An AI-Powered Interactive Interface to Enhance Accessibility of Interview Training for Military Veterans

Rakesh Chowdary Yarlagadda
University of Colorado Boulder
Boulder, CO, USA
rakeshchowdary.yarlagadda@colorado.edu

Pranjal Aggarwal
University of Colorado Boulder
Boulder, CO, USA
pranjal.aggarwal@colorado.edu

Vaibhav Jamadagni
University of Colorado Boulder
Boulder, CO, USA
vaibhav.jamadagni@colorado.edu

Ghritachi Mahajani
University of Colorado Boulder
Boulder, CO, USA
ghritachi.mahajani@colorado.edu

Pavan Kumar Malasani
University of Colorado Boulder
Boulder, CO, USA
pavankumar.malasani@colorado.edu

Ehsanul Haque Nirjhar
Texas A&M University
College Station, TX, USA
nirjhar71@tamu.edu

Theodora Chaspari
University of Colorado Boulder
Boulder, CO, USA
theodora.chaspari@colorado.edu

Abstract

This demonstration paper presents an artificial intelligence (AI)-powered interactive interface designed to enhance interview training for military veterans transitioning to civilian jobs. The interface uses large language models (LLMs) to provide real-time feedback on veterans' responses to common interview questions, classifying answers as under-explained, succinct, comprehensive, or over-explained. The system further offers a justification of its decision, potentially enhancing the user's understanding of their responses and identifying areas for improvement. This tool aims to bridge the gap between military and civilian employment, addressing unique challenges faced by veterans and potentially extending to other sensitive groups in future applications.

CCS Concepts

• **Human-centered computing** → **Interaction design process and methods**; • **Applied computing** → *Interactive learning environments.*

Keywords

Interview training, interactive interface, large language models, speech, language

ACM Reference Format:
Rakesh Chowdary Yarlagadda, Pranjal Aggarwal, Vaibhav Jamadagni, Ghritachi Mahajani, Pavan Kumar Malasani, Ehsanul Haque Nirjhar, and Theodora Chaspari. 2024. An AI-Powered Interactive Interface to Enhance Accessibility of Interview Training for Military Veterans. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI Companion '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3686215.3688371

ICMI Companion '24, November 04–08, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-0463-5/24/11
https://doi.org/10.1145/3686215.3688371

1 Introduction

Military veterans bring diverse experiences, a wide range of skills, and the benefits of their military training to the civilian workforce [13]. They work well in a team, depict a sense of responsibility and accountability for completing the job tasks, are organized and disciplined, and possess a strong work ethic. Yet, many veterans struggle with integrating to the civilian workforce due to several factors [7]. Civilian interviewers with little or no experience in the military domain are often asked to interview veterans, thus, they might not be aware of the unique challenges associated to returning to civilian life [7]. This can result in military veterans being subject to discrimination, negative stereotypes, stigma, and exclusion [8]. In addition, veterans may struggle to articulate the relevance of their military-specific skills and may not be adequately prepared for the civilian job interview [5].

The employment interview is the most common method used to assess a job candidate [6]. Despite their strong qualifications, veterans might depict several weaknesses when engaging in the civilian job interview, including ineffective translation and communication of relevant technical skills acquired in the military, use of military jargon, and over-explaining their responses [12, 13]. Existing programs for preparing veterans for the civilian job interview are limited and often rely on a "one-fits-all" solution. For example, the U.S. Department of Labor offers a one-day employment overview that teaches military veterans how to build a resume and prepare for an interview, along with an e-learning curriculum to support service members and their spouses. While these programs have been effective in helping veterans to find a job after leaving the military or transition to a better job [9], they may not always address the individual needs and backgrounds of veterans, nor provide tailored support necessary for effective job market integration.

Interactive interfaces that rely on artificial intelligence (AI) for facilitating interview training have received increased interest in several research domains, such as affective computing and human computer interaction (HCI). These assistive technologies help the users in practicing their communication skills by allowing them to answer mock job interview questions and provide feedback related to their verbal and non-verbal cues [1, 3, 4]. Hoque *et al.* developed the My Automated Conversation coacH (MACH) system that employed virtual avatars as interviewers to simulate a job interview scenario [4]. Users were asked questions by the avatars and their responses were recorded. By analyzing the facial expression and verbal cues, MACH generated a visual feedback for the users about their performance. Anderson *et al.* proposed a gamified interface named TARDIS where users interacted with a virtual agent to improve their communication skill [1, 2]. TARDIS analyzed the non-verbal cues related to facial expressions and gestures to provide users a quantitative feedback about their performance. Similarly, Yadav *et al.* focused on developing an assistive interface to provide feedback to the users about their conversational engagement and behavioral cues during interviews. However, these systems do not take into account the linguistic content of the responses. That causes the users to miss out on feedback related to the quality of their response. The proposed demonstration aims to fill this gap in the literature via leveraging large language models (LLMs) due to their ability to understand and generate human-like text and provide feedback on a sentence-level or even word-level.

This demonstration paper presents a AI-based interactive interface for interview training of military veterans. The interface simulates an interview scenario. The user views each question and has the ability to verbally respond or type their answer. A backend system that is based on LLMs classifies the degree of explanation of the user's response to each question (i.e., under-explained, succinct, comprehensive, over-explained), and provides a justification of its decision. In this way, the system can analyze the unique verbal behaviors of a veteran user on a moment-to-moment basis offering personalized feedback and adaptive training that can pinpoint specific parts of responses for improvement. This demonstration is discussed in terms of its generalizability to other sensitive groups.

2 Implementation

2.1 Interface Design

We developed a web application utilizing React for the front end and Flask for the back end. The application leverages the Whisper API to enable near real-time audio transcription. When a user submits their response, the back end processes and classifies the response as under-explained, succinct, over-explained, or comprehensive, providing a detailed reasoning for the classification. This interface can be deployed on the cloud for wide accessibility through a web link, while also supporting local deployment to ensure user privacy.

During the simulated interview session, users are presented with a series of eight questions from various categories that simulate common behavioral interview questions. The categories include introduction, mental capability, knowledge and skills, basic personality tendencies, communication and interpersonal skills, leadership, persuasion and negotiation, interests and preferences, and conclusion. The application is designed with a user-friendly interface (Figure 1), featuring intuitive buttons for navigating between questions, starting over, submitting answers, recording responses, and resetting drafts. Users have the option to type their answers or dictate them using the web browser's microphone, and they can combine both typed and dictated responses.

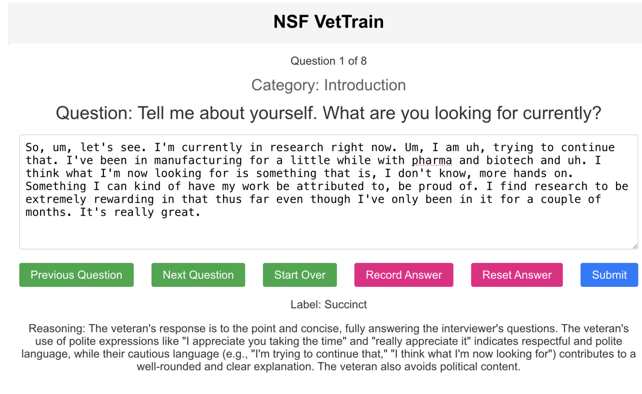

Figure 1: System interface

2.2 System Design

2.2.1 Data. In order to design the AI-based system to identify the degree of explanation in the responses to interview questions, we used the interview dataset introduced in [10] for prompting the LLMs. This dataset consists of data from a study where 38 U.S. military veterans participated in a mock job interview conducted by experienced interviewers via Zoom. Audio recordings and transcripts of the interviews were obtained from Zoom. Collectively, the participants responded to 286 interview questions. To label the degree of explanation of these responses, three annotators were employed. Based on their annotation, each response was labeled as one of the following four possible categories [14]. **Under-explained**: Short and incomplete; **Succinct**: Concise and to-the-point; **Comprehensive**: Detailed and complete; and **Over-explained**: Unnecessarily long.

2.2.2 Automatic speech recognition (ASR). For automatically transcribing the user responses, we evaluated multiple speech-to-text models by calculating their word error rates (WER) and transcription times. Previously recorded participants' audio and corresponding transcripts [10, 14] served as the ground truth for these evaluations. The models compared included Whisper, Faster-Whisper, Distil-Whisper, and Whisper Large-v2. Among these, Whisper Large-v2 achieved the lowest WER of 0.315. However, to balance the trade-off between transcription time and accuracy for near real-time processing, we selected the Whisper Medium model for local deployment. Whisper Medium transcribed an average of 16 minutes of audio in 146.26 seconds, compared to Whisper Large-v2's 205.47 seconds. Despite its slightly higher WER of 0.326, Whisper Medium offers faster transcription due to its smaller model size.

2.2.3 Large Language Models (LLMs) for identifying degree of explanation. Different LLMs were tested across various experimental setups to classify a response into a specific degree of explanation. These classifications were performed through 2-way comparisons: Succinct vs. Under-explained (task 1) and Comprehensive vs. Over-explained (task 2). We used a Chain-of-Thought approach [15] with few-shot learning. In our system, *n-shot* learning means that the prompt includes *n* examples from each class. The experimentation involved modifying the prompts by including various elements such as context from previous questions, domain knowledge on the linguistic characteristics of each type of response, and justifications for the assigned labels of the *n* training examples. The context refers to the number of previously asked questions the model has access to when answering the current question. This context is maintained separately for each participant and increases as the number of questions asked to the participant grows. Domain knowledge was integrated by adding to the prompt specific linguistic characteristics of each type of response, such as the use of tentative language, polite expressions, and language related to achievement and politics, as measured by the Linguistic Inquiry and Word Count (LIWC) [11]. Justifications for each of the *n* training examples were generated by the experimenter according to the annotation manual. The models that yielded the most promising results were integrated into the backend to classify queries from the frontend ASR model into their respective categories.

For task 1, the best model employed Gemini 1.5 Pro, using 1-shot learning and the context of previous questions. The most significant performance improvement was observed when context was provided to the model, yielding macro F1-score of 0.54. This improvement likely indicates that the Gemini model, with its large context window, has a better ability to understand and utilize context. Additionally, the results suggest that integrating domain knowledge to the prompt may not yield significant performance improvement for this classification task. For task 2, we utilized GPT 3.5 Turbo with 2-shot learning. The model considers the justification of labels for 2 examples per class and context of previous questions. The prompt also includes information from domain knowledge. The combination of the three aforementioned design elements yielded a macro F1-score of 0.61 for this task.

3 Discussion

By offering personalized feedback in a controlled environment, the proposed interface can potentially help veterans articulate their skills and experiences more effectively, thus promoting their integration to the civilian workforce. The interface can also help civilian interviewers to better understand the specific challenges faced by candidates from the military, potentially raising awareness about the difficulties these candidates encounter. Yet, the proposed interface has the following limitations that can be addressed in future work. It would be valuable to conduct a formal evaluation of the effectiveness of the interface in interview training and assess not only the classification performance of the interface in unknown data, but also the quality of generated justifications. The interface focuses on the specific population of military veterans, but the framework and methodology could be applied to other sensitive groups, such as individuals with disabilities, those re-entering the workforce after incarceration, and immigrants or refugees. The system can help individuals with disabilities practice articulating their strengths and accommodations needs, assist former inmates in framing their past experiences positively, and aid immigrants in overcoming language barriers and cultural differences in job interviews. Finally, another way to improve the system would be to make the interface adaptive based on the user's skills.

4 Conclusion

This paper presents an interactive interface that is based on LLMs to help military veterans effectively translate their military skills to the civilian job interview. The interface utilizes React for the front end and Flask for the back end. LLMs including Gemini 1.5 Pro and GPT 3.5 Turbo conduct the automatic classification task of the degree of explanation of a response in the interview. Ultimately, this interface can offer personalized feedback to military veterans and help them articulate their skills and experiences more effectively.

Acknowledgments

The authors would like to acknowledge the support of the National Science Foundation (#1956021) for this work.

References

[1] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In *International Conference on Advances in Computer Entertainment Technology*. Springer, 476–491.

[2] Patrick Gebhard, Tanja Schneeberger, Elisabeth André, Tobias Baur, Ionut Damian, Gregor Mehlmann, Cornelius König, and Markus Langer. 2018. Serious games for training social skills in job interviews. *IEEE Transactions on Games* 11, 4 (2018), 340–351.

[3] Arno Hartholt, Sharon Mozgai, and Albert" Skip" Rizzo. 2019. Virtual job interviewing practice for high-anxiety populations. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 238–240.

[4] Mohammed Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM UBICOMP*. 697–706.

[5] Mary Keeling, Sara Kintzle, and Carl A Castro. 2018. Exploring US veterans' post-service employment experiences. *Military Psychology* 30, 1 (2018), 63–69.

[6] Julia Levashina, Christopher J Hartwell, Frederick P Morgeson, and Michael A Campion. 2014. The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology* 67, 1 (2014), 241–293.

[7] Fred Mael, Will Wyatt, and Uma Janardana Iyer. 2022. Veterans to workplace: Keys to successful transition. *Military Psychology* 34, 5 (2022), 516–529.

[8] Charn P McAllister, Jeremy D Mackey, Kaylee J Hackney, and Pamela L Perrewé. 2015. From combat to khakis: An exploratory examination of job stress with veterans. *Military Psychology* 27, 2 (2015), 93–107.

[9] Nicole R Morgan, Katie E Davenport, Keith R Aronson, Kimberly J McCarthy, Julia A Bleser, and Daniel F Perkins. 2022. The relationship between veterans' employment program component use and career advancement over time. *Journal of Veterans Studies* 8, 1 (2022), 147–163.

[10] Ehsanul Haque Nirjhar, Md Nazmus Sakib, Ellen Hagen, Neha Rani, Sharon Lynn Chu, Winfred Arthur, Amir H Behzadan, and Theodora Chaspari. 2022. Investigating the interplay between self-reported and bio-behavioral measures of stress: A pilot study of civilian job interviews with military veterans. In *Proceedings of 2022 ACII*. IEEE, 1–8.

[11] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. (2015).

[12] Deborah Roy, Jana Ross, and Cherie Armour. 2020. Making the transition: How finding a good job is a risky business for military veterans in Northern Ireland. *Military Psychology* 32, 5 (2020), 428–441.

[13] Md Nazmus Sakib, Ellen Hagen, Nidhal Mazza, Neha Rani, Ehsanul Haque Nirjhar, Sharon L Chu, Theodora Chaspari, Amir H Behzadan, and Winfred Arthur Jr. 2024. Capitalizing on strengths and minimizing weaknesses of veterans in civilian employment interviews: Perceptions of interviewers and veteran interviewees. *Military Psychology* (2024), 1–13.

[14] Raghu Verrap, Ehsanul Nirjhar, Ani Nenkova, and Theodora Chaspari. 2022. "Am I Answering My Job Interview Questions Right?": A NLP Approach to Predict Degree of Explanation in Job Interview Responses. In *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*. 122–129.

[15] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *NeurIPS*.