



Understanding Human-AI Misalignment in LLM-Based Job-Seeking Support for Neurodivergent Users

Kaely Hall

School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
khal33@gatech.edu

Marcus Ma

School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
mma81@gatech.edu

Xinyue Zhang

Georgia Institute of Technology
Atlanta, Georgia, USA
xzhang3108@gatech.edu

Vedant Das Swain

New York University
New York, New York, USA
vds@nyu.edu

Jennifer G Kim

School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
jennifer.kim@cc.gatech.edu

Abstract

Large Language Models are often trained on data reflecting neurotypical norms, yet are increasingly deployed to support neurodivergent users in sensitive domains like job-seeking. We examine interactions between neurodivergent job-seekers and a GPT-4-powered career support chatbot through the lens of misalignment. Through analysis of over 300 chat logs and interviews with 15 neurodivergent participants, we found that the chatbot frequently misrepresented users' skills, imposed neurotypical language and expectations, and provided generic or inappropriate advice— even when relevant user data was available. Participants expected the chatbot to interpret implicit insights from their data, however, they sometimes lacked the clarity or confidence to correct the system when it did not, revealing gaps in both AI design and user understanding of system function. Our findings underscore the need for bi-directional alignment between neurodivergent users and LLMs, and call for design strategies that integrate neurodivergent perspectives and preferences to ensure more authentic, personalized, and human-centered AI support.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI; User studies; Empirical studies in accessibility.**

Keywords

Large Language Models, Empirical Study, Neurodivergence, Employment

ACM Reference Format:

Kaely Hall, Marcus Ma, Xinyue Zhang, Vedant Das Swain, and Jennifer G Kim. 2025. Understanding Human-AI Misalignment in LLM-Based Job-Seeking Support for Neurodivergent Users. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*, October

26–29, 2025, Denver, CO, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3663547.3746361>

1 Introduction

Large Language Models (LLMs) are primarily trained on data that reflect normative, neurotypical worldviews [13, 49], despite their ever-growing ability to provide recommendations to users' needs through open-ended and natural conversation [32, 34, 39, 48]. Recent HCI research has found that LLMs not only inherit human biases, but also represent marginalized groups as more homogeneous, a phenomenon that can reinforce stereotypes of these population [49], and might result in outputs that suggest and perpetuate neurotypical norms [18]. In the case of neurodivergent users— individuals whose cognitive functioning diverges from societal norms, including those with ADHD, autism, dyslexia, mood disorders, anxiety disorders, intellectual disabilities, and other natural neurological differences [9]— this imposition of neurotypical norms can result in significant psychological harm [25, 77], as it can cause internalized oppression, where the neurodivergent individual internalizes the constant differences between their worldview and the normative worldview as negative [11, 63]. This can lead to reduced self-efficacy, well-being, and sense of self for this population [65].

Historically, bias emerges in technologies and algorithms when the users have significantly different characteristics from the population that is assumed to use it, such as differences in values, experiences, or world-views [23]. While there has been some HCI work that seeks to capture and design for the values and experiences of neurodivergent technology users [7, 19, 29, 35, 45, 53, 72], there is a lack of extensive research focused specifically on how *misalignments* between user values and system outputs can affect their well-being, self-perception, and interaction outcomes in highly specific contexts. This research gap has significant consequences, as much research has shown that neurodivergent individuals have different lived experiences and preferences in contexts such as learning [78], employment [53], and online housing search behaviors [46], all contexts that have significant effects on quality of life.

LLMs such as ChatGPT are increasingly used by neurodivergent individuals in sensitive domains like employment, relationships, and emotional well-being [18, 19, 33]. In the workplace context,



This work is licensed under a Creative Commons Attribution 4.0 International License. *ASSETS '25, Denver, CO, USA*

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0676-9/25/10
<https://doi.org/10.1145/3663547.3746361>

prior HCI work has shown that LLM-based agents are capable of providing situationally-appropriate emotional support in real time during challenging workplace situations [20], and that neurodivergent individuals want workplace systems that adapt to their unique lived experiences and strengths [29]. Regarding job-seeking and application support, some companies have introduced LLM-powered chatbots designed to support autistic and other neurodivergent users in specific, high-stakes contexts such as job-seeking [1]. However, neurodivergent users often face unique social and institutional barriers to employment [52, 70], and inaccurate or misaligned AI recommendations can intensify these challenges. Since employment plays a critical role in supporting neurodivergent individuals' well-being [26, 56, 66], it is essential that AI systems offer support that is both contextually accurate and aligned to users' values and lived experiences [18, 24].

Yet, while these chatbots may achieve fluency in conversation [41], general pre-trained LLMs often fail to provide the specific, context-aware, and personalized recommendations that neurodivergent users need and seek [38]. Prior work has shown that such systems may suggest masking behaviors or disclose diagnoses in inappropriate contexts, thereby misaligning with users' identities and needs [36]. Compounding this, neurodivergent users frequently exhibit a high degree of trust in technology [19], which increases the risk of harmful real-world consequences when guidance is misaligned [10]. Despite a growing interest in neurodivergent-centered LLM applications, much of the research remains focused on surface-level usability rather than examining *how and why* misalignments arise. We argue that focusing on misalignment offers a powerful methodological and analytical lens. This lens can offer insights into the breakdowns in value representation, interaction expectations, and contextual reasoning within the human-AI interaction. In this study, we use the lens of misalignment to surface and examine these breakdowns in real-world interactions, identifying critical failure points and informing the design of more inclusive, adaptive, and value-aligned AI systems that more accurately reflect the goals and lived experiences of neurodivergent job seekers.

In this work, we aim to understand neurodivergent users' experience and perceptions of their interactions with a neurodivergent-specific job-seeking support chatbot. Through analysis of 348 real-world chatbot interactions from a neurodiversity-focused career support startup company, and interviews with 15 participants, we examine users' values and expectations in the context of the job-seeking interaction, assessing the extent to which the chatbot's responses aligned with them and collaboratively uncovering the misaligned assumptions shaping its outputs. We report that participants frequently experienced misalignments related to the chatbot's overgeneralized advice, neurotypical framing of skills and expectations, and failure to adequately incorporate or interpret profile data. These misalignments often resulting in inauthentic, unhelpful, or even discouraging responses. Ultimately, our contributions to the human-computer interaction (HCI) community include a new perspective to understand human-AI interactions between LLMs and neurodivergent users and design implications to facilitate alignment of outcomes for this population in job-seeking support chatbots.

2 Related Work

2.1 AI Alignment

Human-AI alignment is defined as the design and training of AI systems to exhibit behaviors that are consistent with human values and intentions [37]. Recent HCI research has shown that aligning AI and human interests can increase collaboration outcomes with AI across several dimensions, including specification of goals, execution of processes, and interpretable evaluation [75]. For systems such as chatbots, linguistic alignment (i.e., matching user style and domain knowledge) is particularly important for user interactions to reduce frustration and increase satisfaction [74].

As stochastic parrots—systems that echo statistical patterns from their training data rather than understanding meaning—large language models (LLMs) are fundamentally limited in their ability to genuinely reflect individual needs [13]. Given their scale and training practices, it is impractical to expect such models to perfectly align with any single user's context, such as serving as a job coach specifically tailored to neurodivergent job-seekers. However, existing models can be fine-tuned to retain general language modeling capabilities while demonstrating preferences towards domain-specific specializations. Even in situations where datasets are small (relative to the large-scale text corpora used to train modern LLMs), techniques such as Reinforcement Learning from Human Feedback (RLHF) are effective ways to align LLM outputs with intended responses [50, 61]. However, RLHF requires expert-quality annotations of model responses. Even when the LLM itself cannot be modified, there are still many ways to improve LLM alignment by improving the quality of the prompt fed to the LLM, a subfield of NLP more generally known as *prompt engineering* [68]. For example, including example inputs and outputs in the prompt, known as *in-context learning*, provides information to the LLM for what kind of response to output and improves overall output quality [58]. The value of in-context learning is heavily determined by the selection of high-quality and diverse examples [67, 83], which is of particular importance to ensure that specifications unique to specific niches such as neurodiverse job seeking support are represented. Additionally, when dealing with complex tasks such as job coaching, explicitly asking LLMs to break down reasoning into smaller, explainable steps improves both performance and interpretability of the corresponding output [80].

2.2 Alignment & Neurodivergent Users

While fine-tuning and prompt engineering are effective methods of model adjustment [6, 51, 64], they require specific and extensive data. However, previous HCI research has found that when LLMs are employed to generate descriptions or representations of disabled or neurodivergent groups, they tend to reinforce stereotypical biases, as the training data that is available is predominantly authored or shaped from a neurotypical point of view [5, 24]. Due to the lack of niche and contextualized training data on neurodivergent job-seekers' experiences and perspectives specifically, existing technologies will be aligned to data that is grounded in neurotypical perspectives and assumptions, which can result in poor experiences for neurodivergent users [12, 18, 36].

AI alignment is typically framed as singularly designing AI systems to fit human needs, however, Shen et al. propose that alignment can be **bi-directional** in interactive systems by educating human users about how to better utilize and interpret AI systems [71]. Similarly, Norman posited that if a system can represent its state effectively and that state matches the state of the user, better user experience can be achieved [59]. When applied to LLMs, researchers have found that this remains true, emphasizing that user mental models of the system are equally as important to address as the system’s alignment to the user [55]. Therefore, we adopt a *bi-directional* perspective in our research lens to identify specific instances of misalignment from the perspective of the system *and* the neurodivergent job-seeker.

Bi-directional alignment has been explored in HCI between a human and a computer [22, 79], and between virtual agents [62]. For LLM design, Wang *et al.* found that utilizing linguistic analysis on user inputs can allow a LLM conversational agent to automatically identify the user’s mental model of how it works, and opens design opportunities for the system to help correct their mental model [79]. While this research is advancing the AI capabilities in terms of mutual understanding between the user and the system, it does not take into account communication styles or preferences of neurodivergent users, which could exhibit different characteristics than neurotypical users [17]. Building on this foundation, our work extends the concept of bi-directional alignment by specifically examining *how* misalignments between neurodivergent job-seekers and LLM chatbots emerge, and how LLMs can be adapted to facilitate both human and technological alignment.

2.3 Job-Seeking Support for Neurodiversity

Much HCI research has demonstrated both the need and effectiveness of understanding the neurodivergent community and their relationships to technology [31, 36, 42–44, 47, 53, 54]. Hall *et al.* demonstrated how the employment space is a context where the expectation and reality of neurodivergent employees are different, sometimes adversarial, and how technologies for this group need to be designed explicitly to support neurodivergent employees’ values and work-related goals in order to not undermine or impact their agency [29].

We extend this work by investigating *job-seeking* activities, such as generating cover letters and applying to job posts. We posit that investigation of LLM use in this design space is a good example of misalignments between the technology and neurodivergent users’ perspectives because other HCI research has shown that there exist significant differences between the neurotypical and neurodivergent job-seeking experiences, however, the technologies align more with the neurotypical experience. For neurodivergent job-seekers in particular, searching and filter functions on common job sites reflect neurotypical standards [8], and AI screening and recruitment tools exhibit bias in their algorithms towards this group [60]. These challenges represent a larger problem within society stemming from professional practices being rooted in normative social behaviors and expectations, while not considering or attempting to understand neurodivergent perspectives. We posit that these social and technological challenges result in misalignment in LLM outputs for neurodivergent job seekers, and aim to understand where the

misalignments manifest in real-world, LLM-supported job seeking activities.

Some companies are making efforts to bridge this gap by connecting directly with neurodivergent job-seekers and inclusive employers committed to supporting diversity into their workforce [1–4]. As more companies utilize AI in their platforms and business models to support neurodivergent job-seekers, the importance of developing AI systems that can better align with the unique needs of this group cannot be understated.

3 Methods & Analysis

To conduct this study, we partnered with an online neuroinclusive employment platform with over 46,000 neurodivergent job-seekers [1]. Partner companies post neuroinclusive jobs directly to the platform, and job seekers are matched accordingly. At the time of the study, the platform was piloting an LLM job-seeking support conversational agent, powered by GPT-4, to provide support for neurodivergent users’ job application processes such as cover letter writing, career direction support, and interview preparation. This conversational agent was developed, deployed, and maintained by the platform, not researchers. We will refer to the conversational agent as “the chatbot” in subsequent sections. With IRB approval and the platform’s permission, we collected and examined conversations with the chatbot, and subsequently conducted interviews with real users. Below, we describe additional details of the neurodivergent career matching platform’s chatbot setup.

3.1 LLM Job-Seeking Support Chatbot

3.1.1 Chatbot Persona. Large language models see increased performance when prompted to take on a persona before answering complex questions (i.e., “You are a leading expert in mathematics. How would you prove the following theorem?”) [28, 76]. When users start a conversation with the chatbot in this study, they see a simple initiating question based on one of four conversation topics (*Career Advice*, *Writing a Cover Letter*, *Interview Preparation*, and *Writing a Professional Email*). On the backend, the chatbot is prompted with a complex job-coach persona description and also provided access to information from the user profile, which is created by users upon making an account with the platform. This includes relevant information such as previous job history, work preferences and interests, skills and experience, strengths, and diagnosis information. For privacy and anonymity purposes, exact language and details of the backend prompts to the chatbot are not reported in this paper. See Figure 1 for a visual summary of the backend and frontend processes.

3.2 Recruitment & Chat Log Collection

After IRB approval and approval from the platform’s legal team, the platform provided our research team with de-identified chat logs from the chatbot users who consented to participate in our study. The time period of collection was October 2023 through February 2024. All user names and identifying data within the chat logs were redacted prior to delivery by the platform’s development team. The purpose of analyzing the chat logs was to understand what types of questions neurodivergent job-seekers asked a job-support chatbot, and inform specific questions for qualitative interviews

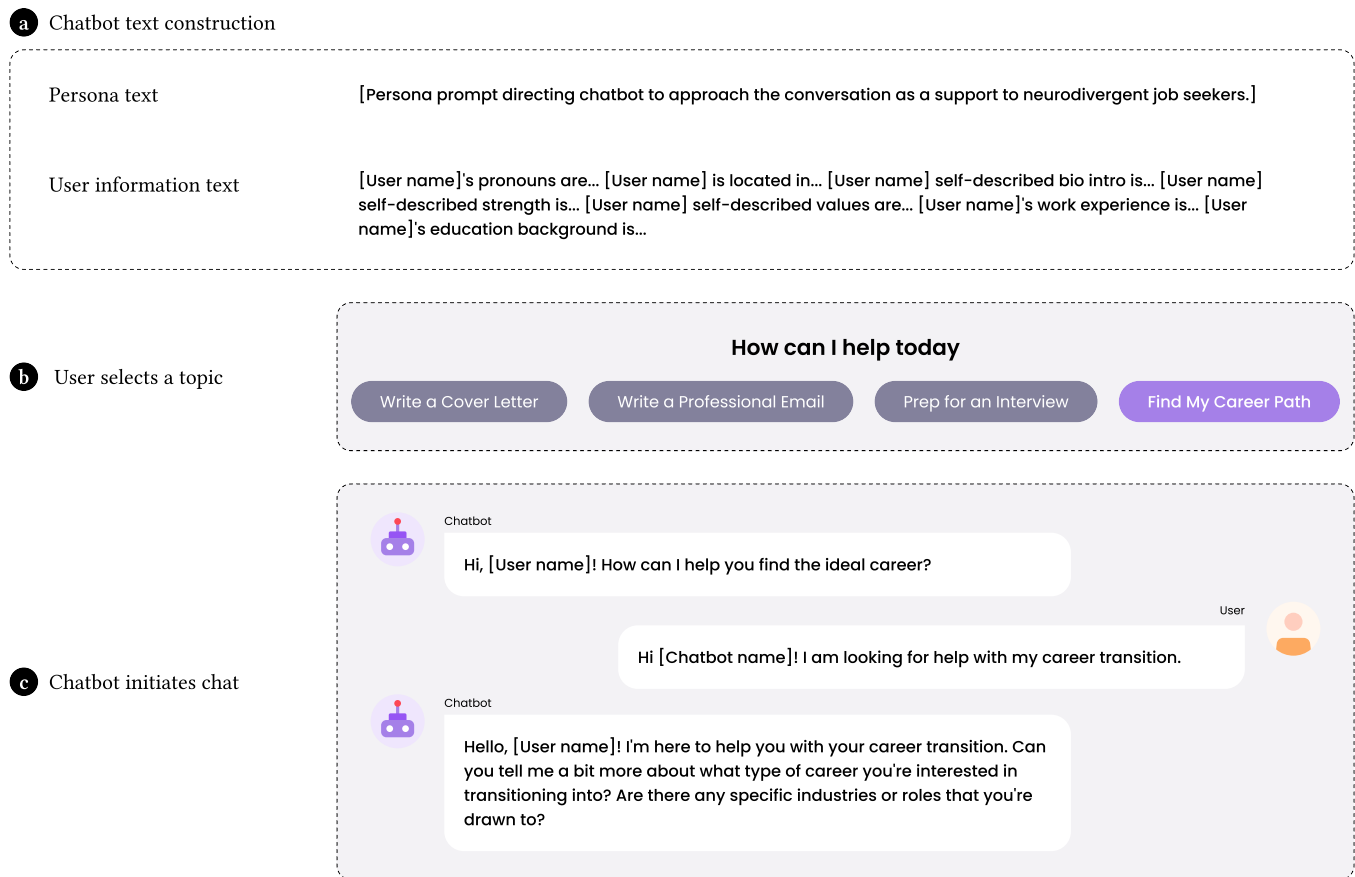


Figure 1: Example of the backend chatbot persona prompting and the front-facing user interaction.

with the participants. Data sets included a de-identified user ID, a de-identified chat session ID, chat topic selected by the user, timestamp information, the user's profile information (including work experience, resume, interests and diagnoses), and message texts to and from the chatbot. Upon initiating a conversation with the chatbot, the user is prompted with a pop-up that asks if they would be willing to participate in a University research study that seeks to understand their chatbot experience. If the user selected YES to this question, researchers assessed their data for inclusion criteria (discussed below), and sent the de-identified user IDs back to the platform team. The platform team then sent an email to the user that included a study description, IRB-approved consent form, and link to sign up for a 1.5 hour study session with the research team. Upon signing up for a session, the user consented to be interviewed and have their chat logs de-identified to the researchers for discussion.

The chat log data set included 348 total user-initiated conversations with the chatbot across 271 unique users, with the majority of users (85.2%) recording one conversation with the chatbot. The total number of unique user messages was 1,100, with an average of 3.16 user-sent messages per conversation. 60.0% of users identified as autistic, with a co-diagnosis of 69.7% with ADHD/ADD, 25.9% with Anxiety, and 17.2% with PTSD.

3.3 Participants & Study Sessions

Inclusion criteria for this study required that the participant have a diagnosis of autism (co-diagnoses were allowed), be at least 18 years of age, and reside in the U.S. or Canada. Diagnoses shared with the research team were derived from the user's profile, and were self-reported. The platform did not require official verification of diagnoses. Participants were compensated with a \$40 Amazon gift card, and a total of 15 participants were interviewed for this study (See Table 1). Study sessions were conducted on the Zoom online platform, and were audio- and video-recorded.

During each study session, the participant's chat log was shared with the participant to refresh their memory of their interaction with the chatbot. These interactions were reviewed line-by-line with the participant, followed by a semi-structured interview intended to understand their experience with the chatbot. Specifically, researchers were interested in understanding what about their interactions with the chatbot did not meet their specific goals or expectations.

3.3.1 Chat Log & Interview Analysis. Our initial goal in reviewing chat logs was to understand the scope of user questions and the types of support they sought from the chatbot. Users selected one of four Chat Topics to begin a session: career advice, cover letter

writing, interview preparation, or professional emails. However, we found that many queries extended beyond these categories. To capture this, we conducted thematic analysis and open coding, identifying sub-themes such as career transition, emotional support, and career matching under career advice, and interview practice and emotional support under interview preparation. This allowed us to understand both the nature of users' questions and the currently unmet support they were seeking. For example, we noticed that when participants asked the chatbot to generate content for them (cover letters, emails), the logs were shorter and less information seeking. We could not infer if these interactions were satisfactory or not. Additionally, it was clear to the research team from the breadth of question types and inputs that the expectations of what information the chatbot could provide varied greatly among the dataset. Below we explain out detailed analysis to inform our research scope and ultimately inform personalized interview questions for each participant.

Chat Log Analysis The chat topic and sub-themes of participant chat logs can be found in Table 1. *Chat topic* indicates which of the four job-seeking topics chosen by the participant to initiate the conversation with the chatbot (*Career Advice*, *Writing a Cover Letter*, *Interview Preparation*, and *Writing a Professional Email*). Of the 348 conversations collected, 15.8% chats were *Writing a Cover Letter*, 70.7% were *Career Interest*, 9.2% were *Interview Preparation* and 4.3% were *Writing a Professional Email*. While these four topics broadly categorized the interests of chatbot users, thematic analysis and open coding [16] of user inputs revealed sub-themes for the *Career Interest* and *Interview Preparation* topics. The *Career Advice* topic included the sub-themes transition support, emotional support, career matching. Transition support inquiries from chatbot users included seeking advice on how to transition and communicate their skills in a new field, such as “*I’m hoping to switch into UX jobs and would love help figuring out next steps*” (P15). The user input “*Hello. I feel sort of stuck at the moment. I thought working as a clinical psychologist would work out for me, but I have been unable to get into a PsyD program. I feel discouraged since this is what I trained for since college*”(P5) is a *Career Advice* topic, however, the user’s language suggests that they are in need of encouragement as well, labeling this input as emotional support. Career matching inquiries occurred when the user implies the chatbot should directly match them with a career direction, such as “*remote job entry level, not dealing with customers*” (P3), or “*I am not sure what career to go into*” (P9). *Interview Preparation* topic included sub-themes practice and emotional support. Practice inquiries occurred when users asked the chatbot to help them prepare for the questions in the interview, such as “*I am doing an interview with a recruiter soon, and she would like to talk with me about her company and some roles they might have open. How should I best prepare for this?*” (P10). Many users expressed anxiety surrounding interviews. If anxiety or stress related to an interview was mentioned, the inquiry was labeled emotional support under *Interview Preparation* topic, separate from *Career Advice*. For anonymity purposes, we do not report content from chatlogs where the user did not consent to be interviewed.

Interview Development & Analysis. For the interviews, we labeled participant chat logs with the above out-lined categories and included general questions in the interview scripts regarding the participant’s understanding of how AI-powered chatbots worked,

what their motivation for using the chatbot was, and what types of information it was or was not capable of providing. Further, we generated targeted questions based on specific moments in each conversation, tailoring interviews to participants’ unique experiences. For example, in the interview of a participant engaged in several messages with the chatbot regarding his mental and emotional state after being laid off, we included more specific interview questions to understand what aspects of the exchange felt meaningful for the participant. This individualized approach ensured interviews were both grounded in the data and responsive to each participant’s lived experience. The interviews also aimed to explore participants’ overall experiences with the chatbot, their perceptions of the chatbot’s responses and recommendations, and their understanding of how to interact with the system to achieve desired outcome.

The first author reviewed the recordings of each interview and conducted a top-down, deductive coding process to identify emergent themes across participants’ experiences, starting with the themes and codes outlined above and understanding each participant experience through the lens of alignment. Our focus on alignment stems from the paper’s core goal: to understand how and why misalignments occur between neurodivergent job-seekers and the chatbot. This lens allowed us to examine user expectations, chatbot responses, and interaction dynamics. Rather than using inductive thematic analysis, we identified themes (inauthentic characterization, perpetuation of neurotypical norms, user-driven alignment) as drivers of misalignment, or how much the LLM output resonates or diverges from user intentions, understandings, and expectations. Following this and on an iterative basis, a second author independently reviewed and collaborated to refine the codes. Final codes and interview quotes were reviewed with the full research team. During this process, qualitative data and illustrative quotes were collected to capture key insights into instances of misalignment between participants’ expectations and intentions and the chatbot’s responses. Themes included inauthentic characterization of participant skills and experiences, perpetuation of neurotypical norms, the chatbot’s failure to capture implicit context, user-driven alignment, and misalignment due to user ambiguity.

4 Findings

In this section, we will present our findings from analysis of participant interviews and related chatbot logs, highlighting misalignments between the participant and the chatbot. First, we provide examples of chatbot responses that do not align with participant self understandings, intentions, and goals. Following this, we describe instances where participants were able to redirect and correct the chatbot’s misaligned responses, and contrast these with interactions where alignment could not be achieved. Finally, we present evidence of misalignment due to participants not having a clear idea of what they need from the chatbot, and why this is challenging for neurodivergent users in particular.

4.1 Misalignment in Job-Seeking Support Chatbot Interactions


4.1.1 Inauthentic Characterizations of Skills and Experiences. We observed misalignments between our participants and the chatbot’s

Participant ID	Age	Gender	Diagnosis	Duration Job-Seeking	Chat Topic & Sub-Themes
P1	31	M	Autism, ADHD, Anxiety	3 months	Career Advice, Emotional Support
P3	30	F	Autism, ADHD Combined Type	5 years	Career Advice, Career Matching
P4	31	M	Autism, ADHD/ADD, Anxiety	1 year	Interview Prep
P5	26	M	Autism, Anxiety	N/A	Career Advice, Emotional Support
P6	35	NB	Autism, ADHD/ADD, PTSD, Anxiety	3 years	Career Advice, Transition
P8	42	M	Autism	3 months	Career Advice, Career Matching
P9	36	F	Autism, ADHD/ADD	1 year	Career Advice, Career Matching
P10	30	NB	Autism, ADD/ADHD	N/A	Interview Prep, Practice
P13	41	M	Autism, ADHD/ADD	N/A	Career Advice, Transition
P14	33	F	Autism	N/A	Career Advice, Transition, Cover Letter
P15	31	F	Autism, ADHD/ADD	3 months	Career Advice, Transition
P16	21	F	Autism, Intellectual Disability	2 years	Career Advice, Transition, Cover Letter
P17	31	M	Autism, ADHD/ADD	3 months	Career Advice, Emotional Support
P18	25	F	Autism, ADHD, PTSD	2 months	Career Advice, Transition
P19	35	F	Autism, ADHD/ADD	3 months	Career Advice, Transition


Table 1: Summary of study participants and chatbot topics. Some participants initiated multiple chats, leading to multiple chat topic themes

responses in instances where participants sought help with career transitions. Specifically, when participants asked the chatbot to generate cover letters for applications to new jobs, a task that required taking their current skills and experiences and applying them to new contexts, **the chatbot tended to overemphasize skills and abilities, even if proficiency was not indicated in the participant’s profile.**

For example, P14 input:

 Hi [chatbot name]! I am a systems analyst but would like to transition into more creative work.

Following this input, she asked the chatbot to create a cover letter for her transition to graphic design work. The chatbot’s suggested cover letter included the following content:

 As a graphic designer, I have honed my skills in creating visually appealing and impactful designs across various platforms. My experience includes designing marketing materials, branding assets, social media graphics, and website design. I am proficient in graphic design software, such as Adobe Creative Suite, and have a keen eye for detail and aesthetics... I possess a Bachelor’s degree in Graphic Design and have completed multiple design projects that have received positive feedback from clients and colleagues.

While this response may read as if P14 is a strong candidate for a graphic design position, it was problematic for the participant. Most concerning is the claim of possessing a degree in Graphic Design. P14’s profile did not include a Bachelor’s degree in Graphic Design or any feedback from clients on design projects. While P14 did have a “Contract Web Designer” position listed in her background, the

only design work listed with it was “Created custom graphic assets and style guide.” The rest of her profile was heavily financial systems analyst work. Regarding this, she explained: “I have web design on there and some graphic design on my resume and so I was like it’s there, you just aren’t quite using it right” (P14). This finding suggests that even when relevant information is available for the chatbot use, it is not applied in the response correctly, creating a breakdown between how the user wants their skills to be represented and how the chatbot portrays them.

Regarding the language suggesting proficiency in tools and areas where P14 had little experience, she stated “I would never put it that way because I’m a systems analyst with freelance design experience” (P14). This can create tension and dissonance for neurodivergent job-seekers. Job seeking is already a challenging experience for many neurodivergent people [21], and seeing a response such as this might discourage them further as they may internalize that the job requires “superior” skills they do not possess. Further, this finding echoes recent HCI work showing that LLM-generated descriptions of disabled or neurodivergent individuals often rely on stereotypes or confidently assert hallucinated information. For example, an AI tool described a blind person as “courageous” and “resilient”—a characterization that a participant critiqued as inappropriate, noting it reflected how neurotypical people tend to frame blindness [5]. From a model perspective, the tendency of the LLM to confidently generate exaggerated or inaccurate information—such as hallucinating a degree—reflects known limitations of current models, which frequently prioritize fluency and coherence over factuality or contextual appropriateness [13, 38]. These challenges

can compound one another, resulting in outputs that grossly misrepresent neurodivergent users' qualifications and experiences.

Ultimately, P14 felt that this cover letter did not represent her skills in graphic design accurately and that the definitive language used to describe her as a strong candidate for a graphic design position did not align with her intended goal, which was a career transition. P14 expected the chatbot to provide a cover letter that accurately represented her level of skill as a graphic designer, while also contextualizing her experience as a financial analyst into the new role.

Additionally, Other participants expressed similar misalignment with over-generalized or over-representative language that they experience during job-seeking activities. Many participants noted that this language is rooted in neurotypical norms. For example, P1 and P5 expressed that when considering job posts to apply to, they are discouraged by language and required skills such as “quick learner,” can “work independently,” and can “thrive in a fast-paced environment.” From their perspectives, this language perpetuates neurotypical work environments and expectations that are not inclusive of their needs. P5 felt that these are inauthentic characterizations of their preferences: they prefer work environments that emphasize supported task learning, guidance on work assignments, and predictability in day-to-day activities. *“Generally I like it when I hear that there’s support for someone going into the position. So for instance, when I see [posts] that [say] ‘the person must be a self starter’ I know that’s not necessarily for me” (P5).*


This finding supports that our participants did not want to overemphasize or mischaracterize their abilities, and instead wanted their skills and abilities represented exactly how they are. However, P1 told researchers that he feels if he does not include this language in his applications, he won't be considered for the position. *“I feel like I’m a good worker. I have skills, and I’ve done good things and that should be enough, but it seems like [in] the work search these days like they want somebody who is who’s ‘quick,’ and it’s like, come on, give me a break” (P1).* Research into neurodivergent perspectives has shown that in job-seeking practices, these pressures can cause dissonance for the individual, and make them believe they have to “mask” their behaviors that do not align with normative expectations [21]. Further, our participants noted that new AI screening tools used in job recruiting specifically look for these types of key phrases. If they do not include the “right” language, even if it is misaligned with their self understanding and work preferences, they feel their application will get kicked out.


Because our participants expressed concern and uncomfortability with over-representation of their skills suggested by the chatbot but pressure to do it anyway, instances such as this could result in adverse outcomes, such as interviewing for a job that the individual is not qualified for or misaligned expectations between the job-seeker and the employer. The potential for this poses great risk, as the neurodivergent population exhibits greater vulnerability to negative work-related experiences [27]. Further, the user may internalize negative feelings toward themselves and their abilities [73], reducing their job-seeking self-efficacy [69]. For instance, if a job-seeking support chatbot provides a cover letter for a user who is applying to a job and “inflates” their abilities or work-style preferences, the user may experience heightened anxiety of actually being able to secure and perform in that position.

4.1.2 Failure to Capture Implicit Preferences. Another misalignment between the chatbot and the study participants was the chatbot's superficial use of the participants' skills and interests in its responses. Users expected the chatbot to possess a deeper understanding of their experiences and work preferences based on their profile data and recommend relevant job opportunities. Instead, the chatbot made surface-level matches based on the participants' self-reported skills and interests. From participants' perspectives, it did not take into account their preferences for factors like work environment or job type (e.g., freelance, corporate, etc.), although they felt these preferences were implicit through their profile data and would therefore be used in the chatbots reasoning.

For example, some of our study participants felt they had good work ethic and skills, but not in a “typical” way, and thus pursued careers that were not “typical,” such as jobs that have defined hierarchical employment structures. P3 did not complete a college degree, stating her diagnosis made traditional education difficult for her and a burden on her mental health. Instead, she acquired experience through non-traditional jobs, such as completing freelance projects. *“It’s easier for me to be self-employed than it is for me to work for someone else, which is common with people with ADHD” (P3).* She has been applying for freelance design work, however, she feels that not having formal training is deterring employers from considering her applications. She utilized the chatbot to help her identify jobs that would be amenable to her “non-traditional” career path and were appropriate for her background and experience.

P3 chose the chatbot topic Career Advice, and the following conversation resulted:

 remote job entry level, not dealing with customers

 Sure, I understand that you're looking for an entry-level remote job that doesn't involve direct customer interaction. Here are a few options that might align with your strengths and preferences:

1. Data Entry Specialist: You mentioned that you're good with computer-based tasks, and this role involves accurately inputting and managing data in databases or spreadsheets.
2. Research Assistant: Since you mention that you enjoy solving complex problems that require research, a remote research assistant role might be a good fit. You could assist with gathering and analyzing data, conducting literature reviews, and organizing research materials.


Although the chatbot seemed to utilize her profile information, P3 was ultimately unsatisfied with this answer. *“It’s all kind of stuff that I’ve heard before, but it’s not so easy to get. It’s not like bad or anything, it’s just not really helpful for me personally” (P3).* She felt that the suggestions were jobs that required college degrees and formal training, and therefore, they did not match her level of experience or job-type preferences. The chatbot did not account for P3's unique context, in which she prefers freelance work and wanted to apply to positions that met her abilities.

Other participants were similarly frustrated with “generic” and “common sense” recommendations from the chatbot. As P13 stated about the response the chatbot provided after inquiring about a career transition, the chatbot gave *“standard information that is repackaged and presented in different ways. It did a good job summarizing the steps, but they are fairly generic steps” (P13).* Similarly,


reflecting on advice the chatbot provided after asking about how he can apply his skills and experience to a new position, P1 stated, “*It’s good general advice, but I was hoping to get something more personal*” (P1). Another participant, frustrated with the same “general advice,” asked the chatbot directly “*How do I use you to get personalized advice?*” (P19).

4.2 User-Driven Alignment

In contrast to the examples above, other interactions between the chatbot and participants achieved more satisfactory and mutually aligned outcomes (P4, P8, P9, P13, P17, P18). Participants were most satisfied with their interaction with the chatbot when they engaged more intentionally and were able to provide richer context throughout the conversation. Participants who shared specific career interests, example job roles, relevant qualifications, and skills they wanted to develop were able to elicit more aligned and actionable recommendations, such as contextualized resume and cover letter suggestions, and concrete next steps for learning position-relevant skills. These more aligned interactions often resulted in longer chats where users were able to redirect the conversation when the chatbot provided responses that were not aligned with their specific needs. In these cases, the chatbot was able to provide suggestions that matched the user’s goals and constraints. For example, after telling the chatbot they want to transition to a new career, P6 entered additional context:


 My area of specialty is human behavior research, research analytics, and leadership. The tech and gaming industry has always appealed to me. But I can work in many different fields. I am avoiding education, mental health, and therapy based industries. Roles in User research, human research, behavior research or science, research design and/or analytics, or product research and design.

In response, the chatbot offered roles such as user experience (UX) researcher, market researcher, product researcher, and data analyst. Unlike P3’s interaction (See Section 4.1.2), where she did not add any specifics about her preferred job-type and expected the chatbot to interpret it from her profile, P6 added specific job examples, and even exclusion criteria. As a result, P6 was much more satisfied with the interaction, felt that these recommendations accurately reflected their experience and personal preferences for job-types. Additionally, each role was accompanied by responsibilities and how the role would utilize P6’s specific experience. For example, for the UX role, the chatbot included the following context:


 In this role, you would be responsible for conducting research to understand user behaviors, needs, and preferences, and using that information to inform the design and development of user-friendly products and experiences. Your skills in research analytics and leadership would be valuable in analyzing and interpreting data, and in guiding and collaborating with cross-functional teams.

The chatbot also advised P6 to update their resume with relevant skills to the position that they wanted to pursue. P6 stated that this response was aligned with their specific job-seeking needs, and they were happy with the level of detail and specificity.


P8 wanted to transition from backend development to frontend development, and asked the chatbot for help with this transition. He added specific job skills that he wanted to learn, and asked for help with the next steps:

 I’m trying to find a role that would give me the opportunity to learn a JavaScript framework such as React or Angular. I want to pivot more towards UI development roles, but my lack of experience with a JavaScript framework is preventing me from getting these roles. What should I do to correct my career path?

The chatbot responded with a list of steps for him to “*overcome [his] lack of experience with these frameworks,*” such as self-study, building a portfolio, open-source contributions, networking, and freelance or volunteer work. P8 described that he was unsatisfied with this initial answer, because it was generic and contained information that he already knew. In response, he redirected the chatbot and added more details about his background:

 I’ve already taken a short course in React a couple of years ago. I don’t have any idea for a project that isn’t something generic like To Do list. I only lack on-the-job experience.

In response, the chatbot focused its answer on specific ways that P8 could get “*on the job experience,*” including personal projects, collaborative projects, freelance work, internships, and creating a Git repository. Each suggestion included added context on how to accomplish the task and how it would benefit his experience. For example, for creating a Git repository, the suggestion that P8 found most useful, the chatbot included:

 Maintain an active GitHub or GitLab repository where you showcase your React projects and contributions. This will allow potential employers to review your code, evaluate your skills, and see your commit history.


By specifying what he wanted more on the job experience and providing the chatbot with his current skill level, P8 was able to enter enough context to receive very contextual and personalized responses and action items from the chatbot. He explained “*this has been my experience with AI bots in general: you have to kind of feed them a little bit more information about where you’re coming from*” (P8).

Both P6 and P8 were exploring software-related roles and approached their interactions with the chatbot with clearly articulated goals, such as transitioning into a new technical position or identifying specific skills to develop, respectively. They utilized the chatbot as a supplement to an already semi-informed job-seeking process, which provided them the ability to critically assess whether the chatbot’s responses aligned with their needs. Notably, their backgrounds in software engineering likely contributed to a higher degree of AI literacy, giving them an advantage in understanding how the system worked, recognizing limitations in its responses, and steering the conversation more effectively. This technical familiarity may not be shared by participants in other fields, potentially contributing to the disparity between those who were able to achieve aligned outcomes and those who were not.

4.3 Misalignment due to User Ambiguity

In this section, we present insights that inform why some of our study participants were unable to direct the interaction with the chatbot and achieve alignment. P3 (see Section 4.1.2), for example, gave the chatbot minimal extra detail or context in her inquiry. In fact, because the chatbot had access to participants' user profiles, several of our study participants felt that the chatbot already had enough information to interpret the context of their unique situation, and should be able to respond to their queries in alignment with their expectations without added information. Regarding this, P1 stated *“Well, I mean, knowing that it had access to my profile, I would think that it could have parsed together a little bit better understanding”* (P1).

Additionally, participants' understandings of what they needed from the chatbot varied widely as well. P6 and P8 had clear directions and motives before engaging with the chatbot, which allowed them to redirect and correct the chatbot when they weren't receiving outputs that aligned with their intentions or expectations. Other participants, however, told the chatbot that they were “lost” (P17) or did not know where to start on their job search journey (P9, P19), and relied more heavily on the chatbot to provide direction for them, often with less detail. For example, P17 input:


 I was recently laid off from my dream job and I am feeling very lost as to what to do next.

This suggests that this participant did not have clear intentions when engaging the chatbot. Because effective job-seeking requires identifying and committing to a clear employment goal to drive search behavior [40], participants who are unsure of their goals may struggle to provide accurate information to a chatbot. Therefore, the chatbot will provide misaligned responses as a result of ambiguous, non-specific inputs.

Further, P17 told researchers that he did not have a clear understanding of the chatbot's capabilities when starting this conversation. He stated *“I've used a lot of different, you know, LLMs and there's always a variety in terms of how much it could do... I didn't know if it was what kind of information it had, or if it was connected to any kind of a database for job-seekers. I wasn't quite sure”* (P17). Additionally, P1 stated: *“It's a powerful tool that has a lot of potential, but ... it's a little bit tricky to know exactly how to use it”* (P1), indicating that some users also do not know how to interact with the chatbot to achieve the best outcome for their needs. Many participants simply did not respond or engage the chatbot further when they received a response that did not match their intentions or expectations, even if the chatbot asked them guiding questions within the response.

These findings suggest that both unclear interaction intentions and unfamiliarity with LLM usage contribute to misalignments between participants and the chatbot because the user is either unable to provide additional context for more specific conversations, or they do not know how to redirect the chatbot when it responds incorrectly. This can be especially challenging for neurodivergent job-seekers, as job seeking can be a potentially stressful domain space [21]. P1, for example, stated that he experiences intense rejection sensitivity due to his diagnosis of autism, and this affects his job-seeking activities. In fact, many of our participants expressed

exasperation in the job-seeking process to the chatbot. P5 input to the chatbot:

 I take things too personally, and I feel like I failed if things don't work out, even if its not the case. Not to mention that job searching is stressful for me too... it is all so overwhelming, and it makes it hard for me to think about what to do.

When a user is in an already heightened state of emotional dysregulation, a poor interaction with a technology that is positioned to support them may exasperate the feelings of anxiety. Further, neurodivergent individuals might internalize misalignments with technology as a shortcoming of their own rather than a socio-technical constraint within the technology. Therefore, it is crucial for a job-seeking support LLM to facilitate an understanding of the reasoning used to generate responses and recommendations. If the user can understand and verify the information that the chatbot is using in its logic, they are better able to discern whether or not the chatbot's recommendation is sound and aligned with their expectations, and the misalignments can be directly attributed to the technology. Designs such as this can reduce the potential for the user to internalize negative thoughts about their perceived shortcomings because the response is framed in a way that externalizes the source of error, encouraging the user to see misalignment as a limitation of the system and not a reflection of their own abilities, knowledge, or worth.

5 Discussion

Understanding how neurodivergent job-seekers interact with the LLM chatbot through the lens of misalignment enables us to identify and understand specific instances where incorporating unique user preferences and profile data is most effective in aligning human-AI interactions. Further, in the absence of training data that represents the neurodivergent experience [21], our work contributes insights into how to design LLMs to scaffold the collection of context-rich, user-driven examples and preferences that reflect neurodivergent lived experiences, enabling more personalized and aligned support. In the findings, we highlighted instances where researchers and designers should pay particular attention to the underlying logic in chatbot outputs, the misapplication or neglect of user-provided context, and the lack of transparency in how recommendations are generated—all of which can lead to misalignments that can potentially undermine users' self-efficacy.

In the following sections, we suggest ideas to mitigate misalignments in interactions between neurodivergent users and LLMs. We expand on previous work in Human-Centered AI (HCAI) and suggest implications for HCI researchers and developers to approach new LLM technologies for neurodivergent individuals from a *bi-directional* perspective [71]. This approach integrates AI alignment, explainable AI, and AI ethics in an effort to design supportive systems that achieve more intentionally aligned interactions, resulting in more satisfactory experiences and outcomes that account for the nuance of diverse perspectives and lived experiences.

5.1 Bi-Directional Alignment to Support Neurodivergent Job-Seekers

To address the misalignments that surfaced in our study, we posit that bi-directional human-AI alignment [71] is an effective approach to achieve aligned LLM responses that support highly specific use-cases such as job seeking support for neurodivergent users. Bi-directional human-AI alignment expands the traditional approach to alignment— where the AI is precisely aligned with human values and social intentions— to include investigation of how *humans can align with AI*. We apply this concept by suggesting how the LLM not only adapts to the user’s stated goals and preferences, but also provides transparency and scaffolding that help users better understand, direct, and correct the system’s behavior. This is especially important for neurodivergent users, whose values, self-understanding, and needs may fall outside neurotypical norms typically represented LLM training data. Further, we present how bi-directional alignment can support greater user agency and facilitate outputs that are more accurately reflective of the user’s authentic self, context, and goals while reducing the risk of harmful or misaligned recommendations.

5.1.1 Misalignments as Opportunities for Intervention. A significant challenge in aligning technology to humans is carefully specifying the correct values of the system [71]. HCI researchers have explored ways of extracting relevant data from particular user groups for specific contexts such as strengths-based support for neurodivergent workers [29], problem solving capacities of non-english speaking parents [82], and value-driven technology designs [81]. In this work, we extend that literature by using misalignment as an analytical lens to surface *where, how, and why* breakdowns occur during neurodivergent users’ interactions with a job-seeking support chatbot. By doing so, we suggest critical opportunities to better utilize user profiles, context, and preferences directly into LLM interactions.

Our findings reveal that misalignment is often rooted in the chatbot’s reliance on normative assumptions about job-seeking behavior and language. For instance, in generating resume or cover letter content, the LLM frequently reproduced neurotypical job-seeking norms, employing language such as “quick learner,” “works well under pressure,” or “thrives in fast-paced environments,” that conflicted with how participants wanted to represent themselves. These terms not only felt inauthentic, but also represented a work environment the participants felt they would not be successful in. Yet, they felt the need to echo these terms to be a “strong candidate.”

This misrepresentation highlights a critical opportunity: rather than optimizing for generalized outputs, LLMs should be designed to integrate individualized user profiles, preferences, and values directly into their reasoning and generation processes. In our more successful interactions (4.2), participants who could articulate specific needs or offer corrective feedback helped guide the chatbot toward more contextually appropriate and affirming responses. In contrast, when users were unsure of what to ask or how to redirect the system (4.3), the chatbot defaulted to generic advice that often perpetuated neurotypical standards and led to feelings of misrepresentation. Therefore, LLMs must also scaffold alignment by prompting for clarification, offering tailored examples, and transparently explaining their reasoning, especially in cases where

user goals are under-specified or ambiguous. In doing so, systems can help users clarify their intentions, while also recognizing and responding to the nuances of neurodivergent self-representation. This approach can support neurodivergent job-seekers in feeling authentically represented without the pressure to mask their behaviors or conform to expectations that conflict with their values and identities [21, 65].

Additionally, we suggest that effective chatbot support for neurodivergent job-seekers could better leverage user profiles and preferences to distinguish between features of a person’s background that can be flexibly re-contextualized and those that are fixed. Based on our findings, participants were dissatisfied with the incorrect applications of the experience listed in their profiles. To remedy this, when a user seeks help with a career transition, an LLM could treat their core qualifications, such as degrees or certifications, as fixed, while recognizing that the application or framing of their skills may be adapted depending on the target role. These distinctions are critical for ensuring the user’s identity and background are represented accurately and ethically. Misrepresenting immutable details can not only jeopardize job prospects but also contribute to negative self-perception, as demonstrated by our participants feeling discouraged when the chatbot recommended jobs that they felt unqualified for even though the chatbot had access to their skills and background.

A hypothetical bi-directional system that could address the above-outlined challenges might be one that highlights aspects of a users’ profile or experiences that could be reframed, accompanied by explanations and interactions that allow the user to approve, correct, and add context the LLM’s logic. Because correcting bias or misrepresentations often places a disproportionate burden on the “othered” individuals— and with LLMs, typically requires specialized prompting knowledge [5]— such a system must proactively surface its own potential for misalignment through explanation and support direct user editing. As prior work has shown, blind participants expended considerable effort providing feedback to mitigate ableist AI responses, exemplifying the significant advocacy some individuals must perform to voice their needs and address equity gaps reinforced by technology. A system designed to do this would support a collaborative alignment interaction, where the user is able to input contextual details directly into the instance of misalignment— supporting transparency, agency, and control within the interaction. For example, consider a user with a history of freelance or part-time work(4.1.2), who might expect the chatbot to infer these preferences from their prior roles and incorporate them into its recommendations. Instead of the system applying its logic in opacity, it could flag the pattern and proactively ask the user if they want another freelance position. In doing so, such systems would support collaborative alignment, allowing neurodivergent users to actively shape the system’s outputs in ways that align with their authentic goals and lived experience.

5.1.2 Technical Alignment of LLMs to Neurodivergent Users. Participant critiques in our study should also be contextualized within the current technical capabilities and limitations of LLMs. Participants expressed frustration when the chatbot overstated their qualifications, misrepresented their skills, or adopted language that felt

inauthentic. While these experiences do indicate instances of misalignment, they also underscore systemic challenges in the design of LLM-based applications. A key tension lies in the tradeoff between broad generalization and context-specific accuracy [14, 38]. Generalization enables LLMs to produce fluent and seemingly coherent responses across diverse prompts by echoing statistical patterns in their massive training corpora [13]. However, this fluency often comes at the expense of grounded reasoning or fidelity to individual user data [15]; a phenomenon exasperated for users whose experiences diverge from normative training distributions [24]. For example, neurodivergent participants in our study frequently noted that the chatbot presented standard job-seeking templates or stereotyped motivational and upbeat job-seeking language, even when their profile data suggested alternative needs or perspectives.

Achieving accuracy would require LLMs to meaningfully apply user-specific context and data—something current models struggle with in the absence of extensive prompting or fine-tuning [13, 38]. For example, in our study, even when personal data was available, the chatbot produced hallucinations or generic outputs, reflecting broader limitations in current models’ capabilities of higher resolution outputs [15, 24]. These issues underscore the need for interaction designs, in addition to better training data, that help users understand model boundaries for more engaged and aware use [14, 30]. Further, interaction designs that facilitate iterative feedback to the model and enhance its adaptivity to user-specific contexts would mitigate accuracy challenges as well.

5.1.3 Alignment through Collaborative Design. Mehergui *et al.* argue that current human-to-AI alignment literature fails to consider or address the phenomena where the user often provides inputs based on their own mental model of how they believe the system works, expecting a specific outcome accordingly. However, it is often the case that the system does *not function* according to the user’s mental model, and thus creates an output that is not in line with the expectation [57]. Our findings align with these claims, as many neurodivergent participants experienced misaligned responses due to incorrect or absent mental models of how the LLM chatbot operated. For example, participants often assumed that the chatbot would automatically interpret and integrate profile information to provide contextually relevant advice when more explicit prompting was necessary to achieve alignment. This mismatch between user expectations and system behavior was especially problematic for participants unfamiliar with LLMs, who were unfamiliar with what extra context would be useful to redirect misaligned responses effectively. Designing systems that mitigate these mismatches is therefore crucial, particularly for neurodivergent users whose expectations, preferences, and interaction patterns often diverge from neurotypical norms.

We suggest that LLM designers address these challenges by incorporating explainability techniques that help surface the chatbot’s reasoning logic during interactions. Techniques such as chain-of-thought prompting [80] could be utilized to break down the chatbot’s recommendations into interpretable steps, allowing users to verify whether the system’s suggestions are grounded in their actual profile data or inferred assumptions. For example, when suggesting a job or crafting a cover letter, the chatbot could include a short explanation: “I recommended this based on your listed skills

in X and interest in Y.” Additionally, building on the hypothetical system introduced above, features that flag patterns in a user’s profile and invite them to approve or correct the system’s logic would foster mutual understanding. These interactive explanations would not only clarify how the chatbot is reasoning and applying user data, but would also allow users to shape outputs that reflect their unique contexts, needs, and self understanding. Further, support systems should help users assess the appropriateness of the response, especially when suggestions overstate their experience or do not match their work preferences. Prompt engineering strategies such as in-context learning [58] and example-based prompting [67] could be applied dynamically by the system to personalize suggestions for less-experienced users. For instance, when users like P1 or P17 express ambiguity or emotional overwhelm, the chatbot could draw on pre-structured, accessible templates designed for users with unclear goals, while gently guiding them toward greater specificity through scaffolded dialogue. Our study findings show that neurodivergent individuals might have atypical career paths, nonlinear work histories, or preferences for alternative forms of communication; experiences that are rarely represented in LLM training data. We suggest that supportive LLM-based systems designed for bi-directional alignment that collaboratively adapts to user preferences and lived experiences can better accommodate these nuances.

6 Conclusions

The intention of this research is to examine the interaction between neurodivergent users and a job-seeking support chatbot, using misalignment as an analytical lens to surface where and how LLMs fall short in capturing user intent, needs, and values. Through our analysis, we identified several patterns of misalignment: instances where the LLM mischaracterized users, applied user profile information in overly generalized or superficial ways, or perpetuated neurotypical job-seeking norms and language that failed to resonate with participants. Contributing to the experience of misalignment, some users were unfamiliar with the technology’s capabilities and therefore did not know what information to put into their queries or how to redirect responses that were inappropriate.

Our contribution to the HCI community is a domain-specific understanding of human-AI complementarity in job-seeking contexts for neurodivergent users. By analyzing misalignments as opportunity for intervention, we show how they can reveal the underlying values, preferences, and interaction norms that can better support this population. Rather than focusing solely on fixing the identified failure points, we advocate for bi-directional AI alignment approaches to create more reciprocal and interpretable systems. This shifts design beyond human-AI personalization as a one-way optimization and toward a model of collaborative alignment, where systems adapt to users unique needs while supporting them in understanding and shaping system behavior, fostering agency and control over outcomes.

Acknowledgments

We would like to acknowledge and thank Mentra [1], the career platform and employment network designed by and for neurodivergent individuals, for partnering with our research team to

accomplish this work. Additionally, this research was funded by (1) the Industrial Technology Innovation Program [P0028404] of the Ministry of Industry, Trade and Energy of Korea, and (2) the internal fund/grant of the Electronics and Telecommunications Research Institute (ETRI) [24RD1500/24BD1300].

References

- [1] 2024. Mentra - Neurodivergent Talent for the Modern Workforce. <https://www.mentra.com/> Accessed: 2024-09-11.
- [2] 2024. Mindshift - Empowering Neurodivergent Talent. <https://mindshift.works/> Accessed: 2024-09-11.
- [3] 2024. The Precisionists - Autism Employment Model. <https://theprecisionists.com/autism/> Accessed: 2024-09-11.
- [4] 2024. SourceAble - Inclusive Talent Solutions. <https://jobs.sourceabled.com/> Accessed: 2024-09-11.
- [5] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.
- [6] Awais Ahmed, Mengshu Hou, Rui Xi, Xiaoyang Zeng, and Syed Attique Shah. 2024. Prompt-Eng: Healthcare Prompt Engineering: Revolutionizing Healthcare Applications with Precision Prompts. In *Companion Proceedings of the ACM on Web Conference 2024*. 1329–1337.
- [7] Elizabeth A Ankrah, Francieli L Cibrian, Lucas M Silva, Arya Tavakoulnia, Jesus A Beltran, Sabrina EB Schuck, Kimberley D Lakes, and Gillian R Hayes. 2023. Me, my health, and my watch: how children with ADHD understand smartwatch health data. *ACM Transactions on Computer-Human Interaction* 30, 4 (2023), 1–25.
- [8] Zinat Ara, Amrita Ganguly, Donna Peppard, Dongjun Chung, Slobodan Vucetic, Vivian Genaro Motti, and Sungsoo Ray Hong. 2024. Collaborative Job Seeking for People with Autism: Challenges and Design Opportunities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [9] Thomas Armstrong. 2010. *The Power of Neurodiversity: Unleashing the Advantages of Your Differently Wired Brain*. Da Capo Press, Cambridge, MA.
- [10] Junseong Bang, Sineae Kim, Jang Won Nam, and Dong-Geun Yang. 2021. Ethical chatbot design for reducing negative effects of biased data and unethical conversations. In *2021 International Conference on Platform Technology and Service (PlatCon)*. IEEE, 1–5.
- [11] Howard S Becker. 1963. Outsiders. *Studies in the Sociology of Deviance* 1 (1963).
- [12] Andrew Begel, John Tang, Sean Andrist, Michael Barnett, Tony Carbary, Piali Choudhury, Edward Cutrell, Alberto Fung, Sasa Junuzovic, Daniel McDuff, et al. 2020. Lessons learned in designing AI for autistic adults. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.
- [13] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.
- [14] Reuben Binns. 2018. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency*. PMLR, 149–159.
- [15] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021).
- [16] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).
- [17] Moira Burke, Robert Kraut, and Diane Williams. 2010. Social use of computer-mediated communication by adults on the autism spectrum. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 425–434.
- [18] Buse Carik, Kaike Ping, Xiaohan Ding, and Eugenia H Rho. 2025. Exploring Large Language Models Through a Neurodivergent Lens: Use, Challenges, Community-Driven Workarounds, and Concerns. *Proceedings of the ACM on Human-Computer Interaction* 9, 1 (2025), 1–28.
- [19] Dasom Choi, Sunok Lee, Sugnin Kim, Kyungah Lee, HeeJeong Yoo, Sangsu Lee, and Hwajung Hong. 2024. Unlock Life with a Chat(GPT): Integrating Conversational AI with Large Language Models into Everyday Lives of Autistic Individuals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [20] Vedant Das Swain, Qiuyue" Joy" Zhong, Jash Rajesh Parekh, Yechan Jeon, Roy Zimmermann, Mary P Czerwinski, Jina Suh, Varun Mishra, Koustuv Saha, and Javier Hernandez. 2025. Ai on my shoulder: Supporting emotional labor in front-office roles with an llm-based empathetic coworker. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–29.
- [21] Jade Davies, Brett Heasman, Adam Livesey, Amy Walker, Elizabeth Pellicano, and Anna Remington. 2023. Access to employment: A comparison of autistic, neurodivergent and neurotypical adults' experiences of hiring processes in the United Kingdom. *Autism* 27, 6 (2023), 1746–1763.
- [22] Bobbie Eicher, Kathryn Cunningham, Sydney Peterson Marissa Gonzales, and Ashok Goel. 2017. Toward mutual theory of mind as a foundation for co-creation. In *International Conference on Computational Creativity, Co-Creation Workshop*.
- [23] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on information systems (TOIS)* 14, 3 (1996), 330–347.
- [24] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Remi Denton, and Robin Brewer. 2023. "I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 205–216.
- [25] Erving Goffman. 2023. The presentation of self in everyday life. In *Social theory re-wired*. Routledge, 450–459.
- [26] John Grant. 2008. Paid work—A valued social role that is empowering more people with an intellectual disability and providing employers with dedicated employees! *Journal of Intellectual and Developmental Disability* 33, 1 (2008), 95–97.
- [27] Sarah Griffiths, Carrie Allison, Rebecca Kenny, Rosemary Holt, Paula Smith, and Simon Baron-Cohen. 2019. The Vulnerability Experiences Quotient (VEQ): A study of vulnerability, mental health and life satisfaction in autistic adults. *Autism Research* 12, 10 (2019), 1516–1528.
- [28] Shashank Gupta, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot. 2024. Bias Runs Deep: Implicit Reasoning Biases in Persona-Assigned LLMs. arXiv:2311.04892 [cs.CL] <https://arxiv.org/abs/2311.04892>
- [29] Kaely Hall, Parth Arora, Rachel Lowy, and Jennifer G Kim. 2024. Designing for Strengths: Opportunities to Support Neurodiversity in the Workplace. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [30] Alex Hanna, Remi Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a critical race methodology in algorithmic fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 501–512.
- [31] Hwajung Hong, Jennifer G Kim, Gregory D Abowd, and Rosa I Arriaga. 2012. SocialMirror: motivating young adults with autism to practice life skills in a social world. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*. 41–42.
- [32] Amanda Hoover. 2023. For some autistic people, chatgpt is a Lifeline. <https://www.wired.com/story/for-some-autistic-people-chatgpt-is-a-lifeline/>
- [33] AMANDA HOOVER and SAMANTHA SPENGLER. 2023. For Some Autistic People, ChatGPT Is a Lifeline. <https://www.wired.com/story/for-some-autistic-people-chatgpt-is-a-lifeline/>
- [34] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems (TOIS)* 38, 3 (2020), 1–32.
- [35] Grace D Jaiyeola, Aaron Y Wong, Richard L Bryck, Caitlin Mills, and Stephen Hutt. 2025. One Size Does Not Fit All: Considerations when using Webcam-Based Eye Tracking to Models of Neurodivergent Learners' Attention and Comprehension. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*. 24–35.
- [36] JiWoong Jang, Sanika Moharana, Patrick Carrington, and Andrew Begel. 2024. "It's the only thing I can trust": Envisioning Large Language Model Use by Autistic Workers for Communication Assistance. *arXiv preprint arXiv:2403.03297* (2024).
- [37] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Kwan Yee Ng, Juntao Dai, Xuehai Pan, Aidan O'Gara, Yingshan Lei, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. 2024. AI Alignment: A Comprehensive Survey. arXiv:2310.19852 [cs.AI] <https://arxiv.org/abs/2310.19852>
- [38] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* 55, 12 (mar 2023), 1–38. doi:10.1145/3571730
- [39] Eunkyung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [40] Ruth Kanfer, Connie R Wanberg, and Tracy M Kantrowitz. 2001. Job search and employment: A personality-motivational analysis and meta-analytic review. *Journal of Applied psychology* 86, 5 (2001), 837.
- [41] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.
- [42] Bogoan Kim, Dayoung Jeong, Jennifer G Kim, Hwajung Hong, and Kyungsik Han. 2023. V-DAT (Virtual Reality Data Analysis Tool): Supporting Self-Awareness for Autistic People from Multimodal VR Sensor Data. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [43] Sung-In Kim, So-young Jang, Taewan Kim, Bogoan Kim, Dayoung Jeong, Taehyung Noh, Mingon Jeong, Kaely Hall, Meelim Kim, Hee Jeong Yoo, et al. 2024. Promoting self-efficacy of individuals with autism in practicing social skills in the workplace using virtual reality and physiological sensors: Mixed methods

- study. *JMIR Formative Research* 8 (2024), e52157.
- [44] Mara Kirdani-Ryan and Amy J Ko. 2024. Neurodivergent Legitimacy in Computing Spaces. *ACM Transactions on Computing Education* (2024).
- [45] Ha-Kyung Kong, Derek Xie, Ankhith Chandra, Rachel Lowy, Arielle F Maignan, Sehoon Ha, Chung Hyuk Park, and Jennifer G Kim. 2024. Co-designing Robot Dogs with and for Neurodivergent Individuals: Opportunities and Challenges. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.
- [46] Ha-Kyung Kong, Saloni Yadav, Rachel Lowy, Daniella Rose Ruzinov, and Jennifer G Kim. 2024. Understanding Online Job and Housing Search Practices of Neurodiverse Young Adults to Support Their Independence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.
- [47] Ha-Kyung Kong, Saloni Yadav, Rachel Lowy, Daniella Rose Ruzinov, and Jennifer G Kim. 2024. Understanding Online Job and Housing Search Practices of Neurodiverse Young Adults to Support Their Independence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [48] Minha Lee, Sander Ackermans, Nena Van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselstein. 2019. Caring for Vincent: a chatbot for self-compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [49] Messi HJ Lee, Jacob M Montgomery, and Calvin K Lai. 2024. Large language models portray socially subordinate groups as more homogeneous, consistent with a bias observed in humans. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1321–1340.
- [50] Xinyu Li, Zachary C. Lipton, and Liu Leqi. 2024. Personalized Language Modeling from Personalized Human Feedback. arXiv:2402.05133 [cs.CL] <https://arxiv.org/abs/2402.05133>
- [51] Karen PY Liu, Denys Wong, Anthony CY Chung, Natalie Kwok, Madeleine KY Lam, Cheri MC Yuen, Karen Arblaster, and Aldous CS Kwan. 2013. Effectiveness of a workplace training programme in improving social, communication and emotional skills for adults with autism and intellectual disability in Hong Kong—a pilot study. *Occupational Therapy International* 20, 4 (2013), 198–204.
- [52] Beatriz López and Liz Keenan. 2014. Barriers to employment in autism: Future challenges to implementing the Adult Autism Strategy. *Autism Research Network* (2014), 1–17.
- [53] Rachel Lowy, Lan Gao, Kaelly Hall, and Jennifer G Kim. 2023. Toward Inclusive Mindsets: Design Opportunities to Represent Neurodivergent Work Experiences to Neurotypical Co-Workers in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [54] Rachel Lowy, Chung-Eun Lee, Gregory Abowd, and Jennifer Kim. 2023. Building Causal Agency in Autistic Students through Iterative Reflection in Collaborative Transition Planning. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2023).
- [55] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.
- [56] David Mason, Helen McConachie, Deborah Garland, Alex Petrou, Jacqui Rodgers, and Jeremy R Parr. 2018. Predictors of quality of life for autistic adults. *Autism Research* 11, 8 (2018), 1138–1147.
- [57] Malek Mechergui and Sarath Sreedharan. 2024. Goal Alignment: Re-analyzing Value Alignment Problems Using Human-Aware AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 10110–10118.
- [58] Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? arXiv:2202.12837 [cs.CL] <https://arxiv.org/abs/2202.12837>
- [59] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
- [60] Selin Nugent, Paul Jackson, Susan Scott-Parker, James Partridge, Rebecca Raper, Chara Bakalis, Alex Shepherd, Arijit Mitra, Jintao Long, Kevin Maynard, et al. 2020. Recruitment AI has a Disability Problem: Questions employers should be asking to ensure fairness in recruitment. (2020).
- [61] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL] <https://arxiv.org/abs/2203.02155>
- [62] Christopher Peters. 2005. Foundations of an agent theory of mind model for conversation initiation in virtual environments. *Virtual Social Agents* 163 (2005).
- [63] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1, 4 (1978), 515–526.
- [64] Jianing Qiu, Lin Li, Jiankai Sun, Jiachuan Peng, Peilun Shi, Ruiyang Zhang, Yinzhao Dong, Kyle Lam, Frank P-W Lo, Bo Xiao, et al. 2023. Large ai models in health informatics: Applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics* (2023).
- [65] D Reeve. 2011. Ableism within disability studies: The myth of the reliable and contained body. In *Theorising Normalcy and the Mundane, 2nd International Conference*.
- [66] Anne M Roux, Paul T Shattuck, Jessica E Rast, Julianna A Rava, Amy D Edwards, Xin Wei, Mary McCracken, and Jennifer W Yu. 2015. Characteristics of Two-Year College Students on the Autism Spectrum and Their Support Services Experiences. *Autism Research and Treatment* 2015, 1 (2015), 391693.
- [67] Ohad Rubin, Jonathan Herzig, and Jonathan Berant. 2022. Learning To Retrieve Prompts for In-Context Learning. arXiv:2112.08633 [cs.CL] <https://arxiv.org/abs/2112.08633>
- [68] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv:2402.07927 [cs.AI] <https://arxiv.org/abs/2402.07927>
- [69] Alan M Saks and Blake E Ashforth. 1999. Effects of individual differences and job search behaviors on the employment status of recent university graduates. *Journal of Vocational behavior* 54, 2 (1999), 335–349.
- [70] Jennifer Sarrett. 2017. Interviews, disclosures, and misperceptions: Autistic adults' perspectives on employment related challenges. *Disability Studies Quarterly* 37, 2 (2017).
- [71] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, et al. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. arXiv preprint arXiv:2406.09264 (2024).
- [72] Lucas M Silva, Francieli I Cibrian, Clarisse Bonang, Arpita Bhattacharya, Aehong Min, Elissa M Monteiro, Jesus Armando Beltran, Sabrina Schuck, Kimberley D Lakes, Gillian R Hayes, et al. 2024. Co-Designing situated displays for family co-regulation with ADHD children. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [73] Katta Spiel, Christopher Frauenberger, Os Keyes, and Geraldine Fitzpatrick. 2019. Agency of autistic children in technology research—A critical literature review. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 6 (2019), 1–40.
- [74] Laura Spillner and Nina Wenig. 2021. Talk to Me on My Level – Linguistic Alignment for Chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (Toulouse & Virtual, France) (MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 45, 12 pages. doi:10.1145/3447526.3472050
- [75] Michael Terry, Chinmay Kulkarni, Martin Wattenberg, Lucas Dixon, and Meredith Ringel Morris. 2023. AI Alignment in the Design of Interactive AI: Specification Alignment, Process Alignment, and Evaluation Support. arXiv:2311.00710 [cs.HC] <https://arxiv.org/abs/2311.00710>
- [76] Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. Two Tales of Persona in LLMs: A Survey of Role-Playing and Personalization. arXiv:2406.01171 [cs.CL] <https://arxiv.org/abs/2406.01171>
- [77] John C Turner, Rupert J Brown, and Henri Tajfel. 1979. Social comparison and group interest in ingroup favouritism. *European journal of social psychology* 9, 2 (1979), 187–204.
- [78] Karen D Wang, Jake McCool, and Carl Wieman. 2024. Exploring the learning experiences of neurodivergent college students in STEM courses. *Journal of Research in Special Educational Needs* 24, 3 (2024), 505–518.
- [79] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.
- [80] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf
- [81] Till Winkler and Sarah Spiekermann. 2021. Twenty years of value sensitive design: a review of methodological practices in VSD projects. *Ethics and Information Technology* 23 (2021), 17–21.
- [82] Marisol Wong-Villacres, Carl DiSalvo, Neha Kumar, and Betsy DiSalvo. 2020. Culture in action: Unpacking capacities to inform assets-based design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [83] Jiacheng Ye, Zhiyong Wu, Jiangtao Feng, Tao Yu, and Lingpeng Kong. 2023. Compositional Exemplars for In-context Learning. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 39818–39833. <https://proceedings.mlr.press/v202/ye23c.html>