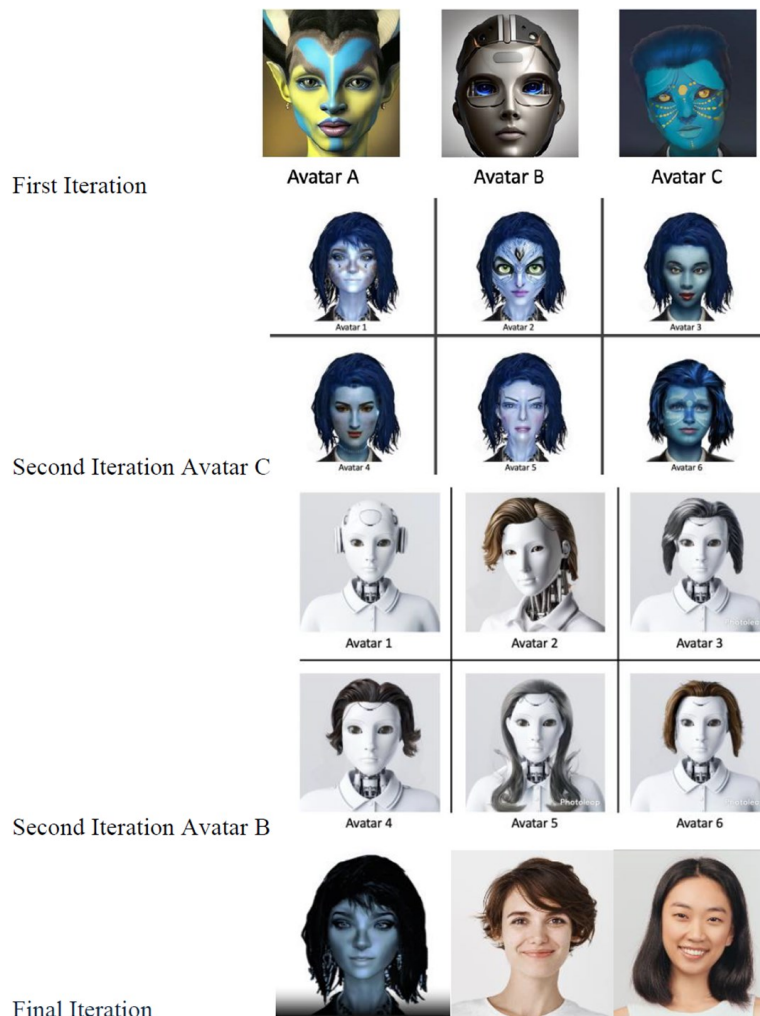
Table 2*Summary of Human Identification of Avatar Traits*

Avatar	ID'd as White	ID'd as Asian	ID'd as Robot	ID'd as None	ID'd as AI	Could not ID	Slight White resemblance	Favored for relatability
White	8	0	0	0	0	0	0	N/A
Asian	0	8	0	0	0	0	0	N/A
Avatar A	1	0	4	2	1	0	0	1
Avatar B	0	0	1	0	5	1	1	7

character's design, that would require context or knowledge about the character's origin or backstory.”

The iterative process and final avatars can be seen in Figure 1. Throughout the process, the nuanced relationship between design choices and viewer perceptions was observed. This iterative, data-informed approach, integrating AI and human feedback, proved essential in navi-

gating these complexities. This rigorous, iterative process underscored the intricate task of crafting an avatar free from racial bias yet personable enough for professional interaction. As avatars continue to evolve, they challenge us to confront and overcome ingrained biases, aiming for a digital representation that transcends racial preconceptions.

Figure 1*Iterative Avatar Creation Process*

Avatar Candidate Qualification Creation Process

A comprehensive job description was crafted using a large technology company's repository of interview questions and job descriptions. Additionally, two distinct video interview transcripts were developed: one with responses that ambiguously matched the job requirements, using interview tips to subtly craft answers that demonstrate potential rather than direct experience; and another with responses that strongly matched the job criteria, demonstrating a clear and direct alignment with the role's expectations.

Creative Reality Studio by D-ID was used to create the video vignettes. "D-ID's Creative Reality Studio is a self-service platform featuring the best generative AI tools to enable users to create videos with moving and talking avatars. Combining the powers of D-ID's deep-learning face animation technology with GPT-3 text generation, and Stable Diffusion text-to-image capabilities to create amazing and cutting-edge videos with the power of artificial intelligence" (D-ID, 2023, para. 1). Uniformity was maintained across all videos by using the same voice, both for the interviewer and the candidate, as well as maintaining consistent expressions.

Pilot One

Pilot one engaged ten participants who, after completing a survey, underwent semi-structured interviews based on a predefined protocol. The interviews were analyzed using NVivo software to identify themes, as illustrated in Figure 2. Feedback on the survey was unanimously positive, praising its structure, clarity of questions, and overall comprehensibility.

During the interviews, participants were shown images of avatars and asked to identify their gender and race. This exercise yielded high agreement rates, detailed in Table 3, particularly for the 'White' and 'Asian' avatars, with 90% identifying the 'White' avatar as female and white, and 100% recognizing the 'Asian' avatar as female and Asian. The 'Blue Avatar', however, was predominantly seen as female but with an 'unknown' race by 90% of participants. Based on this feedback, the avatar designs will be retained for future study phases.

Feedback highlighted the need for more succinct job descriptions. In response, the description has been meticulously condensed to encapsulate essential information more succinctly. Additionally, there was a call for interview responses to include more illustrative examples.

Participants' quotes include:

"I would have preferred to probe with more targeted questions."

"Despite the extensive dialogue, the candidate's failure to provide specific examples left me unconvinced of their suitability for the role."

Responding to critiques about the avatars' speech pace, video speeds were adjusted to 80% of the original. Participants desired evidence of preliminary candidate screening and preferred to influence advancing candidates rather than making final hiring decisions. Consequently, survey

adjustments were made to reflect these preferences, aiming to enhance the study's focus on inclusive hiring and bias reduction.

Pilot Two

Pilot two built on the findings from the first, with ten new participants completing a revised survey and engaging in semi-structured interviews using the same protocol as before. The data were analyzed with NVivo (2020), revealing themes captured in Figure 2. The survey received positive feedback for its design and clarity, similar to the first pilot. Participants showed a high consensus in identifying the gender and race of avatars, detailed in Table 3, for entries 11-20. Specifically, 100% of participants identified both the 'White' and 'Asian' avatars correctly in terms of gender and race. The 'Blue Avatar', while unanimously recognized as female, had 80% of participants uncertain of its race, often labeled as 'unknown'. These findings confirmed the suitability of the current avatar designs for further study stages.

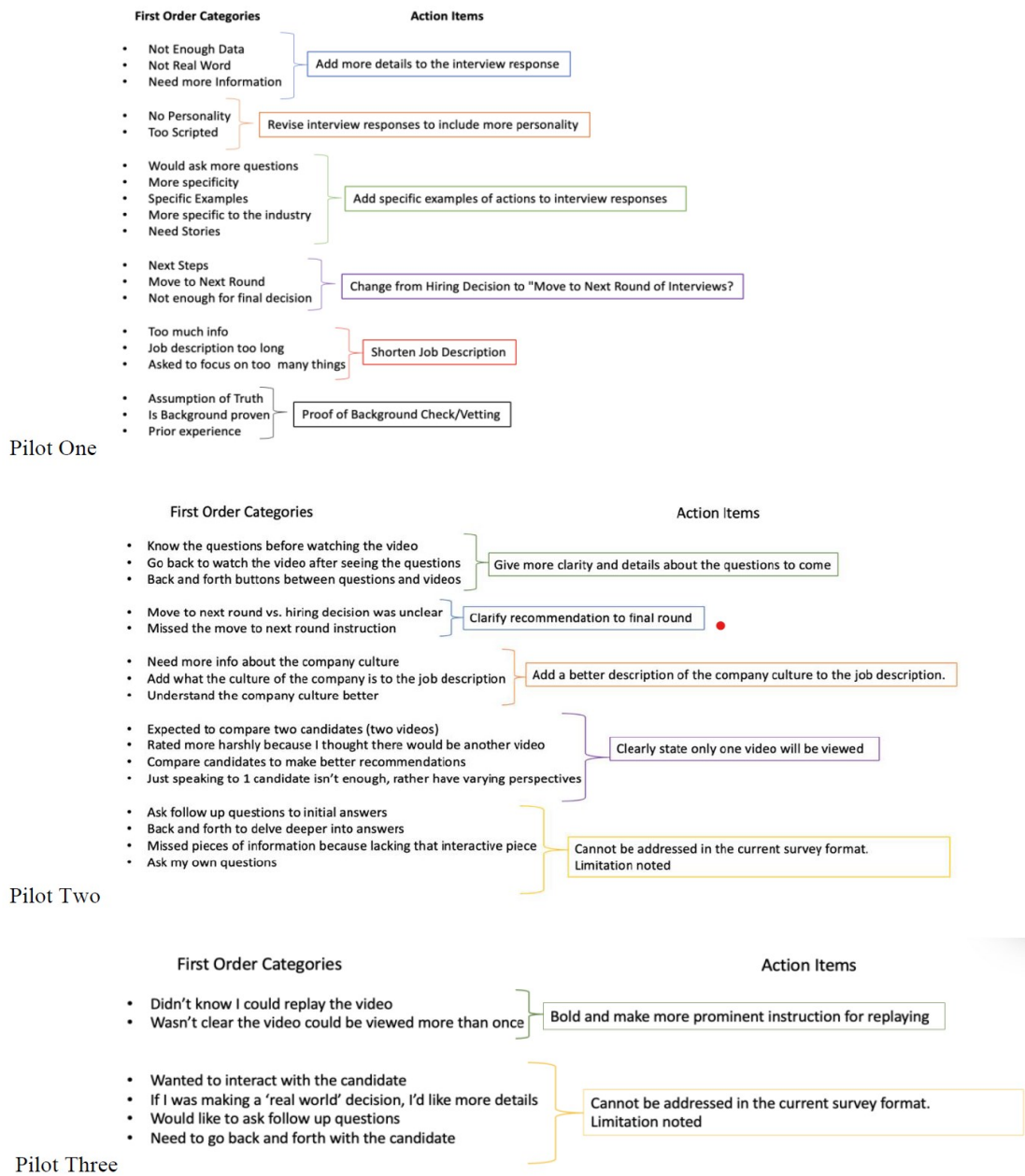
Participants expressed a desire for clearer information on the evaluation criteria before viewing interview videos. One noted, "I wish I had known what the questions were before (watching the interview)." This led to modifications to improve question instructions' specificity and clarity. There was confusion about the interview process's aim, with some unclear if it was for a final hiring decision or advancing to another round, prompting questionnaire adjustments to clarify this.

Feedback also indicated a need for a better understanding of the company's culture to make informed decisions, leading to an enriched job description with more cultural details. Additionally, there was uncertainty about the number of candidates to be evaluated, expecting a format more akin to real interview scenarios. This resulted in guideline revisions for clarity on evaluating a single video. The inability to ask follow-up questions was noted as a limitation, with one participant stating, "without the ability to ask questions I can't make a call on a candidate in general." Though interactive questioning wasn't feasible for this survey, this feedback will be considered for future study limitations.

Analysis showed that participants' confidence in making hiring decisions varied with the candidate's qualification clarity. Those evaluating candidates with a strong match felt more informed (42%) compared to those assessing candidates with ambiguous qualifications (11%), highlighting the importance of clear qualification matching in decision-making confidence.

Pilot Three

Pilot three, continuing from the previous pilots, involved ten new participants who completed a revised survey and engaged in semi-structured interviews using the established protocol. NVivo (2020) software helped analyze the data, revealing themes illustrated in Figure 2. Similar to earlier feedback, participants praised the survey's structure and clarity. The majority accurately identified the gender and race of avatars, with uniform agree-

Figure 2*Iterative First Order Categories and Action Items – Three Iterative Pilot Studies*

ment on the 'White' and 'Asian' avatars as female and of their respective races, as shown in Table 3, for entries 21-30. The 'Blue Avatar' saw consensus on gender but not race, with 70% labeling its race as 'unknown'.

The feedback led to maintaining the avatar designs for future study phases. Participants preferred the strong interview transcript over the ambiguous one, with a 97% agreement rate. Adjustments were made for clearer video replay instructions. The limitation of not enabling follow-up questions was highlighted, with a participant noting, "Interactive conversation is where you can really dig in

and candidates will shine or trip up." This feedback, especially from those viewing the ambiguous fit interview, underscores the importance of interactive elements in interviews, a consideration for future research discussions.

Overall Pilot Results

Final Survey

By the end of Pilot Three, participants demonstrated a consensus on the enhanced clarity of the job description,

Table 3*Iterative Identification of Avatar Features*

Subject	White Avatar Gender Assumption	White Avatar Race Assumption	Asian Avatar Gender Assumption	Asian Avatar Race Assumption	Avatar Gender Assumption	Avatar Race Assumption
1	Female	White	Female	Asian	Female	AI generated mixed race
2	Female	White	Female	Asian	Female	No
3	Female	Caucasian	Female	Asian American	Female	Couldn't Tell
4	Female	White	Female	Asian American	Female	Can't tell
5	Female	White	Female	Asian	Female	Avatar
6	Female	White	Female	Asian/Mix	Female	Really Hard - European Mix
7	unknown	unknown	Female	Asian	unknown	unknown
8	Female	Caucasian	Female	Asian	Female	Couldn't Tell/Mix
9	Female	Caucasian	Female	Asian	Female	Do not know
10	Female	Caucasian	Female	Asian Descent	Female	Cannot Tell
11	Female	White	Female	Asian	Female	No
12	Female	White	Female	Asian	Female	Cannot Tell
13	Female	White	Female	Asian	Female	Can't distinguish
14	Female	White	Female	Asian	Female	Ambiguous
15	Female	Caucasian	Female	Asian	Female	Unclear/ some sort of European
16	Female	Caucasian	Female	Asian	Female	Not really sure
17	Female	White	Female	Asian	Female	No
18	Female	Caucasian	Female	Asian	Female	Gets Harder/ maybe European
19	Female	Caucasian	Female	Asian	Female	Not Sure
20	Female	Caucasian or Latina	Female	Asian	Female	Mixed
21	Female	Caucasian	Female	Asian	Female	probably Caucasian
22	Female	European	Female	Asian	Female	Can't tell
23	Female	Caucasian	Female	Asian	Female	No
24	Female	European	Female	Asian	Female	Alien
25	Female	Caucasian	Female	Asian	Female	Like European
26	Female	Caucasian	Female	Asian	Female	I have no idea
27	Female	Caucasian	Female	Asian American	Female	Don't know
28	Female	White	Female	Asian	Female	Caucasian maybe
29	Female	American	Female	Asian	Female	I don't know
30	Female	Caucasian	Female	Asian	Female	I have no idea

interview content, and survey questions, reflecting the effective modifications made in response to previous feedback. The iterative improvements in survey design, clearer instructions, and more detailed company cultural information contributed to a more straightforward and comprehensible process for the participants, including the

ease of interacting with the IAT components. Overall participant gender and race assumptions were: for the White avatar: 97% female and 97% White; the Asian avatar: 100% female and 100% Asian; and the Blue avatar: 97% Female, 83% Unknown and 17% Maybe European/Caucasian. Thus, these avatars will be used for the experi-

mental study. By the end of the pilot study, the saturation point had been reached (NVivo, 2020), and no additional modifications were needed.

Feelings about the IAT

Responses to the Implicit Association Test (IAT) showcased a range of emotional reactions, highlighting the test's ability to provoke deep personal reflection on ingrained biases. One participant's comment, "I didn't like answering the questions, (but) it's what I was taught," encapsulates the discomfort arising from confronting learned biases, suggesting an internal conflict between current beliefs and ingrained teachings. Another participant described the experience as "awkward and uncomfortable," a reaction that underscores the emotional dissonance and cognitive challenges posed by the test. These expressions, combined with the score distribution in Table 1, reveal the IAT's effectiveness in surfacing subconscious biases, leading to self-awareness and potential discomfort due to the recognition of previously unrecognized or unacknowledged biases. Such reactions emphasize the emotional complexity of addressing implicit biases, demonstrating the IAT's role in fostering personal growth and understanding of deep-seated attitudes and conditioning.

Ambiguous Versus Strong Job Candidate Qualification Manipulation Check

The Mean (M) for perceived job candidate qualifications for the ambiguous qualification manipulation was $M = 5.80$ versus for the strong qualification manipulation it was $M = 7.07$. A one-tailed t-test, found $t(28) = -1.81$ (where 1 = Ambiguous, 2 = Strong), is $p < .05$. This result supports the effectiveness of the vignette depiction of ambiguous versus strong candidates.

Discussion

Proposed Field Experiment

Following the successful pilot, the field experiment will be launched to test the research question: To what extent, and in what ways, can utilizing avatars to mask racial identity during the interview process mitigate implicit in-group bias? The independent variables in this research are candidate race and qualifications, and the dependent variables are hireability of the candidate. The participants'

existing implicit bias as measured by the IAT Test will be used as a moderator.

To achieve a sample size of 1,000 participants, with an equal distribution of 500 White and 500 Asian individuals, recruitment will be conducted through an online crowdsourcing platform (SurveySwap, n.d.). This strategy is effective for engaging a virtual workforce and ensuring a balanced representation of demographic variables. The platform's wide reach and diverse user base will be instrumental in meeting the targeted recruitment numbers, facilitating a well-distributed and ample participant pool for the study.

Participants will read a job description, then watch a video vignette of one candidate answering interview questions. After screening, eligible participants will be randomly assigned to one of six video vignettes in a two (qualifications: clearly strong or ambiguous,) x three (race of candidate – represented as an Asian avatar, a White/Caucasian avatar, and a racially unidentifiable avatar) design like the process conducted by Dovidio and Gaertner (2000) as seen in Table 4. These vignettes will be less than 5 minutes long and participants will be given the opportunity to rewatch if they choose. The participants will then be asked a series of questions designed to assess hireability of the candidate.

Finally, they will be asked to take the IAT to identify the existing implicit bias level of the participants (Project Implicit, n.d.). It is important to note that the IAT is not without its criticisms and limitations. Some argue that the IAT may not always accurately capture implicit biases or predict behavior in real-world settings (Fazio & Olson, 2003).

Research Contribution

The research presented in this paper significantly contributes to the existing literature regarding the development of generative AI avatars to be used for hiring in the workplace. It also shows how there can be a disconnect between experienced hiring manager self-assessments denying racial bias versus IAT results revealing such bias, as seen in Table 1. It begins to delve into an innovative exploration of how these digital personas can be used in hiring situations. Combining candidate avatars with the successful manipulation of perceived strong versus am-

Table 4

Experimental Treatment Combinations

		Candidate Race		
Candidate Qualification		White	Asian	Avatar
	Strong	Hiring-related Variables	Hiring-related Variables	Hiring-related Variables
	Ambiguous	Hiring-related Variables	Hiring-related Variables	Hiring-related Variables

biguous candidate qualifications, sets the stage for testing their influences on interviewer hiring recommendations. Such testing will hopefully extend the scholarly discourse within the realms of discrimination, stereotyping, implicit attitudes, and group identity, as previously discussed in seminal works by Dovidio and Gaertner (2000), and Lee et al. (2015).

Limitations

The limitation of the interviewer not being able to interact directly with candidates represents a significant challenge in the context of this study. Interaction often provides depth and nuance that static survey responses or avatar representations cannot capture. The absence of a dynamic exchange means that evaluators are deprived of the opportunity to probe deeper into a candidate's responses, to seek clarification, or to gauge reactions to unplanned follow-up questions. This restriction can lead to a superficial assessment, potentially glossing over the candidate's full range of abilities, personality traits, and the subtleties of their thought processes. It also means that the spontaneous elements of human interaction, such as body language, tone of voice, and the ability to build rapport, are missing, which could otherwise offer valuable insights into the candidate's suitability for a role.

The exclusive use of the Implicit Association Test (IAT) to measure implicit bias towards Asians may not fully encapsulate the complexity of such biases. A test-retest reliability of each participant's IAT score would ideally substantiate the findings, yet this is formidable to execute. The study focuses on the ramifications of candidate race and qualifications on hiring decisions, a pertinent aspect, yet it does not account for other influential variables like candidate experience, interview performance, and organizational factors that might also sway recruitment outcomes (Bertrand & Mullainathan, 2022).

The pilot study focus is limited due to the exclusive selection of female hiring managers and experienced hiring contributors from a technological corporation in the United States. This sample will be extended across technology corporations in the field experiment which will affect the generalizability of its findings across different industries and organizational contexts, as discussed in Bertrand and Mullainathan (2022). This limitation is particularly notable given the study's pilot sample and future experiment sample do not explore the intersectionality of multiple diversity characteristics, such as race and gender, due to its all-female participant base. Additionally, the focus on only Asian and White/Caucasian participants does not capture the full spectrum of workforce diversity, and the reliance on self-identification for racial categorization might not accurately reflect participants' cultural backgrounds or experiences, potentially introducing bias into the research outcomes. As such this current research underscores the need for future research to address the intersectionality of race and gender, particularly in the context of AI-facilitated hiring practices. Thus, this study sets the stage for more comprehensive inquiries that can provide a deeper understanding of diversity, equity, and inclusion within organizational settings.

Ethical Implications

Venturing into the practice of anonymizing identities during the interview process brings forth an array of ethical considerations that intertwine with challenges pertaining to authenticity, fairness, and ethical conduct in hiring. The rationale for anonymizing a candidate's identity lies in reducing biases associated with visible traits such as gender, ostensibly to foster a more impartial hiring procedure. However, this brings forth complexities concerning the sincerity and comprehensive evaluation of a candidate. The act of masking one's identity may engender a sense of discomfort or inauthenticity, potentially affecting the candidate's performance during interviews adversely. Such obfuscation can hinder authentic interactions between the interviewer and the candidate, veiling the candidate's self-presentation—a critical element in the evaluation process. Given that the perceptions of others partially determine various material, social, and personal outcomes for an individual, the emphasis on being perceived positively is understandable (Brown & Levinson, 1987; Leary et al., 1994). A thorough contemplation of the ramifications of obscuring identity during the hiring process is imperative to ensure the best interests of both the candidate and the organization are preserved. The distribution of pilot sample scores for comfort level when interacting with the avatar supports this. It's important to balance reducing bias with maintaining a transparent and authentic interview process.

Using AI-generated avatars with racial identifiers for candidate representation during interviews poses ethical challenges. This approach may unintentionally emphasize racial distinctions, affecting impartiality in evaluation. The intention to foster inclusivity could backfire by simplifying complex identities into broad, potentially inaccurate categories. Ensuring fairness requires scrutinizing how these avatars are created to prevent embedding or magnifying biases. It's essential for developers and organizations to carefully consider the impact of such technology, prioritizing ethical guidelines to uphold the integrity of the hiring process. Engaging diverse perspectives in the development and review of these systems can help address these concerns, ensuring that innovation does not compromise ethical standards.

Future Research

There are many avenues for future research, extending the finding here and delving into adjacent topics. Introducing interactive Extended Reality (XR) avatars into future studies could mitigate this limitation significantly. By simulating a more realistic interaction scenario, evaluators could engage with virtual representations of candidates in a controlled yet flexible environment. This technology would allow for the nuanced interaction that is currently missing, providing a platform for follow-up questioning and the assessment of non-verbal cues. Through XR, evaluators could experience a candidate's responses in real-time, allowing for an evaluation that more closely mirrors an in-person interview. This could enhance the reliability of the assessment outcomes and

offer a more holistic view of the candidate, thus overcoming one of the critical limitations of the current study design.

Expanding the participant pool would enhance the generalizability of the findings, future research could include participants from a wider range of industries or organizations, encompassing both male and female participants, and participants of other races. This would help to explore the effects of avatars and implicit biases in various contexts. This research focuses on candidate race and qualifications, there may be other variables that influence hiring decisions. Exploring factors such as candidate experience, interview performance, personality traits, or organizational characteristics could provide a more comprehensive understanding of the hiring process and potential biases (Kristof-Brown, 2000).

The potential of artificial intelligence (AI) to assess deeper-level aspects of diversity, such as cognitive thinking styles, values, and attitudes, presents a complex blend of opportunities and challenges. AI's capability to conduct nuanced assessments can offer profound insights into human diversity, enabling predictive modeling and the detection of biases to promote fairness and inclusivity in various societal contexts. However, ethical concerns regarding privacy, consent, and the risk of data misuse, coupled with the inherent complexity of accurately capturing the multifaceted nature of human diversity and the potential for biases within AI algorithms, necessitate a cautious and multidisciplinary approach. Future research must therefore prioritize the development of ethical frameworks, enhance the accuracy and fairness of AI tools, and encourage interdisciplinary collaboration. By addressing these challenges, AI has the potential to significantly contribute to a deeper understanding of diversity, leading to more inclusive and equitable communities.

Exploring the potential of AI to proactively identify and recruit candidates who bring both deep-level and surface-level diversity into organizations represents a compelling direction for future research. This approach to leveraging AI in hiring processes goes beyond merely mitigating biases; it involves actively seeking out candidates whose diverse experiences, cognitive styles, values, and backgrounds can contribute to a richer, more innovative organizational culture. Such a strategy recognizes diversity as a key driver of organizational success, enabling companies to tap into a broader range of perspectives, problem-solving approaches, and creative ideas. By utilizing AI algorithms designed to highlight not just qualifications but also diverse attributes and potential contributions to organizational diversity, companies can foster a more inclusive environment. This not only enhances their competitive advantage but also aligns with broader goals of equity and inclusion. Future studies could investigate the efficacy of AI tools in this capacity, assessing how they can be optimized to support strategic diversity objectives and contribute to the cultivation of a diverse and dynamic workforce.

Conducting a longitudinal study over an extended period could reveal how implicit biases and the impact of avatars and XR in the workplace evolve over time. This

approach would provide valuable insights into the long-term effects and potential changes in biases and behaviors.

References

- AI4.Tools. (n.d.). *ArtGuru*. <https://ai4.tools/ai-tools/artguru-2/>
- Albuja, A. F., Muñoz, M., Kinzler, K., Woodward, A., & Gaither, S. E. (2023). Hypodescent or ingroup over exclusion? Children's and adults' racial categorization of ambiguous black/white biracial faces. *Developmental Science*, 27(2), e13450. <https://doi.org/10.1111/desc.13450>
- ArtGuru AI. (n.d.). *ArtGuru AI*. <https://www.artguru.ai/>
- Bar-Anan, Y. (2020). *Running Project Implicit's IAT from Qualtrics*. <https://minnojs.github.io/minnojs-blog/qualtrics-iat/>
- Bertrand, M., & Mullainathan, S. (2022). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94(4), 991-1013. <https://www.jstor.org/stable/3592802>
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Çoksan, S., & Cingöz-Ulu, B. (2022). Group norms moderate the effect of identification on ingroup bias. *Current Psychology*, 41(1), 64-75. <https://doi.org/10.1007/s12144-021-02091-x>
- Cox, T., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Executive*, 5(3), 45-56. <https://doi.org/10.5465/ame.1991.4274465>
- Csonka-Peeren, J., & Cozzarin, B. P. (2021). A conceptual model of ambiguity attitudes for entrepreneurship experiments. *Journal of Behavioral and Applied Management*, 21(1), 62-80. <https://doi.org/10.21818/001c.29692>
- Derfler-Rozin, R., & Pitesa, M. (2020). Motivation purity bias: Expression of extrinsic motivation undermines perceived intrinsic motivation and engenders bias in selection decisions. *Academy of Management Journal*, 63(6), 1840-1864. <https://doi.org/10.5465/amj.2017.0617>
- Deros, E., Ryan, A.M. (2018). When your resume is (not) turning you down: Modelling ethnic bias in resume screening. *Human Resource Management Journal*, 29(2), 113 - 130. <https://doi.org/10.1111/1748-8583.12217>
- D-ID. (n.d.). *FAQs*. from <https://www.d-id.com/faqs/>
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54(1), 297-327. <https://psycnet.apa.org/doi/10.1146/annurev.psych.54.101601.145225>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27. <https://psycnet.apa.org/doi/10.1037/0033-295X.102.1.4>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social*

- Psychology*, 74(6), 1464-1480. <https://psycnet.apa.org/doi/10.1037/0022-3514.74.6.1464>
- Karras, T., Laine, S., & Aila, T. (2021). A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4217-4228. <https://doi.org/10.1109/TPAMI.2020.2970919>
- Leary, M. R., Tchividjian, L. R., & Kraxberger, B. E. (1994). Self-presentation can be hazardous to your health: Impression management and health risk. *Health Psychology*, 13(6), 461-470. <https://doi.org/10.1037/0278-6133.13.6.461>
- Lee, S. Y., Pitesa, M., Thau, S., Pillutla, M. M. (2015). Discrimination in selection decisions: integrating stereotype fit and interdependence theories. *Academy of Management Journal*, 58(3), 789-812. <http://dx.doi.org/10.5465/amj.2013.0571>
- Lira, M., Egito, J. H., Dall'Agnol, P. A., Amodio, D. M., Gonçalves, O. F., & Boggio, P. S. (2017). The influence of skin colour on the experience of ownership in the rubber hand illusion. *Scientific Reports*, 7(1), 15745. <https://doi.org/10.1038/s41598-017-16137-3>
- Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An emerging theory of avatar marketing. *Journal of Marketing*, 86(1), 67-90. <https://doi.org/10.1177/0022242921996646>
- Microsoft Corporation. (2023). *Output generated by Bing AI* [Artificial Intelligence Software]. <https://www.bing.com>
- Microsoft. (n.d.). *How to create digital art with Bing Image Creator. Microsoft Bing*. <https://www.microsoft.com/en-us/bing/do-more-with-ai/how-to-create-digital-art-with-bing-image-creator?form=MA13KP>
- NVivo (2020). *Version 12. Software. QSR International*. <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>
- OpenAI (n.d.). *ChatGPT Version 4*. [Software]. <https://chat.openai.com/>
- Project Implicit. (n.d.). *Select a Test. Project Implicit - Harvard Implicit Association Test*. <https://implicit.harvard.edu/implicit/selectatest.html>
- Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology*, 32, 299-328. <https://doi.org/10.1146/annurev.soc.32.061604.123132>
- Quillian, L., Midtboen, A. H. (2021). Comparative perspectives on racial discrimination in hiring: the rise of field experiments. *Annual Review of Sociology*, 47, 391-415. <https://doi.org/10.1146/annurev-soc-090420-035144>
- Roberson, Q. M. (2019). Diversity in the workplace: A review, synthesis, and future research agenda. *Annual Review of Organizational Psychology and Organizational Behavior*, 6(1), 69-88. <https://doi.org/10.1146/annurev-orgpsych-012218-015243>
- Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17, 523-534. <https://doi.org/10.1016/j.labeco.2009.04.005>
- SurveySwap. (n.d.). *Home page*. <https://surveyswap.io/>
- Tajfel, H. (1979). Individuals and groups in social psychology. *British Journal of Social and Clinical Psychology*, 18(2), 183-190. <https://doi.org/10.1111/j.2044-8260.1979.tb00324.x>
- Temple University. (n.d.). IRB Forms & Standard Operating Procedures. <https://research.temple.edu/research-compliance/institutional-review-board-irb/irb-forms-standard-operating-procedures>
- Westerman, D., Tamborini, R., & Bowman, N. D. (2015). The effects of static avatars on impression formation across different contexts on social networking sites. *Computers in Human Behavior*, 53, 111-117. <https://doi.org/10.1016/j.chb.2015.06.026>
- Young, A. M. (2022). Women's presence on corporate boards: Addressing director selection, participation, and gender bias. *Journal of Behavioral and Applied Management*, 22(2), 108-129. <http://dx.doi.org/10.21818/001c.37168>
- Zipppia. (n.d.). *Human Resource Specialist Demographics*. <https://www.zipppia.com/human-resource-specialist-jobs/demographics/>

Alexa S. Trifilo (alexa.trifilo@temple.edu)

Gary Blau (gary.blau@temple.edu)
