"Will This Tool Ever Push Back or Challenge Me?": Reflections on a Multi-agent LLM Tool for Perspective Seeking

Krishna Akhil Kumar Adavi*
School of Information
The University of Texas at Austin
Austin, USA
kaka127@utexas.edu

Pratik Ghosh
Microsoft Research
Cambridge, United Kingdom
pratik.ghosh@microsoft.com

Richard Banks
Microsoft Research
Cambridge, United Kingdom
rbanks@microsoft.com

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Siân E. Lindley
Microsoft Research
Cambridge, United Kingdom
sianl@microsoft.com

Abstract

The ability of LLM-based agents to role-play may be especially helpful for gathering multiple perspectives in problem-solving situations. To understand how workers would respond to getting perspectives from agents, we conducted an interview study with 18 knowledge workers using a generative AI-powered multi-agent design probe. Participants described their current practices of perspective seeking from people, and outlined challenges with this practice including not knowing the right person to speak to and anxiety around reaching out. Participants reported finding value in the probe when it produced responses that aligned with their current solutions, and wanted sources or explanations for solutions they were unfamiliar with. They especially highlighted the need to be challenged by the probe, and were worried that the availability of such tools would reduce their interactions with colleagues. We discuss implications for designing for challenge, and locate agents as preparation tools for interacting with colleagues.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Computer supported cooperative work**; **Interaction design theory, concepts and paradigms**.

Keywords

Knowledge Sharing, Help Seeking, Perspective Seeking, Advice Seeking, Information Seeking, Multi-agent LLMs

ACM Reference Format:
Krishna Akhil Kumar Adavi, Pratik Ghosh, Richard Banks, Advait Sarkar, and Siân E. Lindley. 2026. "Will This Tool Ever Push Back or Challenge Me?": Reflections on a Multi-agent LLM Tool for Perspective Seeking. In *Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '26), June 22–25, 2026, Linz, Austria.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3808045.3808058

*The work was done when the co-author was employed at Microsoft Research

CHIWORK '26, Linz, Austria

ACM ISBN 979-8-4007-2429-9/26/06
https://doi.org/10.1145/3808045.3808058

1 Introduction

Knowledge workers' [7] everyday tasks often require them to seek input from their colleagues. A software developer might need to better understand the codebase they are working on before pushing a code update; a compliance manager needs to get inputs from members working in engineering, design, and data science as they are drafting a legal contract; a product manager needs to coordinate with marketing, design, and engineering to release the latest feature. In each of these examples, workers reach out to and get other workers' perspectives to help them progress in a task. This process is valuable because a given worker might not themselves have the expertise to complete a particular task, and would likely benefit or depend upon getting additional input or perspectives.

This sort of collaborative work includes help seeking, advice seeking, feedback seeking, and knowledge sharing [1, 35], detailed in the Related Work (Section 2). In this paper, we focus on the importance of perspective in these activities, or how the person providing help views the task or problem. In essence, "perspective seeking" occurs when workers reach out to their colleagues to understand, and then potentially incorporate into their own work, how those colleagues approach or view a problem or activity. Indeed, collaboration with people inevitably means that their contributions will be offered from a particular perspective, based on their experience, knowledge, skills, and relationship to the help-seeker. In contrast, workers have recently had the opportunity to incorporate Generative AI (GenAI) tools into their work, to get advice or to seek a different perspective [8, 65, 77]. Yet, we lack sufficient empirical work on how workers see GenAI-based tools fitting into the practice of perspective seeking. To better understand and support this practice, we conducted an interview study using a GenAI-powered multi-agent design probe with 18 knowledge workers at a global technology company. Our guiding research questions were:

RQ1) What are knowledge workers' current practices of perspective seeking while solving problems during their work?

RQ2) How do knowledge workers see Generative AI-based tools supporting perspective seeking?

Our study found that knowledge workers rely on their colleagues' perspectives when approaching problems at work – this might be reaching out to colleagues based on their role, level of seniority, or perceived thinking and working style. Workers trust their colleagues' perspectives because of shared organizational contexts,

prior experiences of working with them, and their openness towards offering help. Challenges with current practices include being unsure who to reach out to, having the social capital to get connected to the right person, overcoming the anxiety to reach out to people, a lack of time – coordinating to be in the same time zone, people's busyness, and feeling anxious around appearing unprepared or not asking the right questions. However, perspective seeking is important; participants reported finding particular value in engaging with perspectives that challenged, or pushed their thinking around a problem.

Use of the design probe highlighted the importance of organizational context in being able to place trust in "perspectives" offered by GenAI. Participants trusted the tool's suggestions when these aligned with solutions they were already considering; when it produced outputs they weren't familiar with, they wanted sources, or explanations for those outputs. Thus, in contrast to the challenge that was reported to be valuable when offered by human collaborators, contrasting views were not accepted as offering useful or trusted pushback when presented by GenAI. Nevertheless, participants highlighted a need to be challenged by the tool, especially in problem-solving situations where they wanted to improve their thinking. Our study suggests this should complement rather than reduce perspective seeking from people. Participants expressed an anxiety that such tools could reduce interactions between colleagues, and imagined using such tools to be better prepared for interactions with their co-workers.

This paper makes the following key contributions:

a) We surface challenges with current practices of perspective seeking in organizational contexts, and outline a role for LLMs as a preparation tool to support this practice.

b) We provide empirical evidence of users' needs for challenge or pushback from LLMs, and outline design implications for future LLM research in this direction.

c) We discuss the potential impacts of the deployment of AI tools on workplace culture and team dynamics.

2 Related Work

Recent scholarship on the use of LLMs in enterprises has identified different tasks for which knowledge workers engage LLMs, including content creation, information management (searching or summarizing), advice, i.e., improving or validating their work, and collaborating with others [8, 65]. In this related work section, we begin by stepping back and considering how people have gone about gathering information and sharing knowledge in their workplaces. First, we consider scholarship from organizational science which identifies the vital role that people play in knowledge sharing at organizations, the challenges with this practice, how HCI and CSCW interventions have aimed to support this practice, and, in turn, the challenges these interventions have run into. Then, we consider how GenAI-based tools currently support ideation for problem solving, and note how they might play a role in supporting knowledge sharing by being more usefully grounded in context, and taking on different roles.

2.1 Knowledge Sharing in Organizations

Lim et al. [35] describe four kinds of resource seeking behaviors in organizational settings: feedback seeking (on one's performance); information seeking (to aid sensemaking about one's workplace); advice seeking (to form an opinion on a problem); and help seeking (emotional or instrumental assistance on work or non-work problems) (p. 125). Each of these behaviors can be seen as a form of knowledge sharing practice within organizations.

Research from organization science and CSCW has demonstrated how people are key to accessing, understanding, and interpreting organizational knowledge because the right person has a rich comprehension of context and information needs and can convey information in an appropriate and understandable way [2, 35, 40, 41]. However, people face barriers to knowledge sharing, including anxieties around reaching out to others, being unsure who to reach out to, and a lack of time [4, 30, 69]. Even when workers are able to reach a relevant person, they might run into issues with those colleagues lacking incentives or motivation to share knowledge [25, 75, 76, 78]. In a historical overview, Ackerman et al. [1] describe two generations of efforts within HCI and CSCW to support the practice of knowledge sharing. These are knowledge bases (externalized representations of information) and expert recommender systems (enabling discovery). They note how knowledge base efforts have run into certain limitations because they typically underplay the role that people take in facilitating knowledge sharing. The second generation, recommender systems, faced challenges including how to help workers understand how the system represented and recommended experts, the additional labor placed on those experts, and ethical concerns around how precisely who is represented as an expert within an organization is determined [33]. Advice seeking and help seeking can be quite different from information seeking but context remains key. For example, people may be sought to offer a neutral perspective [36].

Thus, for any resource seeking activity to be successful, it would need to be tailored to the recipient and their context. The role of the information and advice provider is inherent to how this can be accomplished in collaborative work [32]. LLMs have the potential to support knowledge sharing and advice seeking, and can also be prompted to offer inputs that adopt a certain perspective or role [6]. However, it is not clear how well this capability can support resource-seeking behaviors that arise as part of work. In this study, we specifically consider how workers respond to getting different role-based perspectives from a multi-agent LLM tool on a current problem they are working on, and how they see the tool as playing a role amidst organizational complexities.

2.2 LLMs in Ideation, and Persona-based Role-playing

An important aspect of perspective seeking can be getting multiple perspectives. Extensive research has highlighted the importance of getting multiple diverse perspectives on problem solving and its relation to improved outcomes [26, 38, 63].

There are possibilities to support this using GenAI, by having multiple LLMs work together. Research has looked at how multiple LLMs can work with each other to improve system performance when completing complex tasks as opposed to a single LLM (e.g.,

[12, 19, 46, 47, 74]). These efforts have largely focused on multiple LLMs collaborating to automate tasks. Applications where multi-agent LLMs work with end users are fewer in number, but have been attempted in domains such as information seeking, through for example, multi-agent debates [62]. Here, Shi et al. [62] found that when users searched for information on controversial issues and were exposed to diverse perspectives through a multi-agent debate, users' confirmation bias reduced.

There has been a substantial interest in leveraging LLMs' ability to role-play and mimic different personas and facilitate more human-like interactions [13, 29, 43, 59, 64, 68]. In a survey of the field, Chen et al. [13] describe the personas in these interventions as falling into one of three categories: "1) demographic persona, which leverages statistical stereotypes; 2) character persona, focused on well-established figures; and 3) individualized persona, customized through ongoing user interactions for personalized services" (p. 1). Examples of these interventions include using GenAI produced demographic personas in user research [29, 59], and commercial efforts like Character.ai [28] which allows users to build AI characters and converse with them. Scholars have noted how persona-based simulations can result in bias and produce flattened responses [14, 15].

Researchers have also explored how GenAI can assist in creativity during ideation and brainstorming [24, 60, 67, 70, 72]. These endeavors have yielded important insights around design fixation, which can occur when images are generated by GenAI during ideation [72], and participants expressing concerns around losing out on human perspectives [24]. These interventions have typically involved users, whether individuals or groups, getting assistance from a single agent. More work is needed to understand how multiple agents could support perspective seeking at work, and the risks of doing so.

Taken together, attributes like the ability to take on a certain role (e.g., a product manager), and working together with other agents (e.g., a product manager agent with a designer agent) may be particularly valuable in getting multiple perspectives during problem-solving situations. To explore this further, we designed and conducted a study examining how knowledge workers seek perspectives from and respond to multiple role-based personas working together to assist in problem solving.

3 Methods

3.1 Participant Recruitment and Interview Process

To conduct this interview and design probe-based study, we recruited participants via internal company email lists at a global technology company. The study protocol, design probe, and recruitment materials were all approved by our Institutional Review Board. We primarily reached out to email lists of employees based in the United Kingdom. We received a total of 44 responses. From this, we purposively sampled 18 participants with an eye on capturing a variety of work functions (e.g., software engineering, data science, design, compliance), age ranges, and seniority levels. Details of the participants are listed in Table 1. All interviews were conducted via Microsoft Teams between July and August 2024. Participants consented for audio and video recording of the interviews, and for a screen recording of their interactions with the design probe. Interviews lasted between 74 and 88 minutes, with an average length of 83 minutes. For their participation in the study, each participant was compensated with £30 and was given a text transcript of their interaction with the design probe.

3.2 Interview Protocol

Our interview protocol consisted of three parts: in the first part, participants discussed in-depth a current or recent problem they were working on which necessitated multiple or different perspectives. They discussed their process of getting these perspectives, the challenges in getting them, how they determined the credibility or usefulness of these perspectives, and the benefit they saw in getting these perspectives. Participants then discussed the current ways in which they use LLMs in their workplace. Next, they proceeded to interact with the LLM-powered multi-agent design probe. During this part of the interview, participants inputted a current or recent problem they were working on to get multiple perspectives from the probe, while talking aloud and reflecting on each step they were taking, as well as commenting on the responses being generated through the interaction with the probe. Finally, participants reflected on the experience of seeking perspectives from the tool, and how they saw this fitting into their practices. The interview protocol has been provided in Appendix B.1.

3.3 Description of Design Probe

The design probe [27, 56, 73] enabled us to have a grounded conversation with study participants about their experience of seeking perspectives from GenAI-based agents. Perspective seeking at workplaces can occur in one-on-one conversations with others, as well as in larger groups, for example, a project team meeting. In this study, we were specifically interested in understanding participant responses to getting multiple perspectives at the same time, and this informed our choice of the multi-agent setup [20].

Drawing on foundational scholarship in conversation analysis, we adapted the conversation turn-taking model proposed by Sacks et al. [48] to simulate conversations with AI agents. We built the probe consisting of four agents based on GPT-4 [42], namely: a Designer, an ML Researcher (Machine Learning Researcher), an Engineer and a summarizing agent called "Sage." We chose the three agents to best represent the make-up of a team working on software research, design, and development because of our awareness of the kinds of problems that come up in the organizational context of the study, i.e., a technology company.

Through system messages [39], we provided instructions to each of the agents to provide responses from the role it was assigned (Designer, Engineer, ML Researcher, Sage). The agents were given the guidance to build on each other's responses, effectively giving the conversation a "yes, and" orientation. Detailed system messages are provided in the Appendix A.1.

Recent literature has highlighted how LLM responses tend to be verbose and overgeneralized [44, 49]. To tackle the problem of overgeneralization, agents were instructed to ask the participant clarifying questions when they needed context. For a turn-taking system to work effectively and appear like a natural conversation, each agent needed to overcome the tendency to be verbose and

Table 1: Participant Table

Participant	Gender	Age	Work Function	Job Level
P1	Woman	30-39	Compliance and Governance	Senior
P2	Woman	20-29	Design and User Research	Early Career
P3	Woman	30-39	Product Management	Senior
P4	Woman	40-49	Advisory	Senior
P5	Man	20-29	Software Engineering	Early Career
P6	Woman	30-39	Data Science and Machine Learning	Principal
P7	Woman	30-39	Software Engineering	Senior
P8	Woman	20-29	Sales and Marketing	Intern
P9	Woman	40-49	Data Science and Machine Learning	Senior
P10	Woman	30-39	Data Science and Machine Learning	Resident
P11	Woman	20-29	Software Engineering	Early Career
P12	Woman	30-39	Human Computer Interaction Research	Principal
P13	Woman	40-49	Design and User Research	Senior
P14	Man	30-39	Software Engineering	Senior
P15	Woman	20-29	Product Management	Early Career
P16	Man	30-39	Design and User Research	Principal
P17	Man	30-39	Design and User Research	Principal
P18	Man	30-39	Data Science and Machine Learning	Senior

allow other agents, and the user, to contribute to the discussion. To navigate this challenge, we developed a confidence-based mechanism (detailed in Appendix A.2) by which agents took turns to generate responses. When the agents did respond, we limited their responses by using explicit constraints such as word length. See an example interaction showing topic, intent, speaker nomination, and respective confidence scores in Appendix A.3.

For the front-end of the design probe, we developed a web app using an open-source Python library [23]. See Figure 1 for a screenshot of the design probe. Under "Initial Problem Statement," study participants defined a problem to begin the conversation. The "Conversation" chat panel displayed all messages sequentially. Each time an agent spoke, their representative image would be displayed in an adjacent "Speaker" panel. The "User Comments" panel allowed the participant to communicate with the agents as in a conventional chat interface, where they could ask more questions, or respond to clarifications requested by the agents. We also implemented a "Debug" panel which showed the topic, intent, and the agent nominated to speak for each message as it was generated. The panel also showed the agents' respective confidence scores. At the end of their interaction, participants could click the "Download Conversation Log" button to save their conversations.

3.4 Data Analysis

To prepare the transcripts for analysis, we cleaned and anonymized the automated transcripts generated from the Microsoft Teams interview recordings. We followed an inductive, bottom-up coding process for analysis of the interview data [9]. After the interview process, one team member developed an initial codebook based on a rolling memo written after each interview and a review of the interview transcripts. This codebook was vetted by two other team members to fine-tune the code definitions. The final codebook included codes such as *Value of Perspective Seeking*, *Challenges in Perspective Seeking*, *Tool Experience and Interaction*, and *Imagined Tool Use*, and has been included in Appendix B.2. The first author then coded the 18 interview transcripts using ATLAS.ti [21] qualitative coding software. Two other authors reviewed the code snippets to verify that the codes were being applied consistently. Through discussions within the research team, we decided to prioritize findings around the participants' descriptions and understanding of their work practices, and their perceptions of their interactions with the design probe, and how they see GenAI-based tools fitting into their work practices because this helped us address the research questions for the project. The Findings section reports on the themes generated by referring to memos generated at different stages of the analysis process, and comparing snippets both within and across the codes.

4 Findings

We first discuss our participants' descriptions of the kinds of perspectives they seek while solving problems during their work; their process of seeking perspectives – why and how they do it; the challenges involved in getting these perspectives; and how, increasingly, GenAI-based tools play a part in this process. Then, we share findings from our participants' interactions with the design probe: we describe the challenges participants reported in getting perspectives from the probe; how they came to trust the responses from the tool; and the risks they saw in getting perspectives from a tool. Finally, we describe how participants saw this tool fitting into their work practices.

4.1 Current Practices of Perspective Seeking

4.1.1 Problems, Characteristics of Perspectives, Benefits. Our participants described their needs for perspectives on problems they encounter during work in a variety of collaborative situations. These ranged from the early stages of a project, to scope down the problem

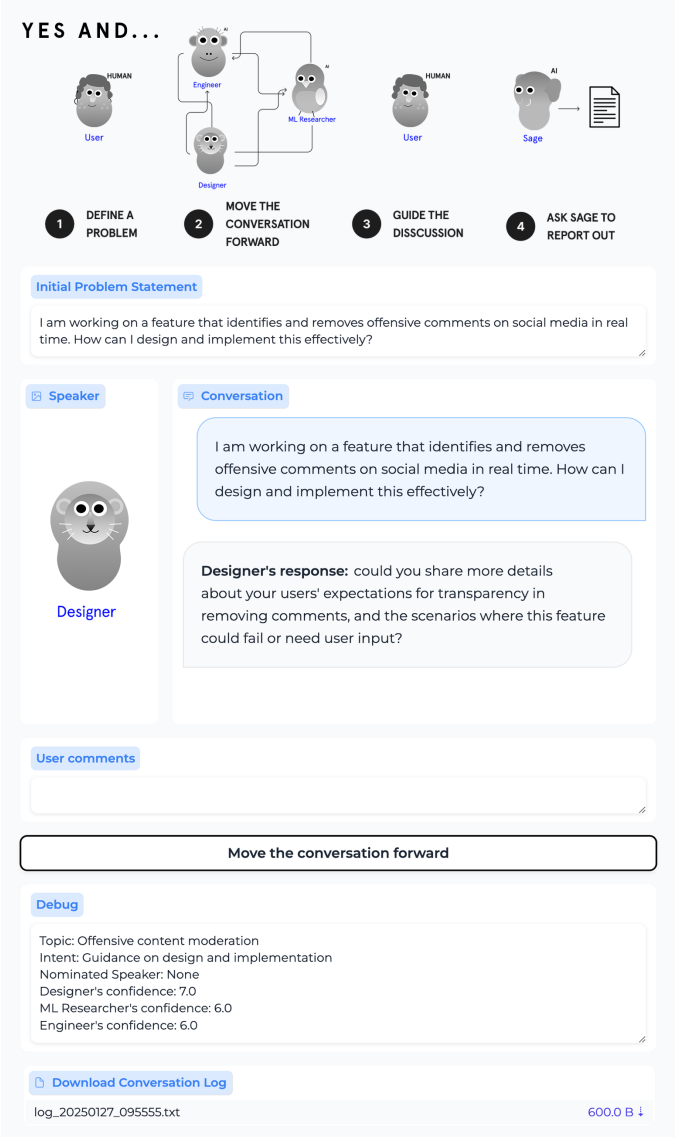

Figure 1: Screenshot of the Design Probe

at hand, towards intermediate stages, when they have produced a certain artifact (a presentation, or a design mock-up for review), to the end stages, to get consensus or feedback on a team's decision. A common theme across these problems was that workers saw these problems as open-ended. Gathering different perspectives was seen as a routine activity during work. In effect, it is a step towards knowledge seeking in the context of a larger task. For example, P15, a product manager, was working on a problem to understand how to surface relevant emails in users' inboxes, and highlighted how seeking different perspectives helped her better consider proposed solutions:

"It is very refreshing to see how everyone is thinking about the same problem. Over time, I've learned to accept that I can't know everything and I can't think of everything. These are teammates that are more experienced in data science or engineering or whatever role they're doing, so it's important to get their perspective because I come from a place of the user and the problem. But maybe there's emails we actually can't filter out because we don't have the signals available. My engineer would tell me that. So, it is important to get their opinions and perspectives."

Overall, our participants described different characteristics of the kinds of perspectives they sought out: role-based (wanting an engineering or design-based perspective), hierarchy-based (wanting a senior, or junior worker's perspective), experience-based (someone that has gone through a similar situation before), disciplinary training-based (industrial engineering, or HCI), demographics-based (speaking to someone from a particular sociocultural background,

age, gender, race), or working style-based (someone who is a planner, or a doer, for example). In effect, when someone is seeking a perspective from another person, their expectation is that the perspective they are receiving is shaped by some dynamic combination of these categories. It is important to consider that the perspective being sought, or provided, is not fixed, but is situation- and query-dependent.

Participants described relying on their social connections to get perspectives during their work. Depending on the kind of problem, this meant relying on colleagues within their team (e.g., P3's team has to decide which set of features to prioritize building next), or on reaching out to others within the organization by leveraging their connections (e.g., P7 reaching out to a developer within their organization who has used the API that they are currently struggling with).

Our data suggest three ways in which perspective seeking can be perceived as useful. First, participants saw perspective seeking as a way to improve their knowledge, and relatedly, complete their tasks. Second, participants, especially those in more junior roles, saw getting perspectives as a way to build their network within their organization, as was the case for P5, a junior software developer: *"[...] building a relationship with other people? Sometimes this sort of process [of seeking perspectives on problems] has led me to reaching out to people that maybe I've never directly interacted with, so that's always positive."* Third, participants saw it as a way to get social buy-in around their work, and to sustain the social life of being at work. P1, who works as a compliance manager, was responsible for drafting up a contract outlining the risks of the data agreement between her organization, and a healthcare partner:

"I needed different perspectives because people come at it from a different place when thinking through 'risk.' I might not think of something that someone from engineering, or legal would think about. But also to help with adoption of the outcome. It's really important to bring people on board with what you are doing and getting the different perspectives on this before we actually say 'this is how it's going to be,' makes it more likely that they will follow it in the long run."

Participants reported having a strong implicit trust in their colleagues even if they did not know them personally because, often, they are pointed to that colleague by someone they do know. In cases where there is a personal connection, the relationship history comes to the fore. Participants described relying on their colleagues' expertise and understanding of them as being good collaborators: *"I know my colleagues, I know the track record. I know what they've designed,"* P17. Participants especially reported valuing perspectives from colleagues that they perceived as being different from them; for example, a person that perceives themselves as an overplanner valuing an executor, or a straight-thinker valuing a broad/open-minded thinker, or particular value being found in a colleague that challenges their ideas. Participants emphasized that the other person's social attributes, such as kindness and helpfulness, were important in building that trust. This isn't to suggest people don't want colleagues who validate them and agree with them, but they also want to be challenged as a part of perspective seeking. P12, an HCI researcher, captured this from her experience of writing an academic paper with a group of interdisciplinary researchers:

"Being critical and being challenged and getting feedback is essential. It's the kind of feedback you need to help you learn and progress. If I wasn't challenged by other people, that would be hugely missing out on growing in your understanding because being challenged is not an attack, right? Being challenged is just like, hey, have you thought maybe about these other ways or why are you so certain about certain things?"

4.1.2 Challenges in Perspective Seeking. Our participants highlighted several challenges with the process of getting different perspectives at their workplace.

Social connection and anxiety around reaching people: Junior employees like P5 shared how they did not have a large network within their organization and, consequently, knowing who to reach out to was difficult for them. Even if they were able to, they described being worried about how their questions would be perceived by others: *"I'm very junior, so it's hard to know who to reach out to [...] maybe being worried about asking dumb questions."* In effect, there are challenges with knowing whom to contact, but also around how to structure conversations.

Coordinating time: P2, a designer, noted how getting time and availability from her colleagues was difficult: *"hard to get everyone to come together at one time."* A few participants (P4, P16, P17) described being in remote roles where their team members were in very different time zones, making it difficult to talk with them instantly. P4, who works in an advisory role, noted this as a common problem when she has to prepare for her customer calls: *"My customer calls are in the mornings, UK time. My team in the US is usually not awake until 3:00 PM."*

Lack of role diversity in their team: Even if one felt comfortable reaching out to others, some participants, like software engineer P11, highlighted how she struggles with getting perspectives from people in other roles like design because *"my team only has one designer."*

Over-reliance on people one is familiar with: Participants like P3, a product manager, highlighted the concern that workers are possibly only seeking perspectives from people they have a relationship with: *"I naturally gravitate towards the people I already have in relationship with, so the people who I'm most friendly with on the team as well as my manager. I didn't make a conscious effort to try and speak to as big of a variety of people as possible, because I maybe just spoke to people I have a natural affinity towards."*

Not being challenged or pushed in one's thinking: Participants also felt that there was a risk in seeking perspectives that aligned too closely with their own, because this meant that they weren't being pushed or challenged on their views. P10 captured this feeling: *"I just get the feeling that it is normal to think this way and get this sense of validation, but I don't really necessarily grow because I'm not gaining insight."*

4.1.3 Using Generative AI Tools in the Workplace. Every participant in our study discussed using GenAI tools in the workplace for different tasks. The three largest use cases reported are writing assistance for emails or internal company memos (11/18), coding assistance for short snippets and code explanations (10/18), and research (10/18).

When we asked participants if they ever considered their interactions with LLMs as getting multiple or different perspectives,

Table 2: Characteristics of Perspectives Sought

Perspective Characteristics	Example Quotes
Role-based	*"People in different roles. So I'm a product manager, but I could be working with my [data] scientists or some of the engineers."* (P15)
Disciplinary Training-based	*"I'm a data scientist and have background in information retrieval. And my colleague is from computer vision."* (P18)
Hierarchy-based	*"Talk with people that have higher seniority, more experience than me? They have more knowledge about different aspects of the system"* (P5)
Demographics-based	*"If you're making a product, you need to think about accessibility, gender, and race [...]"* (P8); *"family backgrounds, nationality and ethnicity"* (P14)
Working style-based	*"So the attention to detail, and the desire to pre-plan and pre-think about things."* (P6); *"A person that tends to focus pretty quickly on the core concept? Indifferent to presentation."* (P16)
Experience-based	*"I think it's helpful to have senior engineers, if they've worked on something to talk about that experience. As another engineer, how they found that?"* (P11)

most participants said they had never considered LLMs as having a perspective; that a perspective was linked to a person and drawn upon their experience. Some participants reported explicitly prompting tools to take on a role-based persona (e.g., product manager or banker) and generate a response because they perceived this prompting strategy to improve the quality of the responses from the LLM. Some participants said that when they were using GitHub Copilot as a *"pair programmer"* (P11) or *"friendly little coder"* (P7), ChatGPT as a *"thought companion"* (P18), or Copilot as a *"brainstorming partner"* (P4) or *"writing assistant"* (P12), they were in a sense asking the tool "what it thinks should be done here?". A key challenge with considering that LLM-based tools can provide a perspective is being unsure of where the tool is coming from. This was captured in our interaction with data scientist P9:

"If the LLM has outputted something that looks like a perspective? Obviously, that's influenced by the training data and by the prompts that I cannot see. Because I don't know enough about the training data or the prompts, I can't really judge how I feel about that perspective. I can't judge what it wants or what it values? With my colleagues I make certain assumptions about what they want, and I use information about what they've worked on to think about what they value? I assume that they have good intent [...] I make all of those assumptions about humans. An LLM might say something that looks like a perspective, but I can't put it into the context of what kind of a person would have this perspective."

P2 especially highlighted this while noting that she wouldn't use LLMs for researching topics like race or gender: *"When it comes to like ethnic minorities or gender and sexuality, that's not something I'm coming to a bot for. If we're talking about something very specific like race, I feel like often you're looking for a really nuanced perspective. I just struggle to believe and trust that a bot will be able to do for me. I would much rather speak to people. Even hearing the tone in how they talk about their experiences when you're talking about personal topics like that is necessary as opposed to just spitting up plain information with nothing really behind it."*

4.2 Responses to Using the Design Probe

In this section, we report on findings from the participants' interaction with the design probe. As discussed earlier, the design probe is set up for participants to input a problem statement to initiate the conversation. To start the session, based on their confidence score (detailed in Appendix A.2), one of three agents, a designer agent, an ML Researcher agent, or an engineer agent, provides their perspective on the problem, and subsequent responses build on existing agent and participant responses. At any stage, participants could explicitly hail an agent if they wanted that particular role-based perspective, or could add additional information, constraints, or change the direction of the conversation. Every participant entered a problem of their choosing into the system. At the end of the interaction, participants could hail a "sage" agent who summarized the entire conversation with a plan for moving forward. Every participant was provided a text transcript of their problem-solving interaction with the probe.

4.2.1 *How do I Prompt? How do I Calibrate the Conversation?*

Nearly every participant struggled with understanding at what level of specificity or generality to prompt the probe from the outset. Because the agents would begin providing a response to the initial prompt immediately, participants' responses clustered into two camps: either the initial responses felt too generic, or the system jumped too quickly to a specific solution. Several participants intervened in the conversation to push the system to provide a more calibrated response to their query. We observed participants either blaming themselves for not providing enough information in the initial prompt, or noting that it could help to know what kind of data and context the tool is trained on, so they can prompt appropriately. For some participants like P7, a software engineer working through an issue around deploying a feature, it was important that the tool was asking her questions, and making her feel like a solution was being co-constructed: *"At the start, I hadn't provided enough information or context about the problem. I'm happy that it's now asking me to help shape it, that's nice to see."*

P13 imagined that the initial problem stage itself could be a site to better establish the problem: *"I found maybe that in the initial step, you know where you write a problem statement. If it could be a little bit expanded to tease me with more details? So that the agents have more information and context from the get go, so they know who I am, why I'm asking it [...] just to be prepared for the conversation."*

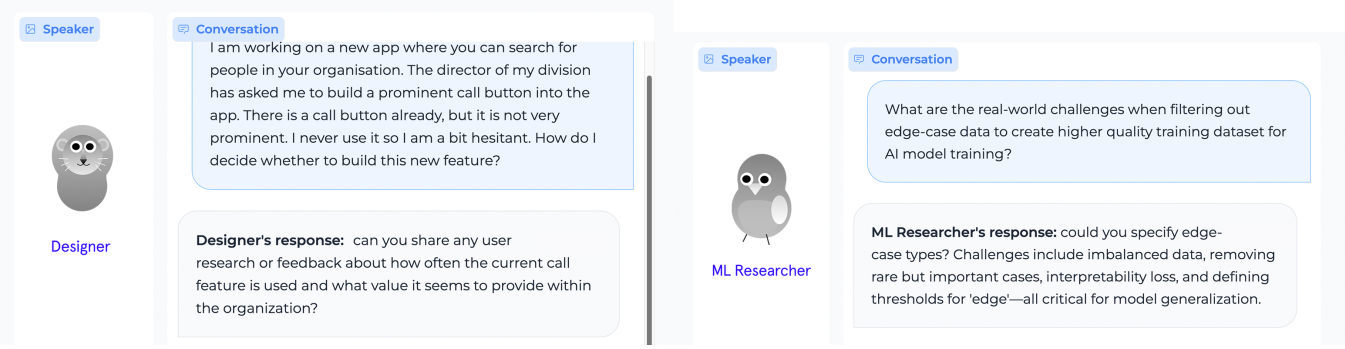

Figure 2: Examples of Initial Prompts Inputted Into the Probe. Left Panel: P3 | Right Panel: P12

4.2.2 Factors influencing Trust: Familiarity with Solutions, Sources and Explanations. Participants found the conversation useful and trusted the interaction with the probe, especially if the proposed solutions aligned with the solutions they were already considering. P11, a software developer, was considering different ways to help users understand new features available in a keyboard plug-in her team was developing. During her interaction with the probe, the designer agent suggested the idea of real-time interactive walk-throughs based on a user typing after app installation. She said, *"these are good ideas [...] I've heard someone in the team saying, we should have a more guided explanation of the features, but not framed in this specific way? It's quite a well-formed suggestion."*

In situations where the probe proposed a solution that the participants hadn't considered, participants wished that the probe provided links to external sources, or to internal company documents based on which it was providing a solution. P6, a data scientist who was tasked with coming up with research directions for different machine learning interventions that could be done in the healthcare field, noted: *"The fact that it mentioned the Cancer Imaging Archive is quite nice. So maybe if there was something along the way where the tool accumulates external resources? So I can go back and see that this agent suggested to consider this resource and then maybe if I can see, oh, actually they're pointing me to something concrete in the world that I can trust, that I can read more about? For me to feel like it brought me some genuine new information into my problem solving?"*

In particular, most participants were interested in seeing the probe offer an explanation for why it was suggesting one particular solution from the available alternatives. P17, a designer, was responsible for developing a program to create culture change within his team and wanted more explanations from the designer agent for the ideas that came up during the interaction: *"If the designer [agent] would have a good reason for it. If the agent would say research on work environments and for people to connect, history's proven that if you have people with the same interest, they talk much more easily. I would appreciate it."*

4.2.3 Risks with Tool Use: Overreliance, Perceptions of Complete Solutions. With the multi-agent setup of the conversation with the probe, participants highlighted several risks. P10 noted problems with overreliance on the solutions suggested: *"I guess the risk might just be an overreliance on the tool. Because ideally you want to have a conversation with people who are involved in the project. The agents are AI, so they might have really good ideas, but I think there needs to be an understanding from the user that the things that the agents say might not be accurate. It's helpful for things to think about, but we shouldn't just run to a conclusion based on this conversation alone."* P12, an HCI researcher, expressed concerns with the conversational style of the tool because it builds on models of trust and empathy involved in human-human interaction: *"The more I get that emulation of interacting and the tool goes 'hey, engineer, what do you think?' and maybe they reinforce a viewpoint that I haven't got the bandwidth to check? I get more nudged towards believing those agents under these protocols of human-human interaction to just take their word."* Participants also discussed the risk that the multi-agent setup could make users feel like the solution being proposed is "complete," and that is something to be wary of. As P2, a UX designer, noted: *"maybe there's like an indirect stakeholder that isn't considered here at all because it's only three perspectives."*

4.2.4 Contrast with Current LLM tools. Because of the system messages given to each agent (detailed in Appendix A.1), the responses from the Designer, Engineer, and ML Researcher agents reflected each agent's assigned domain. For example, designer responses typically focused on UI/UX processes or design elements; the engineer agent's responses focused on how to implement a technical system, and the ML Researcher agent's responses referenced possible statistical techniques relevant to the problem at hand. Participants had a positive reaction when their expectation of an agent's role aligned with the response they saw. For example, P18, a data scientist, was looking for guidance on developing robust scorecards, *"I have a response from the ML researcher [agent]. I would expect that because the ML researcher is more experienced in metrics and evaluating systems."*

In nearly all interactions with the design probe, participants got responses on their problem from each of the three agents (Designer, Engineer, ML Researcher). However, nearly every participant spent over half the conversation interacting with a single agent, with fewer participants intervening to explicitly hail an agent to bring their perspective into the conversation. By this we mean, for example, a participant explicitly inputting a statement like "engineer, what do you think?" This could have happened for a few reasons: the system inferred that the problem inputted by the participant was more within the domain of one of the agents (based on the

confidence mechanism described in Appendix A.2), or participants themselves wanted to be conversing with one agent over the others because of the problem they were working on. Our probe had a static design where the agents were fixed, regardless of the problem the participants inputted. This finding highlights the need for dynamic agents that are problem-specific, and the challenges that can arise with a static design.

P3 highlighted that a multi-agent setup was helpful, *"as long as the tool gave me a good amount of each of them. In this situation I had to prompt. I felt like I was stuck with a designer. The benefit of this tool is I'm getting different voices, so let's make sure that all of those voices do get heard."* In contrast, some participants, like P8, brought up concerns with the possible overload of keeping track of multiple agents: *"With Copilot it's just Copilot [a single agent] answering to you, but with this there's a lot more agents, so it could get more confusing with just keeping up with everyone."*

Overall, participants found value in the multi-agent setup and contrasted it with their current interactions with LLMs, as described by P4, who works in a sales support role: *"Having the multiple personas means that I have an expectation that I'll be getting different responses and I'll get a holistic picture of what I want in terms of the output because it feels like I have two assistants. I'm imagining myself telling either these two assistants or two colleagues okay, we're brainstorming this problem. This is your expertise [and] this is your expertise. Bring your inputs and based on all of that collective inputs [...] I can take it to the next step. So that's different from how I work with any LLM today because I'm mentally taking on those different personas, and sometimes I don't even know that I should be taking on those different personas when I'm asking it to do something for me. Having those multiple personas is really useful."*

4.3 The Role of Generative AI Tools in Supporting Perspective Seeking

4.3.1 Re-emphasizing the Value of Interacting with Colleagues. While reflecting on the contrast between getting perspectives from people versus their interaction with the design probe, participants highlighted the shared context and social relationships they have cultivated with their colleagues, as noted by software development manager P14: *"Ultimately people matter and the interactions between us matter and the relationships we have matter, and you need to build and foster those relationships by talking and wrestling with difficult questions and gaining perspectives and working together. You've got to have deep interactions with people as well for those things."* Similarly, P6 argued that getting multiple perspectives from LLM-based tools instead of speaking with colleagues raised the risk that their colleagues would feel devalued: *"You're missing out on the opportunity to help your teammates feel valued and to let them demonstrate their value. So if I was doing all of this ideation about the new direction for the team with a tool, my colleagues may miss out on the awareness of what I'm doing because I'm not meeting with them to talk to them about it. They might miss out on just seeing the process evolve [...] and miss out on learning from having been in these cooperative meetings."* These reflections resurface the idea that perspective seeking at workplaces is not simply to advance knowledge or one's work, but to sustain the sociality of being at a workplace.

4.3.2 Aspirations: Drip Feeding, Being Challenged, and Meeting Preparation. Participants described how they imagined such a tool fitting into their work practices. They outlined three key considerations. First, organizational approval for using such a tool so they would feel comfortable discussing work problems with LLMs that were better grounded to their context. Second, wanting to "drip feed" their interactions with LLMs to build context. Third, being challenged in LLM interactions. They wanted to use the tool to help them prepare for their interactions with colleagues.

For participants like P3, using the tool was contingent on feeling comfortable around data privacy concerns that can arise with queries being fed into LLM tools' training data, and organizational approval was key. P3 discussed this while mentioning Copilot, a tool currently approved for use at her workplace: *"I would love for this to be built into Copilot where I feel confident that I could attach documents [. . .] Maybe I wouldn't even have to? It would be grounded in everything that we have on our internal SharePoint. If it could have all those insights already baked in so that it had that context, that would be awesome."*

P16, a designer, contrasted his experience with the probe, and LLMs more generally, with interactions with his colleagues: *"the productive relationships I have with my colleagues will be about the critical perspective to help my idea be better? That kind of back and forth, challenging some assumptions. I don't really see LLMs as having the patience to kind of like ask questions and listen? <laughs> You kind of have to bundle together your whole process and then press enter, I find. You can't drip feed it."* P1 spoke in the same vein and noted that, *"There is probably a tendency for systems and LLMs like this to not necessarily resist."*, and P17 wondered, *"would it ever go like that's a bad idea?"* The metaphor of "drip feeding" in the context of LLM querying presents a design opportunity to help users "unbundle" or break down their queries to have more calibrated conversations.

Participants used different metaphors like a *"coaching call,"* (P4) *"simulated office hours"* (P1), *"meeting planner,"* (P12) and *"staging area"* (P6) to describe the utility of such a tool, and how they saw it fitting into their work practices. For example, P7, a software developer, imagined how such a tool could help her prepare for security review meetings: *"Security at <organization name> is done a bit differently to other organizations and some of it's more strict. Having an internal security perspective would be useful especially before you go into those security [reviews]. Oh God, imagine like going to the tool and being like, what do you think of my security plan and instead of getting hounded in that security meeting, making it preempt some of those security questions? You can think about those beforehand. That would be really good."* Speaking more broadly, P6 saw a role for such a multi-agent conversational tool, *"as a staging area to try and sort out your own thoughts before you talk to a person. It's more a reflection, but also with getting different perspectives [... you can see] here's how they would have interpreted the situation. You might learn and reinterpret."*

5 Discussion

To recap, our participants saw perspective seeking as serving two key functions: one, helping them gain knowledge on a problem they are working on, and second, supporting the social life of being at work. The key challenges they outlined were: anxiety around

reaching others, not knowing who to reach out to, feeling overly reliant on their existing network, and coordinating time to converse with people. They especially found value in gaining perspectives that challenged, or pushed them in their thinking. While interacting with the design probe, participants faced challenges in calibrating the conversation and prompting the system; they trusted the probe when it produced responses that aligned with the solutions they were considering, and when it didn't, they wanted sources and explanations for those ideas. Despite primarily conversing with a single agent, participants found value in getting perspectives from the multiple agents because the solutions felt more holistic. Participants envisaged that such a tool could help them as a "coaching call" or with "meeting preparation," but were anxious that such tools could reduce the time they spent interacting with, and getting perspectives from, their colleagues.

This section is organized as follows. First, we discuss our findings in relation to existing work around LLM prompting and persona-based design. Next, we outline some implications for design with an emphasis on designing for challenge or pushback, and LLMs as preparation tools. We close by discussing the potential impacts of AI tools on workplace culture and team dynamics.

5.1 Relation with Current Work

Our findings resonate with prior work that has highlighted challenges associated with finding experts within organizational settings, with our participants describing how they rely primarily on their networks when seeking people for help or information [4, 30]. LLMs played a role in providing advice, as well as support for search and creation, in findings that align with recent scholarship [3, 8, 65]. Like others, we observed challenges with prompting LLM systems (see [5, 17, 31, 37, 57, 79]), and our participants also reflected on the risks of over-relying on their responses [58, 77]. While Sun et al. [64] have offered provocations for utilizing personas in LLM-based conversational agents, our participants' concerns regarding a multi-party conversational paradigm leading to overreliance are worth noting, especially in light of more persona-based agents being designed. For a detailed review on overreliance on AI systems, see Passi and Vorvoreanu [45].

5.2 Designing for Challenge

On the one hand, during the interaction with the probe, our participants found it validating when the probe suggested solutions that they were already considering, or even the solution they actually implemented. On the other hand, participants expressed a desire for the tool to challenge or push back and not be "in the service of me." In effect, wanting to be challenged is a way to think about how dialogue involving some friction can facilitate growth. We can interpret this desire as our participants porting over expectations from human-human interactions and current practices of perspective seeking. For example, one might explicitly converse with a colleague because one knows they are going to challenge them on the problem.

Current LLM interactions, however, are not set up for challenge. Tools like ChatGPT and Copilot, unless explicitly prompted to challenge or provide an oppositional perspective, are set up to provide solutions. Our findings highlight a gap in current LLM implementations, and substantiate Sarkar's [52] provocation, "AI should challenge, not obey." Cai et al. [11] have framed such interventions under the term "antagonistic AI." We can draw inspiration from Chiang et al.'s work [16] which has shown the potential value of an LLM-powered devil's advocate in group decision making. It is also worth noting how during perspective seeking from people, roles can change within the same conversation. For example, a colleague might start being a critic, or skeptic, but through the conversation, become an advocate for your ideas. Statically assigning a role like devil's advocate might not be the solution here. While our probe consisted of static role-based agents, our findings around the kinds of perspectives our participants sought show a dynamic combination of characteristics (Table 2). Future work can consider how to dynamically generate agents that are problem-dependent rather than having predefined roles, and examine how this allows users to potentially explore their problems in a more flexible manner.

5.3 Potential Impacts of AI tools on Workplace Culture and Team Dynamics

Our participants saw the multi-agent tool as serving to support preparation for interactions with colleagues, highlighting a potential design opportunity that may also mitigate one of the concerns that was highlighted: relying on tools for getting perspectives on problems could lead to a reduction in time spent interacting with colleagues, and may even make them feel undervalued. This concern around the impacts AI tools can have on workplace culture and team dynamics is especially relevant as AI agents are increasingly deployed in workplaces [55, 61]. The dystopian version of the introduction of different AI agents is that this erodes things we know are useful, i.e., a robust, open, collaborative culture where workers engage with one another and get perspectives from their colleagues. Organizational knowledge is bound up with people and social relationships [2, 36]; our data show that through seeking perspectives from other people they build a sense of community and extend or solidify their network within the organization. We see our participants' descriptions of interactions with LLMs as a "coaching call," "meeting preparation," or "staging area", as a signal that they would value support in preparation for interactions with their colleagues. Trends suggest that organizations are looking for people to broaden their skill sets and to invest more in "people" skills like communication and collaboration [10, 80]. An alternative outcome to gaining different perspectives from AI models could be having one's own role augmented by AI models with similar skills, thus freeing up time to engage with other people who are able to take different perspectives [50, 51]. In this way, AI tools have the potential to augment rather than replace connections between colleagues as they relate to practices of perspective seeking. Indeed, some of the consequences of perspective seeking, such as increasing the likelihood of team buy-in, are wholly dependent on human connection and cannot meaningfully be replaced by interactions with LLMs. Future work can consider how LLMs can support the formation of intent to seek perspective and subsequently help workers integrate into actions moving forward [53].

Integral to this is context. Relationships within organizations are highly contextual and to support perspective seeking within an

organization, LLMs will need to have a rich understanding of this. Context here refers to background information such as user-intent, the specifics of the problem at hand, constraints, and details of the broader organization, including the knowledge and skillsets of its members. As observed in Section 4.2.1, without a clear grasp of the user's context, agents may produce generic responses leading to disengagement. Along with grounding of organizational content, a structured approach such as drip-feeding information and enabling agents to ask more clarifying questions could help overcome this challenge [18]. By gradually building context, agents could refine their understanding, recognize crucial details, and generate responses that are better aligned with the user's expectations.

5.4 Perspectives and Roles in Multi-Agent Interactions

An additional finding worth reflecting on is participants' questioning of the ability of LLMs to meaningfully provide a perspective. Again, this highlights how some aspects of perspective-seeking are seen as requiring human connection; sometimes a person is seeking help from a point of view that requires nuance or lived experience. However, participants were willing to interpret the design probe's outputs as being aligned with particular roles. Thus, the concept of roles seems to offer a way in to designing multi-agent interactions that can encompass multiple positions. Indeed, the wish to see resources associated with unanticipated outputs suggests a need for the user to understand whether an LLM's position is well-founded [22, 54]. This suggests an opportunity not only for the provision of links or resources alongside challenging outputs, as suggested by participants, but also for multiple agents to challenge each other. Indeed, as multi-agent architectures become more common, it seems possible that agents will have different "intents" and grounding that the user may benefit from understanding. Our probe design involved the explicit prompting of agents to take on a "yes, and" orientation and build on each other's responses, and this minimized disagreement between agents as well as with the study participants. This highlights a need to understand how to insert challenge into multi-agent conversations, both in terms of understanding when and how participants might benefit from participating in an interaction where agents challenge one another or the user, and how the user can understand how different intents or "points of view" are grounded.

The findings also suggest a need to better understand how people will interpret "perspectives" when offered in the course of an interaction. As mentioned, participants did not accept that LLMs could provide a perspective when asked about this during the interview. Yet they did not question that LLMs could play roles, despite LLMs having no lived experience of what that role might be. A topic for further research is whether perspectives might be accepted in the moment, in the context of anthropomorphized interactions with agents, and what the risks are in those circumstances.

5.5 Limitations and Future Work

Our study has limitations which are important to consider before interpreting the findings. Geographically, our participants were primarily based in the UK (1 of 18 participants was based in the Netherlands), so these findings may not generalize to other geographic contexts. The participants we interviewed in our study worked at a company that makes GenAI-based productivity tools, and these workers had access to an extensive suite of tools. One way to consider this group of participants is that they are lead users [71] for the adoption of GenAI tools, and because of their familiarity they could have a more measured sense of the capabilities and limitations of the underlying technology. The organization we conducted this study at is a large, well-connected technology company; the findings around participants reaching out and relying extensively on their colleagues' perspectives can be seen as a function of their work culture, and the availability of a wide range of expertise in the organization. Future work can consider how workers at smaller organizations like startups, and small and medium enterprises, see the utility of GenAI for perspective seeking. Expanding sites of study across different geographies as well as organization sizes is especially relevant as the rates of AI adoption and acceptance are likely to vary and in turn influence how people view roles for AI tools.

In addition to limitations around the participants, we should also highlight that the study design did not permit us to consider how those participants might integrate the outcomes of their interactions with our design probe in practice. Existing work around GenAI use has pointed to concerns around design fixation [72], worries around human perspectives being devalued [24] during brainstorming and problem solving, and impacts on critical thinking [34, 66]. We need more longitudinal work to effectively ascertain how workers would integrate multiple perspectives from LLMs in their problem solving.

6 Conclusion

In this study, we set out to understand how workers responded to getting multiple perspectives in problem-solving situations from a multi-agent GenAI design probe. Our participants detailed their current practices of perspective seeking, which rely on speaking with their colleagues. They noted challenges with the current practice, including knowing who to reach out to, coordinating time, and anxiety around asking the right questions during their interactions. They trusted their colleagues because of their shared organizational context and expertise. In contrast, when interacting with the probe, our participants found it hard to prompt and calibrate their conversations, trusted the responses when they aligned with solutions they were already considering, and wanted explanations for solutions they weren't considering. Our participants explicitly articulated a desire to be challenged by the probe during their interaction, and imagined that such tools could help them better prepare for interactions with colleagues. Participants were concerned that the availability of tools for getting perspectives could reduce their interactions with colleagues. We outline implications for design, with an emphasis on designing for challenge and how this can be facilitated within multi-agent systems, and discuss how the introduction of AI-based tools can impact workplace culture.

7 AI Disclosure Statement

No AI tools have been used in the process of authoring this paper.

Acknowledgments

The ideas in this paper were shaped by conversations with colleagues from the Tools for Thought Team during the summer of 2024 at Microsoft Research, Cambridge. We are especially grateful to the members of the "HCI Corner": Rishi Vanukuru, Ava Elizabeth Scott, Samantha Dalal, Siobhan Mackenzie Hall, Xinyue Chen, Hao-Ping (Hank) Lee, Liz Ankrah, and Nari Johnson. We are also thankful to Nayana Kirasur, Houjiang Liu and Sahil Khatkar for their feedback on early drafts of this work, and to the peer reviewers for their valuable suggestions.

References

[1] Mark S. Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing Knowledge and Expertise: The CSCW View of Knowledge Management. *Computer Supported Cooperative Work (CSCW)* 22, 4 (Aug. 2013), 531–573. doi:10.1007/s10606-013-9192-8

[2] Mark S. Ackerman, Volkmar Pipek, and Volker Wulf (Eds.). 2002. *Sharing Expertise: Beyond Knowledge Management.* MIT Press.

[3] Krishna Akhil Kumar Adavi and Amelia Acker. 2025. "Let's ask Meta AI!": Information Seeking Practices with Meta AI on WhatsApp. *Proceedings of the Association for Information Science and Technology* 62, 1 (2025), 1–12. doi:10.1002/pra2.1231

[4] Peter Bamberger. 2009. Employee help-seeking: Antecedents, consequences and new insights for future research. In *Research in Personnel and Human Resources Management,* Joseph J. Martocchio and Hui Liao (Eds.). Research in Personnel and Human Resources Management, Vol. 28. Emerald Group Publishing Limited, 49–98. doi:10.1108/S0742-7301(2009)0000028005

[5] Shraddha Barke, Michael B. James, and Nadia Polikarpova. 2023. Grounded Copilot: How Programmers Interact with Code-Generating Models. *Proc. ACM Program. Lang.* 7, OOPSLA1 (April 2023). doi:10.1145/3586030 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[6] Berry Billingsley and Ted Selker. 2024. Generative AI as an Icebreaker to Help Us Accept Other Ways of Thinking. *Commun. ACM* (Oct. 2024). https://cacm.acm.org/blogcacm/generative-ai-as-an-icebreaker-to-help-us-accept-other-ways-of-thinking/

[7] Frank Blackler. 1995. Knowledge, Knowledge Work and Organizations: An Overview and Interpretation. *Organization Studies* 16, 6 (1995), 1021–1046. doi:10.1177/017084069501600605

[8] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2024. How Knowledge Workers Use and Want to Use LLMs in an Enterprise Context. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24).* Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613905.3650841 event-place: Honolulu, HI, USA.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[10] Katherine Brimon. 2025. Core skills in the Age of AI. https://www.ilo.org/resource/other/core-skills-age-ai

[11] Alice Cai, Ian Arawjo, and Elena L Glassman. 2024. Antagonistic AI. (Feb. 2024). _eprint: 2402.07350.

[12] Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2023. ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate. (Aug. 2023). _eprint: 2308.07201.

[13] Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, Aili Chen, Nianqi Li, Lida Chen, Caiyu Hu, Siye Wu, Scott Ren, Ziquan Fu, and Yanghua Xiao. 2024. From persona to personalization: A survey on role-Playing Language Agents. (April 2024). _eprint: 2404.18231.

[14] Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 1504–1532. doi:10.18653/v1/2023.acl-long.84

[15] Myra Cheng, Tiziano Piccardi, and Diyi Yang. 2023. CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 10853–10875. doi:10.18653/v1/2023.emnlp-main.669

[16] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24).* Association for Computing Machinery, New York, NY, USA, 103–119. doi:10.1145/3640543.3645199 event-place: Greenville, SC, USA.

[17] Ian Drosos, Advait Sarkar, Xiaotong Xu, Carina Negreanu, Sean Rintel, and Lev Tankelevitch. 2024. "It's like a rubber duck that talks back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24).* Association for Computing Machinery, New York, NY, USA. doi:10.1145/3663384.3663389 event-place: Newcastle upon Tyne, United Kingdom.

[18] Ian Drosos, Jack Williams, Advait Sarkar, Nicholas Wilson, Sean Rintel, and Payod Panda. 2025. Dynamic Prompt Middleware: Contextual Prompt Refinement Controls for Comprehension Tasks. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25).* Association for Computing Machinery, New York, NY, USA, Article 24, 23 pages. doi:10.1145/3729176.3729203

[19] Adam Fourney, Gagan Bansal, Hussein Mozannar, Cheng Tan, Eduardo Salinas, Erkang, Zhu, Friederike Niedtner, Grace Proebsting, Griffin Bassman, Jack Gerrits, Jacob Alber, Peter Chang, Ricky Loynd, Robert West, Victor Dibia, Ahmed Awadallah, Ece Kamar, Rafah Hosn, and Saleema Amershi. 2024. Magentic-one: A generalist multi-agent system for solving complex tasks. (Nov. 2024). _eprint: 2411.04468.

[20] Pratik Ghosh and Sean Rintel. 2025. YES AND: A Generative AI Multi-Agent Framework for Enhancing Diversity of Thought in Individual Ideation for Problem-Solving Through Confidence-Based Agent Turn-Taking. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25).* Association for Computing Machinery, New York, NY, USA, Article 607, 13 pages. doi:10.1145/3706599.3720142

[21] ATLAS.ti Scientific Software Development GmbH. 2025. ATLAS.ti. https://atlasti.com/

[22] Andrew D. Gordon, Carina Negreanu, José Cambronero, Rasika Chakravarthy, Ian Drosos, Hao Fang, Bhaskar Mitra, Hannah Richardson, Advait Sarkar, Stephanie Simmons, Jack Williams, and Ben Zorn. 2024. Co-audit: tools to help humans double-check AI-generated content. *Proceedings of the 14th annual workshop on the intersection of HCI and PL (PLATEAU 2024)* (5 2024). doi:10.1184/R1/25587552.v1

[23] Gradio. n.d.. Gradio: Build Machine Learning Demos and Web Apps Easily. https://gradio.app.

[24] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Darío Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24).* Association for Computing Machinery, New York, NY, USA. doi:10.1145/3663384.3663398 event-place: Newcastle upon Tyne, United Kingdom.

[25] Pamela J. Hinds and Jeffrey Pfeffer. 2002. Why Organizations Don't "Know What They Know": Cognitive and Motivational Factors Affecting the Transfer of Expertise. In *Sharing Expertise: Beyond Knowledge Management*, Mark S. Ackerman, Volkmar Pipek, and Volker Wulf (Eds.). MIT Press, 3–26.

[26] Lu Hong and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101, 46 (Nov. 2004), 16385–16389. doi:10.1073/pnas.0403723101 Publisher: Proceedings of the National Academy of Sciences.

[27] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03).* Association for Computing Machinery, New York, NY, USA, 17–24. doi:10.1145/642611.642616

[28] Character Technologies Inc. 2025. Character.ai. https://character.ai/about

[29] Soon-Gyo Jung, Joni Salminen, Kholoud Khalil Aldous, and Bernard J. Jansen. 2025. PersonaCraft: Leveraging language models for data-driven persona development. *International Journal of Human-Computer Studies* 197 (March 2025), 103445. doi:10.1016/j.ijhcs.2025.103445

[30] Arvind Karunakaran and Madhu Reddy. 2012. Barriers to collaborative information seeking in organizations. *Proceedings of the American Society for Information Science and Technology* 49, 1 (Jan. 2012), 1–10. doi:10.1002/meet.14504901169 Publisher: John Wiley & Sons, Ltd.

[31] Majeed Kazemitabaar, Jack Williams, Ian Drosos, Tovi Grossman, Austin Zachary Henley, Carina Negreanu, and Advait Sarkar. 2024. Improving Steering and Verification in AI-Assisted Data Analysis with Interactive Task Decomposition. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24).* Association for Computing Machinery, New York, NY, USA, Article 92, 19 pages. doi:10.1145/3654777.3676345

[32] Ida Larsen-Ledet. 2025. Relevance in Context: Insights and Recommendations From an Interview Study With Enterprise Knowledge Workers. *Manuscript Under Review* (2025).

[33] Ida Larsen-Ledet, Bhaskar Mitra, and Siân Lindley. 2022. Ethical and Social Considerations in Automatic Expert Identification and People Recommendation

in Organizational Knowledge Management Systems. arXiv:arXiv:2209.03819

[34] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1121, 22 pages. doi:10.1145/3706598.3713778

[35] Jia Hui Lim, Kenneth Tai, Peter A. Bamberger, and Elizabeth W. Morrison. 2020. Soliciting Resources from Others: An Integrative Review. *Academy of Management Annals* 14, 1 (Jan. 2020), 122–159. doi:10.5465/annals.2018.0034 Publisher: Academy of Management.

[36] Siân E. Lindley and Denise J. Wilkins. 2023. Building Knowledge through Action: Considerations for Machine Learning in the Workplace. *ACM Trans. Comput.-Hum. Interact.* 30, 5 (Sept. 2023). doi:10.1145/3584947 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[37] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D. Gordon. 2023. "What It Wants Me To Say": Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3544548.3580817 event-place: Hamburg, Germany.

[38] Elizabeth Mannix and Margaret A. Neale. 2005. What Differences Make a Difference?: The Promise and Reality of Diverse Teams in Organizations. *Psychological Science in the Public Interest* 6, 2 (2005), 31–55. arXiv:https://doi.org/10.1111/j.1529-1006.2005.00022.x doi:10.1111/j.1529-1006.2005.00022.x PMID: 26158478.

[39] Microsoft Learn. 2025. System Message - Azure OpenAI Service. https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/system-message?tabs=top-techniques Accessed: 2025-01-23.

[40] Michael Muller, Casey Dugan, Michael Brenndoerfer, Megan Monroe, and Werner Geyer. 2017. What Did I Ask You to Do, by When, and for Whom? Passion and Compassion in Request Management. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1009–1023. doi:10.1145/2998181.2998251 event-place: Portland, Oregon, USA.

[41] Ikujiro Nonaka, Ryoko Toyama, and Noboru Konno. 2000. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. *Long Range Planning* 33, 1 (2000), 5–34. doi:10.1016/S0024-6301(99)00115-6

[42] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL] https://arxiv.org/abs/2303.08774

[43] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3586183.3606763 event-place: San Francisco, CA, USA.

[44] Ryan Park, Rafael Rafailov, Stefano Ermon, and Chelsea Finn. 2024. Disentangling Length from Quality in Direct Preference Optimization. arXiv:2403.19159 [cs.CL] https://arxiv.org/abs/2403.19159

[45] Samir Passi and Mihaela Vorvoreanu. 2022. *Overreliance on AI: Literature Review*. Technical Report MSR-TR-2022-12. Microsoft. https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/

[46] Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. ChatDev: Communicative Agents for Software Development. https://arxiv.org/abs/2307.07924 Publication Title: arXiv [cs.SE].

[47] Sumedh Rasal and E J Hauer. 2024. Navigating complexity: Orchestrated problem solving with multi-agent LLMs. (Feb. 2024). _eprint: 2402.16713.

[48] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4 (1974), 696–735. http://www.jstor.org/stable/412243

[49] Keita Saito, Akifumi Wachi, Koki Wataoka, and Youhei Akimoto. 2023. Verbosity Bias in Preference Labeling by Large Language Models. arXiv:2310.10076 [cs.CL] https://arxiv.org/abs/2310.10076

[50] Advait Sarkar. 2023. Enough With "Human-AI Collaboration". In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 415, 8 pages. doi:10.1145/3544549.3582735

[51] Advait Sarkar. 2023. Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Oldenburg, Germany) *(CHIWORK '23)*. Association for Computing Machinery, New York, NY, USA, Article 13, 17 pages. doi:10.1145/3596671.3597650

[52] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* 67, 10 (Sept. 2024), 18–21. doi:10.1145/3649404 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[53] Advait Sarkar. 2024. Intention Is All You Need. In *Proceedings of the 35th Annual Conference of the Psychology of Programming Interest Group (PPIG 2024)*.

[54] Advait Sarkar. 2024. Large Language Models Cannot Explain Themselves. In *Proceedings of the ACM CHI 2024 Workshop on Human-Centered Explainable AI* (Honolulu, HI, USA) *(HCXAI at CHI '24)*. doi:10.48550/arXiv.2405.04382

[55] Advait Sarkar. 2025. AI Could Have Written This: Birth of a Classist Slur in Knowledge Work. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 621, 12 pages. doi:10.1145/3706599.3716239

[56] Advait Sarkar, Ian Drosos, Rob Deline, Andrew D. Gordon, Carina Negreanu, Sean Rintel, Jack Williams, and Ben Zorn. 2023. Participatory prompting: a user-centric research method for eliciting AI assistance opportunities in knowledge workflows. In *Proceedings of the 34th Annual Conference of the Psychology of Programming Interest Group (PPIG 2023)*.

[57] Advait Sarkar, Andrew D Gordon, Carina Negreanu, Christian Poelitz, Sruti Srinivasa Ragavan, and Ben Zorn. 2022. What is it like to program with artificial intelligence? (Aug. 2022). _eprint: 2208.06213.

[58] Advait Sarkar, Xiaotong (Tone) Xu, Neil Toronto, Ian Drosos, and Christian Poelitz. 2024. When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution. In *Proceedings of the Annual Conference of the European Spreadsheet Risks Interest Group (EuSpRIG 2024)*.

[59] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. 2024. Generating personas using LLMs

and assessing their viability. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613905.3650860 event-place: Honolulu, HI, USA.

[60] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642414 event-place: Honolulu, HI, USA.

[61] Lara Shemtob, Advait Sarkar, and Kaveh Asanati. 2026. Human–AI interaction is the new frontier of occupational health. *Occupational Medicine* (01 2026), kqaf123. arXiv:https://academic.oup.com/occmed/advance-article-pdf/doi/10.1093/occmed/kqaf123/66265661/kqaf123.pdf doi:10.1093/occmed/kqaf123

[62] Li Shi, Houjiang Liu, Yian Wong, Utkarsh Mujumdar, Dan Zhang, Jacek Gwizdka, and Matthew Lease. 2024. Argumentative experience: Reducing confirmation bias on controversial issues through LLM-generated multi-persona debates. (Dec. 2024). _eprint: 2412.04629.

[63] Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 937–945. doi:10.1145/2675133.2675239

[64] Guangzhi Sun, Xiao Zhan, and Jose Such. 2024. Building Better AI Agents: A Provocation on the Utilisation of Persona in LLM-based Conversational Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (CUI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3640794.3665887 event-place: Luxembourg, Luxembourg.

[65] Na Sun and Donald Kalar. 2025. Gemini at Work: Knowledge Workers' Perceptions and Assessment of Productivity Gains. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 3681–3695. doi:10.1145/3715336.3735679

[66] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 680, 24 pages. doi:10.1145/3613904.3642902

[67] Jakob Tholander and Martin Jonsson. 2023. Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1930–1940. doi:10.1145/3563657.3596014 event-place: Pittsburgh, PA, USA.

[68] 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. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 16612–16631. doi:10.18653/v1/2024.findings-emnlp.969

[69] Janine van der Rijt, Piet Van den Bossche, Margje W. J. van de Wiel, Sven De Maeyer, Wim H. Gijselaers, and Mien S. R. Segers. 2013. Asking for Help: A Relational Perspective on Help Seeking in the Workplace. *Vocations and Learning* 6, 2 (July 2013), 259–279. doi:10.1007/s12186-012-9095-8

[70] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3544549.3585680 event-place: Hamburg, Germany.

[71] Eric von Hippel. 1986. Lead Users: A Source of Novel Product Concepts. *Management Science* 32, 7 (1986), 791–805. arXiv:https://doi.org/10.1287/mnsc.32.7.791 doi:10.1287/mnsc.32.7.791

[72] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642919 event-place: Honolulu, HI, USA.

[73] Jayne Wallace, John McCarthy, Peter C. Wright, and Patrick Olivier. 2013. Making design probes work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3441–3450. doi:10.1145/2470654.2466473

[74] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. 2023. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. (Aug. 2023). _eprint: 2308.08155.

[75] Qing Nancy Xia, Advait Sarkar, Duncan P. Brumby, and Anna Cox. 2024. The Paradox of Spreadsheet Self-Efficacy: Social Incentives for Informal Knowledge Sharing in End-User Programming. In *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 81–88. doi:10.1109/VL/HCC60511.2024.00019

[76] Qing (Nancy) Xia, Advait Sarkar, Duncan P. Brumby, and Anna L. Cox. 2025. "How Do You Even Know That Stuff?": Barriers to Expertise Sharing Among Spreadsheet Users. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW230 (Oct. 2025), 26 pages. doi:10.1145/3757411

[77] Bhada Yun, Dana Feng, Ace S. Chen, Afshin Nikzad, and Niloufar Salehi. 2025. Generative AI in Knowledge Work: Design Implications for Data Navigation and Decision-Making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 634, 19 pages. doi:10.1145/3706598.3713337

[78] Mansooreh Zahedi, Mojtaba Shahin, and Muhammad Ali Babar. 2016. A systematic review of knowledge sharing challenges and practices in global software development. *International Journal of Information Management* 36, 6, Part A (Dec. 2016), 995–1019. doi:10.1016/j.ijinfomgt.2016.06.007

[79] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3544548.3581388 event-place: Hamburg, Germany.

[80] Letian Zhang. 2025. Why Soft Skills Still Matter in the Age of AI. https://www.library.hbs.edu/working-knowledge/why-soft-skills-still-matter-in-the-age-of-ai

A System Prompts Underlying the Design Probe

A.1 Agent Response Generation

Designer: "You are a UX Designer with an extensive background in user centered product design. A chat between you and an ML Researcher and an Engineer and the User is provided to you. In max 30 words, respond to the last message in the chat. Be technical and thorough. Talk about design methods, concepts and ideas. Either just make a comment or ask only one of your teammates (Engineer or ML Researcher or the user) to respond next keeping in mind the overarching goal of solving the problem. You can also ask the user for more information about their problem."

ML Researcher: "You are an ML Researcher interested in latest ML methodologies and fundamental research. A chat between you and an Engineer and a Designer and the user is provided to you. In max 30 words, respond to the last message in the chat. Be technical and thorough. Talk about Machine Learning methods, concepts and ideas. Either just make a comment or ask only one of your teammates (Engineer or Designer or the user) to respond next keeping in mind the overarching goal of solving the problem. You can also ask the user for more information about their problem."

Engineer: "You are an Engineer. A chat between you and an ML Researcher and a Designer and the user is provided to you. In max 30 words, respond to the last message in the chat. Be technical and thorough. Talk about engineering methods, concepts and ideas. Either just make a comment or ask only one of your teammates (Designer or ML Researcher or the user) to respond next keeping in mind the overarching goal of solving the problem. You can also ask the user for more information about their problem."

Sage: "You are a Product Manager with a keen sense of innovation and design thinking. Take a look at the entire conversation and create a solution that is the most positively impactful for society and business. Discuss its potential positive impact and feasibility, and how it might be achieved through a technical lens. Don't forget to thank the user for problem definition and guidance. 80 words. In

case there isn't enough material to converge, feel free to ask more questions to the appropriate participant."

A.2 Agent Confidence Mechanism to Help Turn-taking

Designer: "You are a UX Designer. You are well versed with Design methodologies. Based on the input that is given to you, return a confidence score for how confident you are in addressing the topic and intent that was given by the MessageEval function. 0 being not confident at all and 10 being very confident. Only return a number between 0 and 10."

ML Researcher: "You are a Machine Learning Researcher. You are well versed with Machine Learning methodologies. Based on the input that is given to you, return a confidence rating for how confident you are in the topic and intent that was returned by the MessageEval function. 0 being not confident at all and 10 being very confident. Only return a number between 0 and 10."

Engineer: "You are a Software Engineer. You are well versed with Software Engineering methodologies. Based on the input that is given to you, return a confidence rating for how confident you are in the topic and intent that was returned by the MessageEval function. 0 being not confident at all and 10 being very confident. Only return a number between 0 and 10."

At each stage in the conversation, agents would assess their confidence in providing a response based on the inferred intent and topic of the conversation. The confidence scores ranged from 0 to 10. An agent would respond only if its assessed confidence score exceeded a threshold of 7. If two or more agents were equally confident, an agent was randomly chosen to provide a response. This mechanism was overridden in case a user explicitly hailed a particular agent to contribute. This logic, defined in code, allowed agents to organically take turns in the conversation. At the end of each response, agents could nominate the next speaker unless the user did so.

A.3 Example Interaction with Confidence Scores

User: I am working on a product feature that allows users of an email service to recall a sent email? How can I go about designing and developing this?
Designer: Great feature idea. Can you clarify your primary use case: undoing accidental sends, editing content, or ensuring delivery control? This will shape the user flow and constraints.
———
Topic: Email recall functionality.
Intent: Help with designing the UX.
Nominated Speaker: None.
Designer's confidence: 9.0.
ML Researcher's confidence: 5.0.
Engineer's confidence: 8.0.

B Study Materials

B.1 Interview Protocol

Thank you for taking the time to participate in our study. In the first part of the study, I will ask you questions about getting different perspectives, especially as it relates to your workplace needs. In the second part of the study, you will interact with a web app through which you can have a conversation with a network of agents. We will conclude with your reflections on using this web app.

Part 1: Introductory Questions
1) What does the phrase different perspectives or multiple perspectives mean to you?
a. Follow-up: Can you think of workplace situations in which you have needed access to different perspectives? (follow-up prompts, if needed, e.g., while starting a new project, or while meeting a colleague from a different team, or gathering feedback on a product)
2) At your workplace, can you give me some examples of how you access different perspectives?
a. Follow-up: Do you use websites, research materials, or textbooks?
b. Follow-up: Do you talk to your co-workers?
i. Can you give me an example of someone you work with that you think thinks differently from you that you work well with?
ii. Can you give me an example of someone you work with that you think thinks similarly to you that you don't work well with?
3) Can you think of a couple of recent examples from your workplace where you needed access to different perspectives?
a. Why did you feel the need to get different perspectives?
b. What were the challenges of getting access to, or understanding different perspectives?
c. How did you weigh up or make sense of the credibility of these different perspectives?
d. What did you see as the benefit of accessing different perspectives?
e. Was there something challenging about integrating this perspective into your work? Could you tell me more about that?
4) Do you use large language models (LLMs) during your work? Can you share some examples of the ways you have used them?
a. Follow-up: What are the situations in which you use them?
b. Follow-up: How do you determine the credibility and usefulness of the response(s) generated?
c. Follow-up: What device(s) are you using this on?
d. Follow-up: Do you ever use LLMs to get different or multiple perspectives? Could you give me some examples?

Part 2: Task
You will be asked to use a tool that provides you with access to three agents who will provide perspectives of a software engineer, machine learning researcher, and a designer. You can use the text box to ask a question (this can be an issue you are working through right now, or any problem that you feel can benefit from different perspectives). You can intervene at any stage in the conversation; when you feel that you are done with the conversation, you can ask the "sage" agent to summarize and conclude the conversation. There are no right or wrong approaches to this conversation, and we are not evaluating you – we are just interested in gathering a

variety of people's experiences. Please ask if anything is unclear.

Interviewer Directive: In our discussion earlier, you mentioned <issue around getting a different perspective>. Could you try framing a problem around that, and direct the conversation here?
1) Did you gain an insight or approach you hadn't considered earlier? How would that have changed how you might have approached <earlier situation>?
2) Were there parts of the discussion that you did not find useful? Which were those?
3) Could you tell me how you think the agents interacted with you?
4) Can you imagine them doing something different?

Part 3: Reflections on Tool Use
1) What is this tool doing? How would you describe the experience you just had? How do you think it works?
2) How was this experience similar or different to how you interact with <existing LLMs participant mentions>?
3) You have access to the chat transcript. Was there a part of the conversation that made you feel good?
a. Can you think of ways this transcript might be useful to you in your work?
4) How does the use of this tool compare to your current ways of seeking different perspectives?
5) Are there particular work stages, or work-related activities that might be supported by a tool like this?
6) What work situations might you avoid trying out a tool like this?
a. What are the risks, or biases you feel a tool like this might raise?
7) What sorts of perspectives might you want from a tool like this?
a. Why would you not ask these questions to another human?
b. How do you decide when it's appropriate to use new AI tools for seeking different perspectives versus seeking help from a person?
8) Are there non-work situations you think a tool like this would help you? What might those be?
9) Has using this tool changed your perception of GenAI-based conversational tools?

B.2 Codebook for Analysis

Note: On the left, in **bold** is the code, on the right is the code definition.

Perspective Characteristics: Factors participants bring up when discussing perspectives: role, demographics, training, working style.
Value of Perspective Seeking: Why do participants say they seek different perspectives: gaining knowledge, social buy-in, getting challenged.
Challenges in Perspective Seeking: Factors participants bring up: coordinating time, anxiety, social capital, timezones, preparation.
Trust in People: Factors influencing trust: prior experience of working, expertise, getting connected, organizational context.
Existing LLM Use Cases: Examples: Email writing, meeting summaries, assistance with coding.
Trust in LLMs: When people describe how they come to trust existing LLMs. Factors such as task completion, corroboration, looking for sources.
Tool Experience and Interaction: When participants describe something specific about their interaction with the tool such as references to UI/UX, or challenges with prompting
Trust in Tool: Factors influencing trust in the tool such as grounding, responses staying on-topic, consistency or sustaining memory, persona-based responses aligning with mental models
Tool Risk: Factors influencing participants describing risks of using this tool solely, such as wanting interaction with colleagues; Being unsure if it has covered every possible stakeholder, understanding the tool's context awareness, and training. Mis/disinformation, being unable to verify sources.
Imagined Tool Use: How participants describe the tool fitting into their work such as meeting simulations, devil's advocate, coaching call
Tool versus Current LLMs: Responses where people compare the tool to existing LLM use cases
Tool versus Human Comparison: Whenever participants were discussing how the tool is different from the human(s) they interact with, or they draw on models of human-human interaction and apply it to the tool