Dec 12th, 12:00 AM

Asynchronous Video Interviews and Artificial Intelligence: A Multi-Study Exploration

Tanner Skousen
University of Georgia, tanner.skousen@uga.edu

Jacob Steffen
University of Georgia, steffen.h.jacob@gmail.com

Cherileigh L. Chandler
Brigham Young University, cherileigh.leavitt@gmail.com

Warren Rosengren
BYU, wrosengren16@gmail.com

James Gaskin
BYU, james.eric.gaskin@gmail.com

See next page for additional authors

Presenter Information

Tanner Skousen, Jacob Steffen, Cherileigh L. Chandler, Warren Rosengren, James Gaskin, and Tom Meservy

Asynchronous Video Interviews and Artificial Intelligence: A Multi-Study Exploration

Completed Research Paper

Tanner Skousen
University of Georgia
B425 Amos Hall, Athens, GA 30602
tanner.skousen@uga.edu

Jacob Steffen
Brigham Young University
787 TNRB, Provo, UT 84602
steffen.h.jacob@gmail.com

Cherileigh Chandler
Brigham Young University
790 TNRB, Provo, UT 84602
cherileigh.leavitt@gmail.com

Warren Rosengren
Indiana University
107 S Indiana Ave, Bloomington, IN 47405
wroseng@iu.edu

James Gaskin
Brigham Young University
785 TNRB, Provo, UT 84602
james.gaskin@byu.edu

Thomas O. Meservy
Brigham Young University
782 TNRB, Provo, UT 84602
tmeservy@byu.edu

Abstract

Asynchronous video interviews (AVIs) provide scalable, low-cost opportunities for matching interviewees and organizations. However, the implications of a shift from synchronous interviews aren't fully understood, especially when design choices such as AI evaluations are employed. To better understand the impact of AVIs, we undertook an exploratory qualitative study in addition to an experiment. The first study involves 100 qualitative responses and exploratory quantitative tests on the relationships between coded values and demographic and trait variables of the respondents. Our second study tests the impact of AI feedback using a large online AVI service while accounting for various disadvantaged groups that could experience discrimination in their AVI interactions. We developed 5 propositions regarding the interaction of interviewee traits and AVI design. Additionally, we did not find support that AI feedback increases the performance of interviewees, though we identify several traits that lead to high AI scores and human-rater performance.

Keywords: Asynchronous Video Interviews, Hiring, Social Justice, Artificial Intelligence, Discrimination

Introduction

Asynchronous video interviews (AVIs) are characterized as a one-way qualification assessment in which the interviewee enters an online portal and records a video of themselves responding to a set of predetermined prompts (Lukacik et al. 2020). In contrast to traditional face-to-face or synchronous video interviews (e.g., Zoom interviews), AVIs streamline the application process by replacing tasks typically completed by an employee (e.g., asking questions, setting up appointments) with software functionality. Although the technology is still maturing, the COVID-19 pandemic acted as a catalyst for the popularity of AVIs since many in-person processes were replaced by technology (Rubinstein 2020). Taken together, the overall trend to automate business procedures, the acceleration of software capability, and the heightened awareness of AVIs during the pandemic suggest that AVIs are here to stay, even post-pandemic.

The shift from synchronous interviews to AVIs has several implications for both organizations and interviewees. For organizations, AVIs offers a quick, low-cost mechanism to evaluate a large volume of interviewees. Typically, the technology is provided as a service by a third-party company, such as HireVue, VidCruiter, SparkHire, etc. With the rise of Software-as-a-Service models, companies are flocking to AVI services. For example, HireVue completed its 20 millionth video interview recently (HireVue 2021). However, depending on the implementation, AVIs can also have negative consequences regarding perceptions towards the organization (e.g., dehumanizing a traditionally interactive experience).

For applicants, the outcomes of AVI use are likewise multivalent. Some AVIs provide applicants with question lead-time for preparation, allow for multiple attempts, and can be completed within a generous time window. Empowered by the internet, AVIs also democratize the process since individuals from diverse geographic locations can enter the candidate pool. Further, as witnessed during the COVID-19 pandemic, AVIs can prevent extreme events from halting the hiring process. However, AVIs notably lack the richness and interpersonal connection that is often associated with direct human interaction. This reality frustrates those who rely upon impression management or symbiotic communication patterns. Moreover, some AVI designs do not provide question lead time or the chance for multiple attempts. Additionally, AVIs may disproportionally burden those of lower socio-economic groups since those in such groups are less likely to be able to record their responses in a quiet, distraction-free location with high-quality technology. Understanding the effects of AVIs on disadvantaged groups is of particular importance, especially with recent urgent calls to understand the ways technology affects social justice and disadvantaged groups. The United Nations includes opportunities for work as one of its six key areas of inequality that need to be addressed (Baudot 2006). This research aims to uncover how AVIs may help or hinder those disadvantaged groups.

One particularly important AVI design implementation option is the integration of AVIs with artificially intelligent (AI) evaluation technologies. Some AVI systems allow for the automated rating of applicant recordings. Ratings can then be used to screen potential applicants, or in some cases, provide feedback to the applicants themselves. Some AI features that provide actionable feedback to users include whether one should rerecord with faster speech, change the lighting or sound quality, or avoid negative and filler words. AI technology may improve the processing ability of organizations, but the full impact on an array of potential outcomes is not very well understood. There may be an increased likelihood of false positives and false negatives in the hiring decision-making process. For example, a candidate who is otherwise qualified for the position may be disadvantaged based on exogenous variables (e.g., filming the video in a noisy apartment versus a private home office) that an AI could judge too harshly without nuance. Therefore, AI inclusion into the AVI process is an increasingly critical concern that plays a role in how AVIs help or hinder interviewees seeking employment.

The impacts of AVIs, especially effects of incorporating AI into AVIs, is not well understood for its effects on either the employing organizations or the applicants. As we pioneer into the fourth industrial revolution, business processes—including interviewing—are likely to be increasingly automated. Despite the ubiquity of AVIs in practice, there is a paucity of research on AVIs within IS research. This reality is unfortunate since the IS field is particularly well-suited to address the intersection of business, people, and technology. Our manuscript seeks to explore this imbalance by asking the following research questions:

> *RQ 1: How do interviewees describe their experiences and perceptions of asynchronous video interviews?*
>
> *RQ 2: How do specific AVI characteristics influence perceptions of AVIs?*
>
> *RQ 3: How does AI feedback in AVI design affect the performance of participants?*
>
> *RQ 4: How does the use of AI feedback in AVIs affect disadvantaged groups?*

After briefly reviewing relevant literature, we report on two studies that were undertaken to address our research questions. The first study involves 100 qualitative responses regarding AVIs and exploratory quantitative tests on the relationships between coded values and demographic and trait variables of the respondents. Based on these preliminary findings, we develop five propositions. Our second study was informed by our first study and aims to specifically test the impact of AI feedback in an AVI setting while accounting for disadvantaged groups that could experience social injustice in their interactions with AVIs.

Literature Review

Research on job interviews dates back decades (Caldwell and Burger 1998; Stevens 1998). The majority of interview-focused research has occurred within organizational and applied psychology circles (see McCarthy et al. (2017) for a review), with common research questions focusing on impression management (Wilhelmy et al. 2016) and personality (Hiemstra et al. 2019). However, the proliferation of technology has empowered both practitioners and researchers to explore the effect of technology in this traditional face-to-face process (Woods et al. 2019). Understanding how digitally mediated interviews differ from traditional interviews is a key area of inquiry. For example, studies have shown that synchronous video interviews are not perceived as positively as traditional interviews (Blacksmith et al. 2016). Although IS researchers have previously explored the phenomenon of interviews (Pentland et al. 2018) and asynchronous technologies (Suen et al. 2019a), we are aware of only a handful of studies that have explicitly studied AVIs, as they are a relatively nascent area of research. Most extant research examines participant perceptions of AVIs. For example, some studies indicate that AVIs tend to be rated lower in fairness (Basch et al. 2021), though others note potential positives, such as a positive perception on AVIs providing an opportunity to perform (Zibarras et al. 2018). How positively and fairly AVIs are perceived appears to be related to how an AVI is implemented and its exact arrangement of features, though how exactly each feature of AVIs affects various measures of interviewee perception is not clear (Lukacik et al. 2020).

One important feature of some AVIs that is not yet well understood is automated assessment through artificial intelligence. Artificial intelligence has been used as a tool for decision-making in many business and social processes such as employment (Barocas and Selbst 2016), criminal justice (Corbett-Davies et al. 2017), medicine (Kononenko 2001), and education (Wang 2021). The primary objective of AI algorithms in these scenarios is to make optimal and often complex decisions based on available input data (Gillespie 2014). However, recent evaluations of many AI use case scenarios highlight how these AI systems are biased and unfair (Dastin 2018). This potential bias is concerning if people perceive AI as a fair system and accept their decisions at face value (Araujo et al. 2020). Thus, an important new standard for AI implementations is algorithmic fairness, which can be defined as an algorithm where "otherwise similar individuals should not be treated differently due to having different protected attributes" (Paulus and Kent 2020, p. 3). Paulus and Kent (2020) gloomily share, however, that this ideal standard is a mathematical impossibility. Regardless, researchers have acknowledged the bias in input data and algorithmic design and have started to develop advancements to produce and promote fairer AI systems (Felzmann et al. 2020; Mehrabi et al. 2021). Thus, AI fairness is being handled by many unique solutions such as using constrained optimization rules (Corbett-Davies et al. 2017), employing human-AI hybrid systems (Jarrahi 2018), and simply by designing more transparency into the algorithm (Felzmann et al. 2020).

How AI systems affect the AVI hiring process is still a relatively nascent area of research. Many studies examine AIs from the perspective of the employer, e.g., how they can make better hiring decisions. For example, AI technologies can parse the interviewees' words in addition to analyzing their response structure for the presence of certain phrases as well as the style and complexity of participants' speech (Johnson and Gray 2019). The development of AI tools in AVI contexts is improving, even to the extent that AI tools can now identify personality cues with over 90% accuracy (Suen et al. 2019b). Additionally, IS researchers have developed applications and patents for automatically detecting deception in videos (Proudfoot et al. 2016; Twyman et al. 2016) that can also be utilized to inform hiring decisions. Others have noted the need for domain experts to still be involved in the creation and maintenance of AI tools for hiring (van den Broek et al. 2021). For example, van den Broek et al. (2019) conducted an ethnographic study of hiring algorithms and identified how those algorithms affected fairness perception within a large multinational company. Interestingly, they find that the introduction of AI into the interview process stimulates and shapes individuals' assumptions about ethics both in the context of interviews and in general. One study measured perceived level of fairness and found that AI-scored interviews were perceived by applicants similarly to interviews judged by humans (Suen et al. 2019a), though another study suggests that applicants may be more critical of AI-evaluated interviews in high-stakes situations (Langer et al. 2018).

Because of the novelty of the phenomenon, many unanswered questions remain regarding the impact of AI in AVI contexts. Despite the inroads made by IS researchers who focus on interviews, little empirical data speaks to the perceptions of AVIs from an interviewee's perspective. Our studies aim to uncover additional insights regarding how AVIs are perceived, in hopes to extend prior research that has examined user perceptions of AVIs. Additionally, we seek to understand how AI in AVIs could impact applicants' performance, rather than just help in decision making for the employer. We adopt an exploratory multi-study, multi-method approach to examine our research questions.

Study 1

Since there is scant empirical work on AVIs, in Study 1, we first solicited and inductively analyzed open-ended survey data using content analysis. Content analysis is a research method "for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh and Shannon 2005, p. 1278). Additionally, as part of content analysis, we adopt a quantitative analysis of the qualitative data (Morgan 1993). Through this process, we were able to identify preliminary relationships between themes.

Instrument and Participants

We recruited 100 participants from the online panel Prolific, which is a robust online platform for soliciting survey respondents (Peer et al. 2017). After providing consent per Institutional Review Board oversight, the respondents were presented with a working definition of AVIs and then reported on past AVI experiences. Since we aimed to solicit perceptions of AVIs and lived experiences, we asked different questions based on whether the participants had previously engaged with an AVI as part of a job application. Of the 100 participants, 45 reported that they had been involved with an AVI as part of a job application, and 55 indicated that they had not.

We asked those who had previously interacted with an AVI to respond to the following questions: (1) *Please describe (with as much detail as possible) your experience with AVIs* (2) *How did the format of the AVI shape your attitude towards the* ***organization*** *you were applying to?* (emphasis original) (3) *If possible, would you change anything about the AVI experience?* (4) *What tips would you give to a friend who will be applying to jobs that use AVIs in the interview process?* (5) *Is there anything else that was interesting, frustrating, or surprising about the AVI experience?*

After carefully describing the AVI process, we asked those who had not interacted with an AVI to respond to the following questions: (6) *What would be some of your concerns about using an AVI (as opposed to a traditional face-to-face interview)?* (7) *In contrast to your preparation for a face-to-face interview, what steps would you take (or not take) to prepare for an AVI?*

Furthermore, both groups of respondents were asked the following question: (8) *Do you believe that the AVI format would help or hurt your chances? Why? (if applicable, consider your social confidence, perceived fairness, social anxiety, techno-anxiety, your personality, preferences, etc.).* Finally, respondents were asked questions regarding their social anxiety and technology competence, and if they perceived themselves as a member of an underprivileged group – all of which were theorized to have some potential impact on an application process using AVIs.

Qualitative Analysis

Our analysis started with open coding, where the researcher avoids *a priori* categories and allows codes to emerge from the data (Glaser and Strauss 1967). Responses from each open-ended question were assigned to a member of the research team to complete the open coding. In total, six members assisted in this step. An initial set of patterns and themes were identified for each question, resulting in 148 first-level codes. Subsequently, the team met together to review the findings of the initial open codes. Based on this collaboration, members of the authorship team then re-evaluated and/or re-coded their assigned questions to generate second-level codes that demonstrated common factors influencing the perception of AVIs among respondents. This step reduced the total number of codes to 44 across the eight questions. Researchers then re-coded each question using this reduced set of codes. After the second pass at coding, researchers again met and consolidated these codes into semantically similar groups that helped inform our theory building related to AVIs. Table 1 demonstrates some example excerpts and associated second-level codes that illustrate the range of our findings.

Concept	Description	Example Excerpt
Multivalent Experience	The AVI experience was both positive and negative	"My experience was fine, I had no issues with the technology and was able to answer questions multiple times to get the result I wanted. But I dislike them and would rather have a regular video interview."
Negative Experience	The AVI experience was negative only	"I felt awkward doing it and stressed my answers too much. I felt unnatural and didn't enjoy the process."
Positive Experience	The AVI experience was positive only	"I enjoy them because it is less pressure to respond to questions... I feel like I felt more confident this way."
Caring	A company using AVI is perceived as caring by the respondent	"I liked the organization more because I felt that the organization really cared about me and my health."
Impersonal Attitude	A company using AVI is perceived as impersonal by the respondent	"I have to say it did feel much less personal and I did not feel a strong connection to the company like I normally would in an interview process."
Technical Competence	A company using AVI is perceived as technically competent by the respondent	"I did feel that these orgs were up with the times and technology adept."
Interactivity	The respondent wished the AVI would be a more interactive experience	"I wish it could at least be moderated or have some connection to a live individual."
Transparency	The process of AVI should be more transparent to the applicants	"I would make sure that the participant is more informed by the company about the process."
AVI	The respondent prepared beforehand to improve AVI understanding	"I would suggest to a friend to test the technology before actually attending the interview, that way they can work out any kinks beforehand."
Appearance	The appearance and background that are presented in the video	"Think about the space in which you are presenting yourself to the company."
Technology Concerns	The AVI experience was filled with technology concerns	"I found the app I had to use strange and I didn't feel comfortable with the privacy boundaries around having a third party host the interview platform"
Lack of Human Presence	Without doing an AVI, the respondent would be concerned about a lack of human presence	"A lot of my interview strengths come from being able to ask the interviewer questions which I wouldn't be able to do in an AVI."
Personal Signaling	Without doing an AVI, the respondent would be concerned about the ability to signal his/her strengths	"I would be extremely nervous about completing an AVI. I am not a person who loves the camera and am not sure I come across. In person, I am very warm and attentive, but to be so devoid of a human interviewer would put me at a great disadvantage in interviewing for a job."
Negative Social	AVI would hurt the respondent's chances because of social implications	"I didn't get feedback from a person so I felt more anxious. I think the process was a negative one on my end."
Positive Social	AVI would help the respondent's chances	"It would help because there is less pressure when talking to someone online versus talking to someone in

	because of social implications	person. I get nervous when meeting someone important for the first time (in person), so getting interviewed in a comfortable spot with no face-to-face communication makes me act more like myself with less stress”

Table 1. High-Level Codes

Qualitative Findings

In pursuit of our research questions, the two main coders developed an organizing process model based on the dimensions of stages (1. preparation, 2. performance, and then 3. perception/evaluation) and factors (personal, social, technological, and appearance). The model was then filled out using evidence and examples from the data. The model was presented to the remaining researchers to add, delete, or modify based on their understanding of the codes in the data. After multiple iterations, we settled on a final process model (Figure 1).

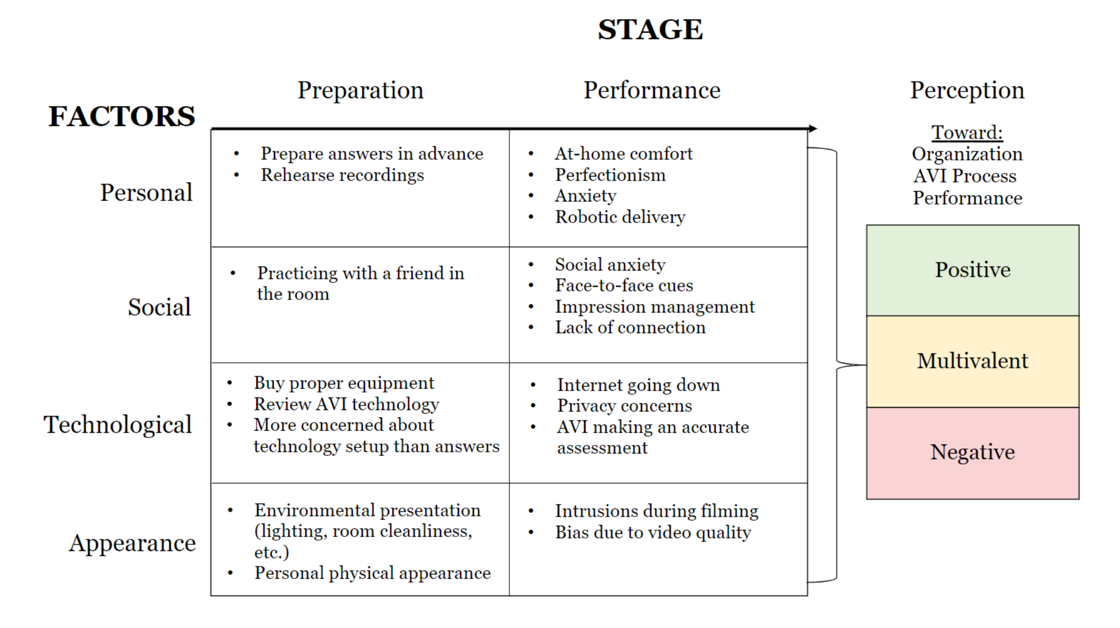

Figure 1. Preliminary Findings

Participating in an AVI typically was described in three stages: what was done to prepare for the AVI (preparation), the actual AVI (performance), and then the attitude toward (1) the organization utilizing AVI, (2) the AVI process, and (3) the user’s performance once the process was finished (perception). Perceptions varied across survey respondents, with some participants indicating that the AVI experience was more beneficial than face-to-face, others indicating that it was more detrimental than face-to-face, and some held both attitudes. Some of these perceptions were drawn from experience, and some were expectations for those that had not experienced AVIs before.

We identified four high-level factors that influenced perceptions at both the preparation and performance stages of the AVI process. These included personal factors such as anxiety or perfectionism, social factors such as the desire for human presence and feedback cues, technological factors such as an individual’s internet connection or the format of the AVI software, and appearance factors such as the video background and physical appearance. Personal factors relate to the ways that AVI affordances affect the mindset of an individual, such as anxiety or perfectionism. Because personal preferences of how interviews should be conducted varied widely, several positive and negative sub-factors can be seen here. The ability to rehearse may be a huge advantage to one, but the fear of a robotic delivery or the need to rerecord videos to perfection may pose a significant disadvantage to another. Social factors relate to how AVIs affect the expected social interactions that would come from interviews. Many lamented the lack of nuanced social cues and the ability to interact more naturally to build a connection and properly communicate. However, others noted how it could ease their social anxiety by decreasing the pressure to perform in front of others. Technological factors refer to any aspect of the AVI process where technology presents an opportunity or concern. The technology could be positive if someone was familiar with AVI software or otherwise tech adept. However, technology concerns often included worrying about internet connection quality, the privacy of recorded videos, and a lack of proper equipment. Appearance refers to the physical appearance of the individual and their environment. Many participants noted the opportunity to properly prepare the environment and appear their best while filming, with some expressing concerns that intrusions, lack of good filming hardware, or video quality could hurt how their performance is judged.

While we do not claim these factors to be exhaustive or mutually exclusive – they illustrate the breadth of factors that could be manipulated and have unique implications at both preparation and performance stages. For example, administrators of AVIs should consider what their processes do that affect a user's preparation, such as providing questions in advance or whether a respondent may be required to have access to high quality computer equipment. Likewise, there are performance stage considerations: what if the internet goes down? Does the AVI change which applicants perform better and how they are affected by social anxiety? Can recordings be discarded and reattempted? Considering these factors is important to ensure they do not result in unintended consequences.

Quantitative Analysis

After determining a set of defined themes in the data, we used content analysis (Hsieh and Shannon 2005; Morgan 1993) to quantify those themes inside the observed responses. Two coders were assigned to each question to ensure the validity of the coding. As indicated above, these coders re-coded all eight questions using the 44 codes derived from the initial open-coding stage. Ultimately, three of the codes were deemed not useful due to their low frequency. Inter-rater reliability scores were calculated for the remaining 41 codes. The average Cohen's kappa was .791. Subsequently, all discrepancies were discussed and resolved to create the final dataset of code counts.

Using these code counts, we explored the relationships between the quantitative coded responses and the characteristics and demographic traits of the respondents. All respondents answered Likert questions on both social anxiety and technological adeptness and then identified whether they identified as a member of an underprivileged minority group. Although space precludes a full report, we next summarize a few of the statistically significant effects we explored via correlations and independent samples t-tests.

Quantitative Findings

First, we found that those in the high social anxiety group more often reported that using an AVI (instead of face-to-face) would help their chances, possibly because of the pressure of face-to-face interviews. Whereas those in the low social anxiety group felt AVIs would disadvantage them because they wouldn't be able to rely on their social skills. Second, a much larger portion of those who identified as less tech-adept had only negative things to say about the AVI experience (on a scale of 1.00, the mean difference was 0.846, $p=0.001$). Similarly, those same respondents were worried that the AVI system (when compared to face-to-face) would suppress their ability to signal their strengths (mean diff=0.625, $p=0.041$). Finally, we can see that those who identify as a member of an underprivileged group are much more likely only to say good things about the AVI experience (mean diff=0.839, $p=0.004$). Similarly, those who identify as a member of an underprivileged group see no need to change anything about the AVI system (mean diff=0.767, $p=0.019$).

Proposition Development

Based on the findings from Study 1, we developed five propositions. First, our preliminary results indicate that AVIs are received more positively by individuals with higher levels of social anxiety. A recent meta-analysis corroborates this observation and demonstrated that anxiety (including state, trait, and interview-specific anxiety) could negatively impact interview performance (Powell et al. 2018). Our findings suggest that AVIs can help those with social anxiety perform better since it decreases the level of social presence and pressure. One respondent remarked: “[The AVI experience] was far less stressful, knowing that the other person wasn’t actually right there.” Accordingly, we propose the following:

> *P1: Individuals with higher levels of social anxiety will have a higher preference for an AVI format.*

A common trend that emerged from our open coding was that individuals who were more socially introverted preferred AVIs since this format allowed them to retry or prepare for the question in advance. Although our initial qualitative questions did not capture how much preparation the AVI afforded, the responses often contextualized the benefits of AVIs concerning preparation. For example, one respondent remarked, “[since] I could potentially redo the recordings I need to send in until I’d felt I had it right, taking away the ‘introvert penalty’ that I normally have during the interview.” Thus, there appears to be a moderating effect of preparation time. Therefore, we propose that:

> *P2: Preparation resources (e.g., time, information) will strengthen the positive relationship between social anxiety and the preference for an AVI format.*

While the absence of someone on “the other side of the table” was a positive factor for those with higher levels of social anxiety, this seemed to be a negative factor for those who relied on in-person cues and interactions. One person remarked, “In general, I have not enjoyed using this type of technology in the interview process, though I do believe it will become a permanent part of hiring for many companies in the future. I miss the opportunity to engage in conversation with my interviewer as well as gauge their reaction to my responses and possibly clarify or elaborate on a point. Also, for some reason, I was much more nervous creating the recordings than I have ever been while in a face-to-face interview.” Another commented, “I was shocked at how cold the experience is. It doesn’t leave you feeling warm and fuzzy.” With these observations in mind, we propose the following:

> *P3: When face-to-face interviews are not an option, individuals with lower levels of social anxiety will prefer AVI experiences with higher levels of social presence indicators in the AVI process.*

Computer self-efficacy (Compeau and Higgins 1995) is an essential component in computer-mediated communication. Since AVIs are still an emerging area within industry practice, there is a high level of variance in the quality of platforms. One participant noted, “Another thing that was an issue was how to navigate the actual company website for this AVI platform. Some companies had very bad interfaces, where I would never know when I’m actually recording my answer.” Understandably, a higher level of familiarity with technology is anticipated to translate into higher levels of AVI self-efficacy, which previous research has shown to be positively related to performance (Langer et al. 2016). Thus:

> *P4: Individuals with higher levels of computer self-efficacy will respond more positively to the requirement to interact with an AVI.*

Common tips on how to prepare for an AVI included “make sure the environment around you is presentable” or “warn others around you to reduce loud noise.” These comments highlight the importance of being able to control one’s environment. For example, those who have access to a well-lit, private office with high-speed internet access are likely to respond more positively to AVIs than those who have to go to great lengths to prepare their environment (e.g., travel to the library, wake up earlier than everyone else in the apartment). Thus:

> *P5: Individuals with higher levels of control over their personal surroundings will respond more positively to the requirement to interact with an AVI.*

Study 2

We conducted a follow-up study to explore the factors that influence the performance and perceptions of participants in an AVI situation. Additionally, we were interested to identify how AI is leveraged and perceived by individuals during the AVI process. Although the purpose of study 2 was not intended to validate the propositions put forth by study 1, our findings outlined in study 1 influenced the design of study 2.

Hypotheses

Theoretically, AI is used to improve the efficiency for the hiring firm as well as the performance of the interviewing individual. In an AVI context, AVI systems incorporate different methods of evaluation. The three evaluation types leveraged in our study were AI evaluation, human rater evaluation, and participant's perceived performance. There are different theoretical mechanisms by which an inclusion of AI in an AVI system would influence these three different performance indicators. For example, AI evaluation metrics can be improved through feedback to the interviewing individual. According to feedback intervention theory (Kluger and DeNisi 1996), when an individual is provided performance feedback on specific task details it enables learning of that task and an increase in performance on specific task details. Therefore, AI feedback in AVIs should improve an AI evaluation through direct training on the details of the algorithm (e.g., how many ums, pauses, etc.).

The overarching goal of an interview is to present oneself in a positive light to the employer. This "goal-directed activity of controlling information in order to influence the impressions formed by an audience" is known as impression management (Schlenker 2012, p. 492). AI feedback of an AVI can assist interviewees to manage the impression that they feel employers will see. AI evaluations can provide the interviewee with an initial, objective baseline that acts as a proxy of an employer. Similar to mechanisms in feedback intervention theory (Kluger and DeNisi 1996), this initial evaluation will motivate an individual to improve their messaging and impression. The difference between the two metrics, however, is that an interviewee will target specific algorithm details for AI evaluations and will target the general perceptions of employers for human evaluations.

Finally, an interviewee's perceived performance is a subjective evaluation by the individual. AI suggestions may improve the performance of the individual. Yet, regardless of improvement, any intervention undertaken, even if it does not actually change the evaluations of human evaluators, can influence a perceived improvement in an outcome by the individual – like a placebo effect (Harrington 1999). In our study's context, an interviewee receiving suggestions from an AI and feeling like those suggestions were implemented can act like a placebo effect. This may influence interviewees to claim that their performance improved. Our findings from study 1 show that those who felt they were able to prepare felt they performed better. Therefore, while interviewees may or may not have improved objective performance in their second response, additional resources from the AI to prepare for the question should increase someone's personal perceived performance.

We collect the AI evaluation, the human evaluator judgments, and the participants' own perceived performance. From each of these types of performance measurements, an improvement measurement can be calculated from the difference between the performance in the second interview and the performance in the first. Because each of the actionable items shown in the AI evaluation feedback is associated and trained based on interviews that are judged by the platform to be ideal, we expect the following for each of our improvement measurements:

> *H1a: Participants that receive AI evaluation feedback will have higher improvement in AI performance scores compared to those that receive no AI evaluation feedback.*
>
> *H1b: Participants that receive AI evaluation feedback will have higher improvement in human evaluator scores compared to those that receive no AI evaluation feedback.*
>
> *H1c: Participants that receive AI evaluation feedback will have higher improvement in participant perceived performance scores compared to those that receive no AI evaluation feedback.*

Additionally, we evaluated the improvement from interview 1 to interview 2 moderated by a variety of demographic, environmental, and personality constructs that are traditionally viewed as factors that lead to a disadvantaging bias. First, we evaluated individuals' general perceived level of personal discrimination. This evaluated aspects such as race, gender, or other personal characteristics that lead to everyday discrimination. Second, we focused on perceived discrimination that is created by using AVIs. For example, this involves people that feel discriminated against based on their technological competence, poor equipment access, or any other reason that disadvantages an individual for interviewing with an AVI instead of face-to-face. Third, our initial study highlighted from interview responses how AVIs could be an ableist system, one which disadvantages people with disabilities. Therefore, we evaluate the disability level of the individual. Finally, since verbal communication is critical in AVIs, we evaluated the English fluency of individuals with the assumption that those who don't speak English as well or have heavy accents may be more disadvantaged than those who don't. The argument is that those who may do worse on initial evaluations will work harder and utilize any resource available to them to improve more than those who do not feel disadvantaged.

H2 a-d: The degree of improvement (on human raters, AI evaluation, and self-reported perceived performance) from AI evaluation feedback will be higher for participants who report (a) higher levels of personal discrimination, (b) discrimination from AVI tools, (c) higher levels of disability, and (d) lower levels of fluency in English.

Instrument and Participants

In this study, we recruited *295* participants from Prolific, the majority of whom (89.3%) were actively seeking a job at the time of the study and were thus focal to our target population. Participants underwent an AVI using the popular AVI software tool, Big Interview. Each participant was asked two questions that are typical of behavioral-type interviews:

1. What skills or characteristics do you bring to the workplace?
2. How do you handle stress and pressure?

Participants were randomly split into two groups where one group received, as a treatment, AI feedback for their response to the first question (before answering the second question) while the control group received no feedback. After the completion of both questions, respondents answered Likert-style questions regarding their disability[1], perceived discrimination (Williams et al. 1997), distractions (self-made), computer self-efficacy (Compeau and Higgins 1995), social presence (Toader et al. 2019), social anxiety (Liebowitz 1987), and personality[2]—all of which are measures with known historical bias that may influence the outcome of both AI and human evaluations (Bendick and Nunes 2012; Srinivasan and Chander 2021). Except for distractions, all measures were used verbatim or adapted from previous scales to apply to an AVI context. To encourage realistic performances during the interviews, the participants were told they would be rewarded with a gift card if they were judged by evaluators as top performers.

The AI feedback generated by Big Interview scored individuals on their pace of speech, "um" counter, vocabulary, filler words, power words, pause counter, eye contact, negative tone, response length, authenticity score, speech volume, and lighting. Each category is rated green (best), yellow, or red (worst). An overall performance badge is provided based on the scores above—gold, silver, or bronze. Interviewees in the condition that received feedback were required to record, in the Qualtrics survey, the insights that the AI provided them to ensure that the feedback was received and processed by the interviewee.

After dropping incomplete responses (a common problem with platforms like mTurk and Prolific), 171 complete responses remained that were then used in the analysis. The majority of the participants were female (60.0%). We had respondents from 13 different countries, most of whom were from the United States (72.7%) and the United Kingdom (13.0%). About half of the participants reported as white (53.0%), with smaller groups reporting as African American (17.1%), Mixed (14.8%), or South Asian (14.8%). For most participants (80.5%), English is their first language.

Data Analysis

In addition to the AI-generated feedback, four members of the research team evaluated the interviews on the following dimensions using a 5-point Likert scale:

- The interviewee conveyed an understanding of the intent of the interview question
- The interviewee responded effectively to the interview question
- The interviewee communicated his/her thoughts clearly
- The interviewee kept the response focused and avoided tangents

[1] https://www.census.gov/topics/health/disability/guidance/data-collection-acs.html

[2] https://www.idrlabs.com/short-big-five/test.php

- The interviewee convinced the potential reviewers to hire the interviewee

These are the same questions the interviewee rated themselves on after each interview. To ensure consistent evaluations across reviewers, the four reviewers first evaluated the same sample of 14 interviews and then compared ratings to discuss discrepancies and create a rubric ensuring minimal unexplainable variability among scores.

All hypotheses were initially tested using OLS regression, with additional relationships being explored through correlations. Predictors (theorized and controlled for) included a measure of perceived discrimination, perceived environmental control, disability, perceived discrimination, perceived social presence, social anxiety, gender, age, ethnicity, English proficiency, non-English accent prevalence, extraversion, agreeableness, conscientiousness, neuroticism, openness, computer self-efficacy, level of distraction during recording, and the length of the video. Verbatim scales are not reported in this manuscript due to space considerations but were adapted from established literature. Where averages were created to summarize latent factors (e.g., for social anxiety), all Cronbach's alphas were above the recommended target of 0.700. The resulting averaged scores all exhibited expected normal distributions. Interaction effects of the treatment and potential disadvantaged status were also examined. Variance inflation factors were examined in models before introducing interaction effects, and no concerns for multicollinearity were present (all VIFs < 3). For the AI performance scores (Model 1), the overall AI score was used, which ranged from 1 (worst) to 3 (best). For the human evaluator scores (Model 2), the "decision to hire" judgment was used, and the participant self-evaluation was used for the final model (Model 3). Regression results are given in Table 2, as an unstandardized regression weight (beta) and standard error (in parentheses).

	(1)	(2)	(3)
	Δ AI Score	Δ Human Evaluator	Δ Participant Perception
AI Feedback	1.532 (2.60)	-3.484 (2.59)	-1.045 (2.41)
Discrimination	0.151 (0.20)	-0.042 (0.22)	-0.162 (0.17)
AI Feedback x Discrimination	-0.290 (0.29)	0.024 (0.29)	0.811** (0.28)
Environmental Control	-0.031 (0.17)	0.008 (0.14)	-0.129 (0.15)
AI Feedback x AVG_Environmental_Control	0.091 (0.23)	0.048 (0.19)	0.163 (0.20)
Disability	-0.248 (0.22)	0.145 (0.28)	-0.101 (0.28)
AI Feedback x Disability	0.177 (0.33)	-0.427 (0.35)	0.342 (0.39)
Perceived Discrimination	-0.042 (0.19)	-0.086 (0.21)	0.124 (0.22)
AI Feedback x Perceived Discrimination	0.148 (0.28)	0.409 (0.26)	0.065 (0.31)
Social Presence	0.066 (0.12)	-0.150 (0.14)	0.065 (0.10)
AI Feedback x Social Presence	0.234 (0.18)	0.041 (0.20)	-0.299† (0.16)
Social Anxiety Fear	-0.032 (0.29)	-0.080 (0.28)	0.008 (0.24)
AI Feedback x Social Anxiety Fear	-0.090 (0.33)	0.314 (0.30)	0.064 (0.32)
English Proficiency	0.101 (0.18)	0.015 (0.20)	0.188 (0.16)
AI Feedback x English Proficiency	-0.085 (0.21)	0.211 (0.23)	-0.296 (0.22)
Accent	-0.034 (0.06)	0.039 (0.09)	0.001 (0.05)
AI Feedback x Accent	0.031 (0.08)	0.036 (0.10)	-0.030 (0.07)
Q1 Video Length Too Long	0.644* (0.26)	-0.106 (0.29)	-0.494† (0.28)
Q1 Video Length Too Short	0.541* (0.23)	0.294 (0.23)	-0.094 (0.24)
AI Feedback x Too Long	-0.134 (0.57)	0.425 (0.60)	0.507 (0.49)
AI Feedback x Too Short	-0.653,† (0.38)	-0.195 (0.32)	-0.024 (0.35)

Ethnicity (Black)	0.289 (0.33)	-0.000 (0.30)	-0.273 (0.37)
Ethnicity (Asian)	0.435 (0.27)	-0.232 (0.29)	0.067 (0.34)
Ethnicity (Hispanic/Latino)	0.299 (0.33)	-0.817† (0.43)	-0.322 (0.26)
Ethnicity (Other)	0.106 (0.46)	0.616† (0.37)	-0.834 (0.52)
AI Feedback x Black	-0.190 (0.48)	-0.373 (0.42)	0.344 (0.53)
AI Feedback x Asian	-0.319 (0.41)	0.075 (0.40)	0.159 (0.45)
AI Feedback x Hispanic/Latino	-0.792 (0.66)	0.794 (0.75)	0.413 (0.41)
AI Feedback x Other	-0.362 (0.65)	-1.225* (0.56)	0.633 (0.93)
Gender (Female)	0.216 (0.18)	-0.026 (0.19)	-0.013 (0.18)
Gender (Trans)	-0.161 (0.59)	-0.298 (0.53)	-0.545 (0.57)
Gender (Other)	-0.583 (0.86)	1.286† (0.73)	0.712 (0.98)
Extraversion	-0.084 (0.05)	0.024 (0.04)	-0.034 (0.04)
Agreeableness	0.007 (0.05)	-0.006 (0.05)	0.009 (0.05)
Conscientiousness	0.002 (0.04)	-0.046 (0.05)	-0.055 (0.06)
Neuroticism	0.036 (0.05)	-0.050 (0.05)	-0.067 (0.06)
Openness	0.074† (0.04)	-0.143** (0.05)	0.056 (0.05)
Age	0.000 (0.01)	-0.001 (0.01)	-0.014† (0.01)
Constant	-2.538 (2.23)	2.151 (2.22)	0.415 (1.75)
Observations	147	161	160
Adjusted R^2	-0.056	0.028	-0.012
Statistical Power	0.316	0.205	0.139

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2. OLS Results for improvement on AI Score, Human Evaluator, and Participant Perception

Study 2 Results

Surprisingly, the treatment of receiving AI feedback did not indicate any significant direct improvement across any of the three measurement types. However, there is an interesting interaction effect between feedback and discrimination. Discrimination measured the degree to which participants felt they had a disadvantage in using the AVI technology to succeed in answering their questions. Given that this question was asked after receiving AI feedback and answering both questions, it appears that those that received feedback and felt less discriminated in their ability to take advantage of the technology rated themselves as performing better by .81 points ($p <.01$) for every point increase in disadvantage they felt. However, this only translated to an increase in perceived performance, but not judged or AI-evaluated performance.

Hypothesis	Evidence	Interpretation
H1a. AI Feedback → ΔAI Perf	β = 1.532	Correct direction, but not significant
H1b. AI Feedback → ΔHuman Perf	β = -3.484	Not supported
H1c. AI Feedback → ΔSelf Perf	β = -1.045	Not supported
H2a. AI x Disc → ΔPerf	β = -0.290 (ΔAI) β = 0. 024 (ΔHuman) β = 0. 811** (ΔSelf)	Supported for self-evaluation improvement
H2b. AI x AVI Disc → ΔPerf	β = 0.148 (ΔAI)	Not significant

	β = 0.409 (ΔHuman) β = 0.065 (ΔSelf)	
H2c. AI x Disability → ΔPerf	β = 0.177 (ΔAI) β = -0.427 (ΔHuman) β = 0.342 (ΔSelf)	Not supported
H2d. AI x English → ΔPerf	β = -0.085 (ΔAI) β = 0.211 (ΔHuman) β = -0.296 (ΔSelf)	Not supported

Table 3. Results of the Hypotheses with Interpretation

Alternative Explanations of the Hypotheses

Upon reviewing our results, we considered other potential factors that could have affected our hypotheses. For example, when the AI is not custom fit to the job posting but is standard out-of-the-box, there is a potential negative effect of AI feedback on human evaluations. This is because participants might focus too much on interview elements that are not important for job success, but that are emphasized by the AI. Big Interview, the AVI software we used, is a linguistic parser and does not evaluate semantic content, meaning that the AI does not grade the subject based on what they say but instead how they say it. The AI we used focuses on elements such as the rate at which the interviewee speaks, the number of "ums" recorded, filler words, eye contract, length of answer, pauses taken, and the amount of power words used. However, it ignores context. To this AI, the evaluation criteria for a blue-collar construction job vs. a white-collar knowledge worker would be the same. While these indicators are important for portraying a professional image, they are not the only (or even most important) things that matter when interviewing for a position in an organization. Humans have a tendency to focus effort on what is being measured (Kluger and DeNisi 1996). When we offer feedback to a subject that informs them that they are talking too slowly and have a significant amount of filler words, the subject is likely going to focus more on fixing those two issues which may result in less attention to the actual question they are attempting to answer.

Another possible explanation could be undermining signals from the AI. For example, AI feedback could make participants more aware of shortcomings. This awareness could increase anxiety in the interviewee leading to a decrease performance in a second interview or rating a second interview worse than they would have originally. These alternative explanations highlight the dual influence that AI feedback in an AVI context can provide. Such dual effects may muddy the outcomes. Lastly, it may be difficult for someone to change their communication habits (such as speed, volume, and filler words) within such a short span as was provided during this quasi-experiment.

Of course, we can also attribute a lack of findings to a lack of statistical power. None of our models achieved adequate statistical power (0.80), and therefore, while significant effects may exist among the variables analyzed, our model and sample might simply be inadequate to detect such effects.

Additional Exploratory Findings

Our modest findings above led us to explore the data beyond our hypotheses. We first produced a correlation matrix with all variables to discover those that shared the most variance with our performance variables. We then conducted a simplified path analysis in AMOS informed by the correlation matrix. Our path model included two dependent variables: AI score and human-rated performance (both after the second interview). The predictors included: social anxiety avoidance, social anxiety fear, extraversion, disability, and age. We tested our model with all data, as splitting the data into treatment and control groups resulted in too small sample sizes to observe significant effects. A post-hoc power analysis showed that we had sufficient power to detect significant effects for human performance evaluations (0.914), but not for AI performance evaluations (0.631). Our path analysis revealed five interesting findings.

First, social anxiety fear had a positive effect on AI score (0.306, $p=0.05$), which was contrary to what we anticipated. One explanation for this effect is that fear may serve as a motivating factor for the individual to intensify effort during the interview. This is meaningful because AVIs may not only be a format that individuals with higher social anxiety prefer (P1), AVIs may be a format in which their social anxiety fear seems to prompt them to perform better. Social anxiety and fear did not affect human-evaluated performance.

Second, although social anxiety fear had a positive effect on AI score, social anxiety avoidance had a negative effect on AI score (-0.367, p=0.03). Where fear may have served as a motivating factor for those with high social anxiety to perform well in the interview, those with high avoidance may have sought to simply finish the interview so it was no longer confronting them. Social anxiety avoidance also harmed the interviewee's human-evaluated performance score (-0.314, p=0.08). These findings imply that an AVI-formatted interview may not be the best interview medium for those with high social anxiety avoidance. A similar study would need to be conducted with other interview mediums to conclude which one is most advantageous to those with high social anxiety avoidance, recognizing that the most advantageous one may not be the preferred one.

Third, while extraversion had a predictably positive effect on human-evaluated performance (0.049, p=0.09), it did not affect AI score (-0.018, p=0.49). This is not surprising as recruiters can often become attracted to individuals with extroverted tendencies (e.g., charisma). Charisma may win over humans, but it does not seem to influence AI evaluations. This may level the field for those who are not naturally charismatic.

Fourth, age did not affect AI score, but age harmed human-evaluated performance (-0.01, p=0.04). Older respondents were judged more critically by human judges, but the AI also overlooked this individual factor. Thus, AI feedback may help to avoid age discrimination in hiring.

Fifth, through the medium of the AVI, disability did not seem to factor into performance evaluations. This result is exploratory and as such we are cautious to make definitive statements or claims. However, if AVIs can help to prevent discrimination based on disability, that would be a step in the right direction.

Discussion & Implications

We completed two studies and a post-hoc exploratory analysis that help to understand individual perceptions of AVI systems as well as how personal characteristics influence the preference for and success of using an AVI system. Additionally, while prior literature has studied how AI tools in AVI systems can be used as an evaluation metric (Suen et al. 2019a; Suen et al. 2019b), we adopt a different perspective by evaluating how AI tools can be leveraged by the interviewee to enhance their performance and their perception of the AVI process. Particularly, to address key societal issues of fairness and equity, we evaluate how disadvantaged groups perform in an AVI process and how they react to the inclusion of an AI tool in that process. The findings outlined above have potentially important implications for people who design AVI platforms, companies that leverage AVI platforms to conduct interviews, and individuals who use AI in AVIs to improve their performance.

First, we demonstrate that individual characteristics and situations are important considerations for AVI design that improve both individual AVI performance and perceptions of the AVI experience. Some key insights are that social anxiety has bifurcating effects on AVI performance. Whereas we supposed those with social anxiety would generally prefer AVI (as opposed to in-person interviews), we found that those with social anxiety fear perform better than those without this fear. However, those with social anxiety avoidance perform worse. This finding adds to prior work that is trying to improve fairness in hiring with new technologies (Ochmann and Laumer 2019). Our study demonstrates that a single method or system will ultimately have difficulties of achieving fairness because of the unique and diverse situations and personalities of jobseekers. Thus, we recommend adopting and implementing hybrid options to increase fairness and opportunities.

Second, we found that extraversion is ignored by AI evaluations, at least in the context of our study. This result highlights that the different evaluation mediums examine different factors when making decisions. While human evaluators may see extraversion as a positive signal, perhaps due to charisma, the AI does not take this into account. Thus, one cannot charm their way into a job when the AI is doing the evaluation. With an increase in AI-evaluated software tools (Suen et al. 2019b), this may balance the scales that primarily lean toward extroverted individuals (Caldwell and Burger 1998). While this may be one way to increase fairness in the hiring process (Suen et al. 2019a), it questions if AI is targeting the correct factors for a job position. Organizations should consider this if a certain job requires charismatic qualities.

Third, our findings suggest that the AI feedback feature appears to be failing in its intended role to help candidates improve their interview skills. The lack of time given between questions and the limited number of questions may be limitations that do not allow individuals to improve noticeably. However, the process that we had our interviewees undertake likely mirrors the time that would be allotted for an actual interview. The AI feedback feature's effect on interview performance may need to be used over time with several interviews to show a significant impact. AI feedback developers may also consider the features to add/remove from the AI score to maximize performance. Future work should seek to identify which features are most successful in improving the hiring success of an individual in addition to the quality of response.

Fourth, our findings suggest that interviewers should not rely solely on AI scores generated from AVI platforms to make hiring decisions, as these decisions did not always align with human feedback. However, in some cases, the AI feedback ignored conventional biases while humans did not. Thus, AI feedback should help to inform human evaluations but should not entirely replace it (Haesevoets et al. 2021; van den Broek et al. 2021). We recommend that companies be aware of the limitations specific to AI feedback or hiring software they use and subsequently leverage a hybrid algorithm in the hiring process that utilizes both AI feedback and human judgement to make hiring decisions.

Conclusion

Through our multi-study, multi-method exploration, we have uncovered a variety of insights regarding AVIs and the utilization of AI feedback within the AVI process. As automation continues to impact business processes and procedures, a more complete and nuanced understanding of AVIs is paramount. Our study may guide future AVI researchers toward interesting research questions and important societal implications. In contrast to many other business processes, AVIs are often the first exposure that potential hires have to an organization. These first impressions can have long-term effects on both hired and non-hired interviewees. Additionally, past research has routinely demonstrated (and condemned) racial, gender, religious, and different types of bias in the hiring process (Bendick and Nunes 2012). The design of the AVI can systematically benefit or harm certain groups (e.g., lower versus higher levels of social anxiety), and a deeper understanding of AVI design can prevent unwarranted bias. Lastly, by investigating the interviewee perceptions of AVIs, we contribute to "last mile" academic research that seeks to "[use] scientific knowledge and methods to address important unsolved classes of problems for real people with real stakes in the outcomes" (Nunamaker et al. 2015, p. 11). Our research agenda is closely tied to the lived experiences of individuals and organizations and is prepared to make practical recommendations to enhance the positive and decrease the negative components associated with AVIs.

References

Araujo, T., Helberger, N., Kruikemeier, S., and de Vreese, C. H. 2020. "In Ai We Trust? Perceptions About Automated Decision-Making by Artificial Intelligence," *Ai & Society* (35:3), pp. 611-623.

Barocas, S., and Selbst, A. D. 2016. "Big Data's Disparate Impact," *California Law Review* (104:3), pp. 671-732.

Basch, J. M., Melchers, K. G., Kurz, A., Krieger, M., and Miller, L. 2021. "It Takes More Than a Good Camera: Which Factors Contribute to Differences between Face-to-Face Interviews and Videoconference Interviews Regarding Performance Ratings and Interviewee Perceptions?," *J Bus Psychol* (36:5), pp. 921-940.

Baudot, J. 2006. "Social Justice in an Open World the Role of the United Nations." New York: United Nations.

Bendick, M., and Nunes, A. P. 2012. "Developing the Research Basis for Controlling Bias in Hiring," *Journal of Social Issues* (68:2), pp. 238-262.

Blacksmith, N., Willford, J., and Behrend, T. 2016. "Technology in the Employment Interview: A Meta-Analysis and Future Research Agenda," *Personnel Assessment and Decisions* (2:1), pp. 12-20.

Caldwell, D. F., and Burger, J. M. 1998. "Personality Characteristics of Job Applicants and Success in Screening Interviews," *Personnel Psychology, Inc.* (51:1), pp. 119-136.

Compeau, D., and Higgins, C. 1995. "Computer Self-Efficacy: Development of a Measure and Initial Test," *MIS Quarterly* (19:2), pp. 189-211.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., and Huq, A. 2017. "Algorithmic Decision Making and the Cost of Fairness," in: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 797-806.

Dastin, J. 2018. "Amazon Scraps Secret Ai Recruiting Tool That Showed Bias against Women " in *Ethics of Data and Analytics: Concepts and Cases,* K. Martin (ed.). New York: Auerbach Publications, pp. 296-299.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., and Tamo-Larrieux, A. 2020. "Towards Transparency by Design for Artificial Intelligence," *Sci Eng Ethics* (26:6), pp. 3333-3361.

Gillespie, T. 2014. "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality, and Society,* T. Gillespie, P.J. Boczkowski and K.A. Foot (eds.). Cambridge, MA: the MIT Press, pp. 167-194.

Glaser, B. G., and Strauss, A. L. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research Observations*. Aldine.

Haesevoets, T., De Cremer, D., Dierckx, K., and Van Hiel, A. 2021. "Human-Machine Collaboration in Managerial Decision Making," *Computers in Human Behavior* (119).

Harrington, A. 1999. *The Placebo Effect: An Interdisciplinary Exploration* Harvard University Press.

Hiemstra, A. M. F., Oostrom, J. K., Derous, E., Serlie, A. W., and Born, M. P. 2019. "Applicant Perceptions of Initial Job Candidate Screening with Asynchronous Job Interviews," *Journal of Personnel Psychology* (18:3), pp. 138-147.

HireVue. 2021. "Hirevue Completes Its 20 Millionth Video Interview."

Hsieh, H.-F., and Shannon, S. E. 2005. "Three Approaches to Qualitative Content Analysis," *Qual Health Res* (15:9), pp. 1277-1288.

Jarrahi, M. H. 2018. "Artificial Intelligence and the Future of Work: Human-Ai Symbiosis in Organizational Decision Making," *Business Horizons* (61:4), pp. 577-586.

Johnson, T. K. M., and Gray, R. 2019. "Use Your Words: Text Analysis in Selection and Assessment," *34th Annual Conference of the Society for Industrial and Organizational Psychology*, National Harbor, MD.

Kluger, A. N., and DeNisi, A. 1996. "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta Analysis, and a Preliminary Feedback Intervention Theory," *Psychological Bulletin* (119:2), pp. 254-284.

Kononenko, I. 2001. "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective," *Artificial Intelligence in Medicine* (23:1), pp. 89-109.

Langer, M., König, C. J., and Fitili, A. 2018. "Information as a Double-Edged Sword: The Role of Computer Experience and Information on Applicant Reactions Towards Novel Technologies for Personnel Selection," *Computers in Human Behavior* (81), pp. 19-30.

Langer, M., Konig, C. J., Gebhard, P., and Andre, E. 2016. "Dear Computer, Teach Me Manners: Testing Virtual Employment Interview Training," *International Journal of Selection and Assessment* (24:4), pp. 312-323.

Liebowitz, M. R. 1987. "Social Phobia," *Modern Problems in Pharmacopsychiatry* (24), pp. 141-173.

Lukacik, E.-R., Bourdage, J. S., and Roulin, N. 2020. "Into the Void: A Conceptual Model and Research Agenda for the Design and Use of Asynchronous Video Interviews," *Human Resource Management Review* (32:1).

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., and Ahmed, S. M. 2017. "Applicant Perspectives During Selection: A Review Addressing “So What?,” “What’s New?,” and “Where to Next?”," *Journal of Management* (43:6), pp. 1693-1725.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. 2021. "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys* (54:6), pp. 1-35.

Morgan, D. L. 1993. "Qualitative Content Analysis: A Guide to Paths Not Taken," *Qualitative Health Research* (3:1), pp. 112-121.

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., and Schwabe, G. 2015. "The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research," *Journal of Management Information Systems* (32:3), pp. 10-47.

Ochmann, J., and Laumer, S. 2019. "Fairness as a Determinant of Ai Adoption in Recruiting: An Interview-Based Study " *DIGIT Proceedings*, Munich, Germany.

Paulus, J. K., and Kent, D. M. 2020. "Predictably Unequal: Understanding and Addressing Concerns That Algorithmic Clinical Prediction May Increase Health Disparities," *NPJ Digit Med* (3), p. 99.

Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research," *Journal of Experimental Social Psychology* (70), pp. 153-163.

Pentland, S. J., Twyman, N. W., Burgoon, J. K., Nunamaker, J. F., and Diller, C. B. R. 2018. "A Video-Based Screening System for Automated Risk Assessment Using Nuanced Facial Features," *Journal of Management Information Systems* (34:4), pp. 970-993.

Powell, D. M., Stanley, D. J., and Brown, K. N. 2018. "Meta-Analysis of the Relation between Interview Anxiety and Interview Performance," *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement* (50:4), pp. 195-207.

Proudfoot, J. G., Jenkins, J. L., Burgoon, J. K., and Nunamaker, J. F. 2016. "More Than Meets the Eye: How Oculometric Behaviors Evolve over the Course of Automated Deception Detection Interactions," *Journal of Management Information Systems* (33:2), pp. 332-360.

Rubinstein, P. 2020. "Asynchronous Video Interviews: The Tools You Need to Succeed."

Schlenker, B. R. 2012. "Self-Presentation," in *Handbook of Self and Identity,* M.R. Leary and J.P. Tangney (eds.). The Guilford Press, pp. 542-570.

Srinivasan, R., and Chander, A. 2021. "Biases in Ai Systems," *Communications of the ACM* (64:8), pp. 44-49.

Stevens, C. K. 1998. "Antecedents of Interview Interactions, Interviewers' Ratings, and Applicants' Reactions," *Personnel Psychology, Inc.* (51:1), pp. 55-85.

Suen, H.-Y., Chen, M. Y.-C., and Lu, S.-H. 2019a. "Does the Use of Synchrony and Artificial Intelligence in Video Interviews Affect Interview Ratings and Applicant Attitudes?," *Computers in Human Behavior* (98), pp. 93-101.

Suen, H.-Y., Hung, K.-E., and Lin, C.-L. 2019b. "Tensorflow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews," *IEEE Access* (7), pp. 61018-61023.

Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., and Rădulescu, A. T. 2019. "The Effect of Social Presence and Chatbot Errors on Trust," *Sustainability* (12:1).

Twyman, N. W., Proudfoot, J. G., Schuetzler, R. M., Elkins, A. C., and Derrick, D. C. 2016. "Robustness of Multiple Indicators in Automated Screening Systems for Deception Detection," *Journal of Management Information Systems* (32:4), pp. 215-245.

van den Broek, E., Sergeeva, A., and Huysman, M. 2019. "Hiring Algorithms: An Ethnography of Fairness in Practice," *International Conference of Information Systems*, Munich, Germany.

van den Broek, E., Sergeeva, A., and Huysman Vrije, M. 2021. "When the Machine Meets the Expert: An Ethnography of Developing Ai for Hiring," *MIS Quarterly* (45:3), pp. 1557-1580.

Wang, Y. 2021. "When Artificial Intelligence Meets Educational Leaders' Data-Informed Decision-Making: A Cautionary Tale," *Studies in Educational Evaluation* (69).

Wilhelmy, A., Kleinmann, M., Konig, C. J., Melchers, K. G., and Truxillo, D. M. 2016. "How and Why Do Interviewers Try to Make Impressions on Applicants? A Qualitative Study," *Journal of Applied Psychology* (101:3), pp. 313-332.

Williams, D. R., Yu, Y., and Jackson, J. S. 1997. "Racial Differences in Physical and Mental Health: Socio-Economic Status, Stress and Discrimination," *Journal of Health Psychology* (2:3), pp. 335-351.

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., and Anderson, N. R. 2019. "Personnel Selection in the Digital Age: A Review of Validity and Applicant Reactions, and Future Research Challenges," *European Journal of Work and Organizational Psychology* (29:1), pp. 64-77.

Zibarras, L., Patterson, F., Holmes, J., Flaxman, C., and Kubacki, A. 2018. "An Exploration of Applicant Perceptions of Asynchronous Video Mmis in Medical Selection," *MedEdPublish* (7).