

Preregistration

The Cognitive, Behavioral, and Strategic Impacts of Voice AI: Evidence from Job Interviews

Brian Jabarian Luca Henkel

March 7, 2025

1 Project Description

Recent technological advances in generative artificial intelligence (AI) enable large language models to hold natural conversations on complex topics. In this project, we want to study the consequences of replacing human-to-human conversations with human-AI conversations in the context of job interviews. For this purpose, we partner with a hiring company to conduct a natural field experiment. In the experiment, we vary whether a human or AI-powered voice agent conducts the job interview with applicants. We are interested in two main questions. First, how is the effectiveness of the hiring process affected by replacing human recruiters with AI? Second, how do humans respond to facing a conversation with an AI?

2 Experimental Design

2.1 Setting

We partner with a global recruitment solutions provider (hereafter referred to as “the company”). Large US-based firms outsource (parts) of their job hiring process to the company. The jobs advertised online by the company are entry-level jobs, such as a customer service representative position. Applicants submit their materials through the company’s portal. Then, the company’s recruiters conduct an initial screening. If the recruiter judges an applicant suitable, they are invited by text message for a phone interview. There are two modes of interviews depending on the job and recruitment center: either the interview is conducted remotely, or in one of the company’s recruitment centers. Complete interviews typically last between 10 and 30 minutes and serve to assess the candidate’s background, experience, and fit for the job. After the interview, a recruiter reviews both the transcript and audio recording of the interview and decides whether to extend a job offer to the applicant or not. If an applicant receives and accepts an offer, there are additional background and medical checks. Once passed, the applicant receives job training and starts their job with the respective firm.

2.2 Treatments

Our treatment variation concerns the interview stage, specifically who conducts the interview. Applicants who successfully pass the initial screening are subsequently randomized into one of three experimental conditions: *Human interviewer*, *AI interviewer*, and *Choice of interviewer*.

Human interviewer. In the *Human interviewer* condition, applicants are interviewed by a human recruiter for their job interview. Human recruiters in the organization follow a structured interview script designed to ensure a standardized interview process. The script includes essential questions tailored to the candidate’s background, such as gaps in employment or transitions between jobs, while also allowing flexibility for recruiters to adapt their approach as needed.

AI interviewer. In the *AI interviewer* condition, applicants are instead interviewed by a voice AI agent based on a proprietary large language model. The voice AI agent follows a scripted conversation that mimics the same questions and follow-up strategies used by human interviewers.

Choice of interviewer. Lastly, in the *Choice of interviewer* condition, applicants can choose whether a human or a voice AI agent will interview them.

Importantly, in all experimental conditions, a human interviewer reviews the conversation after the interview and makes a hiring decision, irrespective of whether a human or AI conducted the interview.

2.3 Sample Size

Our unit of observation will be individuals applying to the company. In a typical (pre-intervention) month, the company receives applications from around 9,000 individuals. About 20% of those complete a job interview. Of those who complete a job interview, around 20% receive a job offer. We plan to run the experiment for three months starting March 7, 2025. That is, we include in the experiment all applicants who applied between March 7 and May 7, 2025. The expected sample size is thus around 27,000 applications, 5,400 interviews, and 1,080 job offers. The actual sample size depends on several external factors such as demand from the firms commissioning the hiring company. Therefore, the sample size is not directly in the researchers’ control, meaning the actual sample size may be (substantially) higher or lower than expected. In case of unexpectedly low application numbers, we may extend the running time of the experiment.

The randomization is implemented by the company using a computer code. The code was developed with help from the researchers. The randomization weights are set by the company and outside of the researcher’s control. We expect all three conditions to include more than 10% of applicants in a given month.

Existing data. We received and analyzed pilot data from the company. In particular, we received preliminary data from applicants applying prior to March 2025 that spans most variables described in Section 3.1.2. We analyzed the pre-experimental data to ensure sufficient data availability and quality for the main experiment.

Stopping rule. We will terminate data collection if the company forbids access to the data. We will also not realize the study if the company does not deliver the data in the agreed-upon form described in this document.

2.4 Criteria for Excluding Observations

While we will not exclude any applicants, for our analysis described in Section 3.2 we focus on applicants who are invited for interviews. Furthermore, for the *Choice of interviewer* condition, we restrict our analysis to applicants who actually make a choice between the AI and human recruiter.

3 Hypotheses and Analysis Plan

3.1 Variables

3.1.1 Pre-interview applicant variables

- Gender
- Engagement score
- Whether the applicant was invited for an interview

Some of the variables will only be available for a subset of applicants. Wherever applicable, we will use the pre-interview variables for heterogeneity analysis.

3.1.2 Key outcome variables

Interview outcomes

- Whether the interview was successfully completed (0/1)
- Whether the applicant receives a job offer (0/1)
- Interviewer score (1/2/3)
- Interviewer comment (open-ended text)
- Whether the applicant accepts an offer conditional on receiving one (0/1)

Length of recruitment process

- Time from initial application until the interview takes place (in days).
- Time from initial application until a final decision has been made (in days).

Job variables

- Whether the applicant started their job at the respective firm (0/1)
- Retention rate: whether the applicant still works at the firm after starting their job.

3.1.3 Interview transcript variables

The company may share with us the transcripts of the interviews in text form. If shared, we will analyze the transcripts using text analysis. In particular, we will focus on the following dimensions.

- Vocabulary richness
- Filler words & hedging
- Sentiment polarity (positive, neutral, negative)
- Turn-taking behavior
- Response length
- Question-Answer Alignment: i) whether the candidate provides an answer aligned with the question asked by the recruiter; ii) whether the recruiter understands or not the answer provided by the applicant
- Redundancy of information shared
- Conversation Frictions (Interruptions & Overlaps)

3.1.4 Interview audio file variables

The company may share the audio files of the interviews with us. If shared, we will focus on the following dimensions when analyzing the audio files.

- Speech Rate
- Pause Frequency & Duration
- Turn-Taking Balance Smooth
- Speaking Time Proportion
- Speech Volume (dB Level)
- Pitch Variability
- Voice Tremor Detection
- Time to First Response (Post-Question Latency)

3.1.5 Survey variables

After completing their interview, applicants are invited to complete a short survey. In the survey, we elicit the following variables. Until otherwise noted, the variables are measured using Likert scales.

- Recommend company to other (0-10)

Assessment of recruiter by applicants

- Recruiter knowledge of company
- Recruiter knowledge of the role
- Relevance of questions asked for role
- Degree to which applicant time was valued

We create a recruiter assessment index variable by averaging across all four individual recruiter assessment items.

Interview experience

- Naturalness of the interview
- Comfortableness during the interview
- Stressfulness of the interview

We create an interview experience index variable by averaging across all four individual interview experience items.

Interview fluency

- Follow-up flow
- Frequency of feedback

Interview discrimination

- Discrimination experience during the interview
- Perceived fairness

AI impact

- Magnitude of the impact of AI on society
- Magnitude of the impact of AI on personal situation

We create an AI impact magnitude index variable by averaging across the two individual magnitude items.

- Direction of the impact of AI on society
- Direction of the impact of AI on personal situation

We create an AI impact direction index variable by averaging across the two individual direction items.

AI knowledge Subjects face a quiz of five multiple-choice questions. In each question, we provide subjects with a statement about AI, and they need to select the correct response from four potential answers. We construct our AI knowledge variable as the number of correct responses.

AI usage We provide subjects with four commonly used AI tools and ask for each how frequently they use them. We construct our AI usage index by summing up the frequencies.

3.2 Hypotheses and Analysis

Main focus. Our main hypotheses focus on whether behavior in the interviews and the outcomes of interviews differ when an AI voice agent is used instead of an human recruiter. To test for differences, we compare the key outcome variables described in Section 3.1.2 between the *Human interviewer* and *AI interviewer* condition. For binary variables (e.g., job offer rate), we will use two-sample tests of proportions. For numeric variables (e.g., interviewer score), we will use t-tests. We will also separately conduct these analyses depending on the interview mode (called or in the recruitment center).

Interview transcript variables. We will analyze the content of the interview’s transcripts and the audio files using ML-techniques. We will focus on comparing how the content differs between our treatments along the variables described in Section 3.1.3 and 3.1.4.

Choice of interview. We will analyze the determinants of choice based on pre-interview variables and the survey variables as well as how the outcome variables outlined in Section 3.1.2 differ depending on the choice of the interview. We then also relate these results to the two direct treatments by comparing the performance (i) of those choosing the AI interviewer with those in the *AI interviewer* treatment and (ii) of those choosing the Human interviewer with those in the *Human interviewer* treatment.

3.3 Conflict of interest reporting

The authors declare that they have no conflict of interest. In particular, no researcher involved in the covered research has received financial or commercial benefits from the company. No third party had the right to pre-review this pre-registration.