Dittos: Personalized, Embodied Agents That Participate in Meetings When You Are Unavailable

JOANNE LEONG, Microsoft Research & MIT Media Lab, USA

JOHN TANG, Microsoft Research, USA

EDWARD CUTRELL, Microsoft Research, USA

SASA JUNUZOVIC, Microsoft Research, USA

GREGORY PAUL BARIBAULT, Microsoft & Hewlett-Packard, USA

KORI INKPEN, Microsoft Research, USA

Imagine being able to send a personalized embodied agent to meetings you are unable to attend. This paper explores the idea of a *Ditto*—an agent that visually resembles a person, sounds like them, possesses knowledge about them, and can represent them in meetings. This paper reports on results from two empirical investigations: 1) focus group sessions with six groups (n=24) and 2) a Wizard of Oz (WOz) study with 10 groups (n=39) recruited from within a large technology company. Results from the focus group sessions provide insights on what contexts are appropriate for Dittos, and issues around social acceptability and representation risk. The focus group results also provide feedback on visual design characteristics for Dittos. In the WOz study, teams participated in meetings with two different embodied agents: a Ditto and a Delegate (an agent which did not resemble the absent person). Insights from this research demonstrate the impact these embodied agents can have in meetings and highlight that Dittos in particular show promise in evoking feelings of presence and trust, as well as informing decision making. These results also highlight issues related to relationship dynamics such as maintaining social etiquette, managing one's professional reputation, and upholding accountability. Overall, our investigation provides early evidence that Dittos could be beneficial to represent users when they are unable to be present but also outlines many factors that need to be carefully considered to successfully realize this vision.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: meeting agents, human-AI-interaction, embodied agents, meetings, collaboration, video-conferencing, generative artificial intelligence, personalization, future of work

1 Introduction

Work is increasingly characterized by highly diverse and global teams, with team members spread out across different physical spaces and time zones. Especially since the COVID-19 pandemic, there is evidence that people are involved in more meetings with more people [12]. This demand for

scheduling meetings makes it more likely for conflicts to arise, due to impacted calendars, more flexible hybrid work schedules, and global time zone differences, besides the usual challenges of dealing with times when people are out of the office. These conditions make it difficult for people to attend all of the meetings which they are requested or interested to attend.

Meanwhile, generative AI technologies are enabling agents that can engage in human conversation patterns, as has been demonstrated in text chat services like ChatGPT [34]. Emerging from this, scores of new chatbots are arriving, such as ones to support specific tasks (e.g., OpenAI's GPTs [35]), and even some being built upon the personalities of recognizable celebrities to inform and entertain [28].

While such offerings, at present, interact only via text, we can expect that generative AI will soon provision embodied conversational agents of the sort that have been previously explored in research prototypes [7, 10]. This leads us to the question – what if embodied conversational agents can be modeled after anyone – *even you*? This is a provocative, yet intriguing concept, particularly for the future of work.

Imagine a future where, in situations when you are unable to attend a work meeting, you could send a personalized embodied agent to interact on your behalf. It could visually resemble you, sound like you, and possess the knowledge you would want to carry into the meeting. The agent would be able to partake in regular conversation with your colleagues, appearing much like a remote participant over a video call via voice, facial expressions, and body gestures. It could listen to what the team has to say, ask clarifying questions, answer questions on your behalf, and raise points that you would want your team to consider. Afterwards, the agent could share with you a summary of key takeaways from the meeting, personalized from your perspective.

If such personalized embodied digital agents were to exist, in what situations might people want to deploy or interact with them – if at all? What should the scope of their capabilities and responsibilities be? Furthermore, what effects would such agents have on team meeting dynamics? In this paper, we seek to explore how such ***personalized embodied agents could interactively represent a particular person in meetings*** (distinct from generic meeting agents that typically help to record, transcribe, and summarize notes in meetings such as Otter.ai[1] and Zoom's AI Companion[2]).

To investigate this space, we conducted two empirical investigations with participants recruited from within a large technology company. We first conducted six focus group sessions (n = 24). Through these discussions, we gathered initial reactions to the idea and identified key areas of opportunity and concern. We then conducted a Wizard of Oz (WOz) user study (10 groups, n = 39), in which working teams were recruited to experience meeting with two variations of personalized embodied agents that represented a known colleague over a video call. While both variations could express the knowledge, interests, and preferences of the colleague, the ***Ditto*** was designed to visually and vocally resemble the colleague whereas the ***Delegate*** presented as a representative acting on the colleague's behalf. Behind the scenes, the real colleague collaborated with a researcher to puppet the Ditto and Delegate and drive this illusion.

Based on survey and interview data from this investigation, we identified insights on how personalized embodied agents could affect the team's interaction and interpersonal dynamics, and discuss the implications it has on future work at the intersection of AI and meetings. Our paper contributes: 1) an exploration of personalized, embodied agents in an inherently collaborative setting (a meeting), which goes beyond most prior work that focused on dyadic interactions with agents; 2) a focus group that elicits people's perceptions of what it would be like to interact with a

---

[1]https://otter.ai/
[2]https://www.zoom.com/en/ai-assistant/

personalized, embodied and mimetic agent representing an absent meeting participant; and 3) a user study that compares two approaches to representing the digital agent, as a Ditto that looks and sounds like the absent meeting participant, or as a Delegate that serves as an advocate for the absent meeting participant.

2 Related Work

In this paper, we explore the design space for an embodied digital autonomous agent to help people participate in meetings even when they cannot be there live. This research fits within the scope of AI-Mediated Communication (AI-MC) wherein a computational agent operates on behalf of a person through generating messages to accomplish interpersonal communication goals [16]. We review prior works that explore asynchronous and digitally-mediated forms of meetings, identify issues around agents that represent people in meetings, and discuss concerns raised about deploying agents in an interactive social setting with people.

2.1 Enabling Absent Participation in Meetings

An increased demand for meetings coupled with a growing reliance on geographically distributed teams [47] makes it difficult for people to coordinate their schedules for meetings. Prior works have explored some avenues for overcoming incompatibilities in meeting availability (e.g., scheduling conflicts, timezone differences). While many tools have enabled people to interactively meet across the boundaries of space in the CSCW Matrix [20], there has been less innovation in enabling meeting *interactively* across time. It has become fairly common to record meetings for later review or document meetings in shared collaborative space (e.g., whiteboard, Miro), and LLM-driven extensions to these services are emerging in which non-embodied generic agents can listen in on meetings to provide summaries, outline action items, and answer people's queries about what has happened in meetings via a chat interface (e.g., Otter.ai and Zoom's AI Companion). Still, prior research has shown the value of the process of interacting in meetings, beyond just the documented outcomes [26]. In our research, we explore enabling a computer-mediated form of an otherwise absent group member to interact with others who are meeting synchronously.

Time Travel Proxy, by Tang et al. [46], enabled an absent colleague to send a video-proxy of themselves to a meeting via prerecorded videos. Those attending the live meeting would be able to see a video avatar of the time-shifted collaborator on a tablet device, listen to their prerecorded video messages, and record responses to those messages. This system afforded some degree of social presence of the asynchronous meeting participant, and also helped ensure that their perspective was considered by the live attendees. The KIZUNA system [31] afforded a similar sense of asynchronous presence in the domestic context of a time-shifted tele-dining system that enabled people to enjoy a meal together through sharing video recordings. They even explored the benefits of adapting the replay speed of their eating partner's dining video to match their own eating speed. Fender and Holz [13] created a system that allowed in-person interactions to be recorded and played back by a user in Virtual Reality (VR). With this system, a person could work undisturbed even while people visited to show them various objects. The person could then catch up on what they missed through immersive playback of the recordings.

While all of these prototypes enabled interacting with asynchronous partners through replaying videos, they did not allow generating any new interactions beyond what was prerecorded. In the current research, we wanted to explore how generative AI-powered digital representations of people could actually create new, spontaneous interactions with time-shifted meeting participants rather than being limited to what was prerecorded.

2.2 Virtual Agents and Digital Avatars in Meetings

Researchers have been exploring autonomously interactive digital agents in meetings since at least the 1990's. Nagao and Takeuchi [29] created an early version of an autonomous conversational agent that attempted to identify misunderstandings between conversants and used voice, facial displays, and head and eye movements to communicate with others. More recently, Cassell [8] reflected on embodied conversational agents that autonomously take turns in conversations and enact hand gestures. She argued that representation design choices affect how users attribute intelligence to agents and interact with them. She concluded that it makes sense to represent an agent as a human in cases where social collaborative behavior is key. In fact, according to the computers are social actors (CASA) paradigm [30], people generally have a tendency to apply human social rules and expectations to computers and agents when they exhibit human-like attributes. Several works since then have sought to outline and explore key design parameters for visual representation and conversational engagement of digital avatars and agents that participate in human conversation.

2.2.1 *Visual Representation*.

Agents and avatars can be represented in a variety of ways, which can shape patterns of communication, satisfaction, and perceived sense of presence. Avatar visual fidelity has been classified into three criteria [14, 25]: *anthropomorphism, realism, and truthfulness*. Anthropomorphism describes where the avatar is on a non-humanoid to humanoid spectrum. Realism maps to how realistic it is, ranging from having little detail to being highly photo-realistic. Truthfulness describes how dissimilar or similar the avatar looks to the target user. Shami et al. [44] explored preferences between abstract dot-, picture-, and cartoon avatar- representations of people in online corporate meetings, and found that the avatars in particular encouraged socializing and informal information exchange between attendees. Phadnis et al. [38] explored how realism affects acceptability of avatars for work meetings. By using animated GIFs, they investigated five levels of photo realism for a manager, known colleague and unknown colleague. In all scenarios, higher realism was favored, but fully realistic avatars were sometimes considered uncanny. Similarly, Inkpen and Sedlins [19] explored users' comfort with different styles of avatars for work communication, and found that people wanted avatars to resemble them, but be more stylized rather than photo-realistic. Groom et al. [15] conducted an empirical study on the effects of embodied agents' behavioral realism on users that compared three theories: *Realism Maximization Theory, Uncanny Valley Theory, and Consistency Theory*, and found that people responded best to an embodied agent when it was moderately realistic, with some inconsistent behavior, to avoid evoking the uncanny valley. Pan and Steed [36] conducted a comparison of how visual appearances as an avatar, video, or a humanoid robot affected users' trust. By successively presenting pairs of representations, and seeing which one participants elected to get advice from in answering a challenging question, they developed an indicator of which representation was trusted more. Overall, they found that people trusted video representations and the robot about the same, but had less trust in the generic avatar that lacked identity cues.

A particular representational choice of interest in collaborative settings is that of a "doppel-ganger"—a virtual representation that looks like oneself, but can act independently [1]. Doppel-gangers can enable thoughts and interactions specific to a particular person, such as one that is absent from a meeting. Lucas et al. [23] found that people operating a doppelganger of themselves in an immersive maze navigation task felt more motivated and engaged. However, more research is needed to understand how people interacting with the doppelganger feel, especially in a collaborative context. Hitada et al. [17] found that people found it eerie to interact with a doppelganger of themselves, especially when they directly interacted with it or witnessed it autonomously acting in VR. Their experiences indicate that it is especially important to consider how the person being modeled by the doppelganger reacts to its use.

Research so far suggests that realistic-looking agents enable engaging interaction with users as long as they do not trigger the uncanny valley. More research is needed in doppelganger representations that look like specific people, both by the people being mimicked and by those interacting with the likeness. Our research explores how new generative AI technologies add to the realism of autonomous agents, and compares a doppelganger representation with a more generic agent representation.

2.2.2 Conversational Interaction. How digital autonomous agents conversationally interact with people in meetings is another key design factor. The conversation dynamics of an agent have a strong influence on the quality of a meeting. We are interested in seeing how agents affect interactive, cooperative conversations in small group meetings. While much of the work on conversational agents so far has focused on text chat conversations, some of the learnings transfer to conversational interaction in general.

Clark et al. [11] conducted semi-structured interviews to understand what people want from Conversational Agents (CAs). They identified that conversation broadly has two purposes: *social* and *transactional* purposes. The central need in social purposes is to establish, maintain, and build a social bond. Transactional purposes comprise goal-oriented communication, such as exchanging information for a task. Their interviewees primarily thought of speaking with CAs for transactional purposes, as they were perceived to be unable to support social and emotional connections.

A more recent study looked at the communication strategies of humans interacting with AI agents as a team [49]. In a Wizard of Oz study, people were teamed with AI agents in a multiplayer, immersive, first-person game. They communicated with the AI agent via text chat, wherein the agent was either proactive or not. They found that AI agents should be proactive, balance efficiency with sociability, provide immediate responses to humans, and leverage patterns in repeated tasks to streamline the communication. These findings also seem relevant for designing a speaking conversational agent. For instance, prior research in human-robot interaction (HRI) and human-agent interaction (HAI) [3, 32, 45] have found that short response times are important for satisfying conversational interactions.

More recent research has applied generative AI to enable interactive conversations with digital agents. A digital sculpture with the appearance of a human face called Stanley applied traditional acting method prompts to a large language model [18]. They explored engaging audiences with spontaneous and emotional expressions to see how they would react to an artificial human. FurChat uses an LLM to drive a digital receptionist that is able to interactively converse with people through verbal and nonverbal cues [10], exploring whether the digital receptionist can evoke more natural conversation with people. While these early research efforts have yet to publish studies on their interactions, they are demonstrating how generative AI can create embodied conversational agents.

Taken together, we see how people's reactions to different representations are sensitive to differences in context and task. In our research, we want to explore how personalized and embodied meeting agents could interact autonomously with the other meeting participants, and examine the reactions of both the people being represented by the agent and those who interact with the agent in a meeting. Our research investigates the implications of the design choices identified in prior research that need to be considered in developing this agent.

2.3 Concerns Around Deploying Social Agents

Deploying autonomous agents in settings that involve interacting with people raises many concerns, particularly as they begin to reach the general public. The chatbot Tay that had to be shut down within a day of its public release in 2016 is a cautionary tale wherein interactions with malicious actors in a public deployment led it to generate inappropriate tweets [48]. More recently, the use

of first person pronouns in ChatGPT responses: "I won't be able to help you with that request," has sparked debate about anthropomorphic AI [42], and highlights the need for more thorough investigations into how algorithmic processes should be shaped to appear and interact with people.

McIlroy-Young et al. [27] termed AI models that are personalized to resemble specific individuals and spontaneously interact with people as *mimetic models*. They grouped uses of mimetic models into three categories: 1) as a means to prepare for a real-life interaction (e.g., rehearsing with a mimetic model of someone you want to present to), 2) as an end for accomplishing a task with the mimetic model instead of a person, and 3) as a copy of yourself to accomplish work the way you would (e.g., generating email replies), as a force multiplier for your productivity. They used a variety of scenarios to surface a number of subtle ethical and social considerations.

Lee et al. [22] termed AI models that digitally replicate people's appearance and behavior as *AI clones*. They used speculative scenarios to identify three major concerns with AI clones. Doppelganger-phobia described the fear that AI clones might exploit and displace the identity of an individual. Identity fragmentation described the concern that creating replicas of a living person threatens their own cohesive self-perception and unique individuality. Living memories described the risk that interacting with a clone of someone over time could distort one's perception of the person, or even create an over-reliance of interacting with the clone instead of the person.

Research has only begun to explore the ethical and social issues around deploying autonomous AI agents in social situations. Our research explores what concerns are specifically raised in the context of work meetings when an agent is used to represent a person who cannot attend the meeting live.

3 Concept and Terminology

In this paper, we were inspired to investigate the potential for leveraging **personalized embodied agents** to serve as a representative or proxy for a human individual who is unable to attend a meeting. We use the term **Source** to refer to the person that would be represented by the agent. Those that would interact with the agent in a meeting are referred to as **Collaborators**. Although agents could conceivably interact with other agents in the future, we limited the scope of our investigation to involve only one agent per meeting. As part of our investigation, we explored two versions of personalized embodied agents in work meeting contexts. The **Ditto** was a *mimetic* [27] agent that possesses the Source's knowledge and resembles the Source in appearance and sound. We also explored a **Delegate** as an alternative, which has the Source's knowledge, but does not mimic or resemble them in any way. The Delegate approach stemmed from the existing practice of sending another colleague to attend meetings on someone's behalf. Examples of a Ditto and a Delegate are pictured in Figure 1a and b respectively, and can be viewed in the accompanying video figure.

4 Investigation One: Focus Groups

As an initial step in our investigation, we hosted a series of focus groups with the goal to elicit people's initial reactions to the concept of Dittos, as well as to stimulate discussion to capture potential key opportunities and concerns. This could inform in what situations such agents may be used, and what the scope of its responsibilities should be.

4.1 Method

IRB approval was attained prior to conducting the focus group sessions. Participants were recruited via email within a technology company to form groups of 3-5 people with similar backgrounds. Each participant completed a consent form in advance, and received a $50 Tango gift card as a gratuity. Session times were negotiated based on their overlapping availabilities. All sessions

were hosted by two researchers online over Microsoft Teams and were recorded. To promote an active and lively discussion, all participants were requested to keep their cameras on and their microphones unmuted.

At the beginning of the session, the researchers provided a basic overview of the Ditto concept. This included that: (1) a Ditto would be fully autonomous and could attend a meeting on their behalf if they were not available, (2) it would be personalized such that it would visually resemble and sound like them (but not be a deepfake), (3) that the Ditto could interact by listening and speaking to other meeting attendees in natural language, and (4) that after a meeting they would in some way be able to learn from the Ditto what had happened in the meeting they had missed. Each participant was able to ask clarifying questions at this stage before proceeding.

A set of open-ended questions were prepared for the focus group sessions, pertaining to the concept of Dittos. Questions were sectioned into three main categories: ***sending their own Ditto***, ***speaking with someone else's Ditto***, and ***general thoughts/concerns***. The first category included four main topics, including: *types of meetings that they would/would not choose to send a Ditto to, factors that would influence their choice, what Dittos should/should not be able to say or do, and potential pros and cons for sending Dittos*. In the second category, questions included: *what would they want to be able to ask/tell someone else's Ditto, circumstances in which they would prefer to speak to a Ditto over the Source (or vice versa), and potential pros and cons of speaking to someone else's Ditto*. The last category was to ensure that any other miscellaneous comments could be captured. See Appendix A for the list of questions. While these questions established a loose structure for discussion, participants were encouraged to respond directly to each others' comments in active conversation, and the experimenters asked follow-up questions to gain deeper insight into points raised by participants during each session. A bottom-up approach to thematic analysis [4, 5] was taken to analyze the focus group discussions. A single researcher that was present at each focus group session sectioned the automatic meeting transcriptions into quotes, and manually generated initial codes (e.g., *respect, hierarchy, social dynamics, etc.*). The same researcher then collated them to conceptualize themes. These themes were reviewed and refined with the input from three others on the research team to arrive at the final set of themes below.

4.2 Participants

A total of 6 groups comprising 24 participants (8 female, 15 male, 1 preferred not to answer) were recruited for the focus group sessions. 1 participant was aged 18-24, 4 were 25-34, 4 were 35-44, 10 were 45-54, 2 were 55-64, and 2 preferred not to say. There was 1 group of 3, 4 groups of 4, and 1 group of 5 participants. We engaged with people from various backgrounds to gather a diverse range of perspectives. Their occupations spanned different roles including: *administrative worker, product/program manager, researcher, industrial designer, data scientist, engineer, etc.* Group 1 involved a **time-distributed group**, with many members in the UK and 1 member in Australia. The group shared that the Australian member would often record a video clip of herself talking to share with her teammates, since it is too difficult to meet synchronously due to the time difference. The team would watch her recordings at the beginning of their synchronous meetings together to incorporate her point of view. Group 2 comprised **administrative** workers. Groups 3 and 4 were mostly **high-level engineers** in product teams that develop productivity and collaboration applications. Group 5 involved people with experience in **avatars and virtual representations**. Group 6 comprised largely of people with **AI expertise**, as they actively work in this domain as part of their jobs. Participants will be referred to using their group number and letter A-E distinguishing the individuals among the group.

4.3 Key Takeaways

Overall the concept of Dittos was quite provocative and elicited a range of different initial responses. Participants responses leaned towards **conservative use cases**. From a Source's perspective of **sending** a Ditto, many felt that Dittos should play more of a passive listening role (e.g., all-hands meetings, a town hall) (2C, 3A, 6A, 6B, 6C) rather than be an active contributor in meetings (e.g., 1-1 meetings with a manager or an intern) (1C, 2B, 4B, 6A). They also mentioned that they would use Dittos in low-stakes rather than high-stakes meetings (1C, 1D). With regards to responsibilities, participants expressed a preference for Dittos for gathering information and reporting back (1D, 2B, 6A, 6D), sharing information by sticking to facts and scripts (1B, 1E, 1C, 2C, 6A), answering very simple and straightforward questions (1E, 3A, 6A), or offering very simple forms of help, such as programming and debugging support (6D). Many people rejected the idea of Dittos being able to brainstorm or give ideas (1B, 2B, 6A), *"think for itself"* (4D), or speculate (1C), represent their opinions as part of making decisions (1B, 1D, 3B), or build social relationships (i.e., should avoid small talk and gossip or discussing anything really personal) (1C, 3A, 3C). They were also concerned about protecting sensitive data, and worried that Dittos could pose a risk to information being leaked to the wrong parties (1B, 2C, 3A, 3C, 4A, 4B). From a Collaborator's perspective of **interacting** with a Ditto, people's boundaries sometimes softened. People could imagine reaching out to someone's Ditto in urgent situations (5B). They also mentioned that running a practice presentation (5B), bouncing off ideas (5B), or asking questions instead of disturbing the Source (1C, 2A, 4A, 5B) would be good reasons to interact with someone else's Ditto.

In many cases, participants were concerned about the **social acceptability** and downstream implications of sending and encountering Dittos. 5B commented that it would be annoying if a Ditto showed up to a meeting instead of the real person, and others (3C, 6A, 6B) worried that it could appear rude to others to send their Dittos. 3C imagined it could feel like a *"slap in the face"*, while 6B worried it would be perceived as *"blowing [someone] off."* However, some noted that Dittos would be socially acceptable due to illness (2A) or obvious time-zone incompatibility (1B, 4B), if their absence would be an obvious blocker to someone else's work (1A), or if it were extremely difficult to re-schedule a meeting (6C). Additionally, hierarchy and meeting size (2A, 3A, 6C) would be factors in determining the appropriateness of using a Ditto. Some participants said they would not send a Ditto to meetings with higher-level managers (1A, 1B, 1C).

Participants' comments also delved into key opportunities and risks. Some opportunities included that in their absence, it could ask questions and generate personalized summaries that could help make progress in their work (2C, 3D). Others mentioned that it would have superior attention, memory, and multi-tasking capabilities to help them collect and retrieve important information (2A, 2B, 2C, 4D). **Reputational risk** was a key risk mentioned about using a Ditto (3A, 3C, 3D, 4D, 5A, 6A). Beyond fears surrounding potential rudeness, some participants were concerned that it may have inappropriate behavior, which could lead to negative projections on their real selves.

Group 6 considered alternate **visual representations** for Dittos with respect to reputational risk. 6A believed that cartoon-like representations could help bypass *"... the uncanny valley [and] resolve some of that reputational risk..."* and 6B thought it would help people *"distinguish between is this teleoperated... or is this... an AI operated kind of thing."* However, 6B also believed that some resemblance would be useful: *"... it could be a bit distracting for other people in the meeting if it was really anything... [I'd] have to keep track that this spirit animal is [Source] and this thing is whatever."* 6C added, *"I wouldn't want a deepfake... you don't actually know if the person's there or not right at some point."* Ultimately, this group arrived at a shared opinion that Dittos should likely resemble the Source, but be clearly visually distinguishable from the Source or a teleoperated avatar in meetings.

In contrast, others questioned if Dittos should look like them at all. 1C stated *"I would reframe the whole idea almost cause I think the idea of a Ditto and it being a representation of me feels icky, especially with all the stuff that's happening in AI at the moment."* Three groups (3, 5, 6) wondered if a Ditto should instead be trained to serve on their behalf as a Delegate. 3C contrasted Dittos with the **practice of sending somebody else** from your team to go in your place:

> *"... you don't ask them to go and wear a mask. Imagine you could get somebody on your team to put on a little rubber face and go in looking like you. So when it's looking like you, you're really asking it not to be your delegate, you're asking it to be you, which I think is weird. I think maybe I'd rather it was my delegate and that everybody understood it was something to do with me and it was a limited delegate."* (3C)

3D agreed with this idea, saying *"I would feel more comfortable if it didn't look like me because I have that kind of plausible deniability when it goes and makes a mistake that I can say - oh yeah, it wasn't really representing me. It's just like I sent a proxy..."* Comments from group 5 also resonated with this concept. 5C stated *"I think I would want my Ditto to very clearly show up as NOT me, but a representative of me."* and 5D agreed, *"A Ditto could be your chief of staff. But it's super clear that it's not me or pretending to be me, but it's somebody that I have a relationship with that I trust, right?"* Group 6 also questioned the idea of the Ditto visually resembling the Source. 6D began contemplating, *"I think if it's visually similar to me, it's representing me wholly versus if it's different visually... "* Building on that, 6C added *"I think that would make me more comfortable too - like think of it as my someone who I've trained to try to represent me as opposed to actually being me.* However, 6C saw nuance there: *"I think there's a lot of pros, like there's a lot of good aspects to having it be me in terms of how people would talk to it. But there is that risk that, I would feel that there would be a much higher threshold of quality and capability."*

4.4 Reflections and Next Steps

Conversations with the different focus groups brought to light a broad spectrum of considerations, opinions, opportunities, as well as concerns. Through discussions with one another, participants identified some potential **benefits** of using Dittos. They believed it could help their teams overcome otherwise impossible scheduling conflicts, help them bypass or mitigate inefficient meetings, serve as an effective aid for collecting and summarizing information, and overall help them or even their colleagues' save time and effort. At the same time, they saw the Ditto concept as being accompanied by several possible **risks**. Many groups expressed concern about the social acceptability around using a Ditto. Contemplating this from both Source and Collaborator perspectives, they worried that it could trigger offense. Additionally, the groups were unsure what expectations they could set regarding the Ditto's performance, and were notably worried about Dittos saying incorrect things, or behaving inappropriately and causing reputational harm. To limit this risk, they envisioned using Dittos in low-stakes situations where they could tightly control what the Ditto would be able to say. In line with this, some groups pushed back on the idea of the Ditto resembling the Source, since they felt this would heighten the risk of reputational harm. Nevertheless, a couple of participants speculated that visual resemblance could be beneficial to the group's interaction dynamics, so long as it would still be visual distinguishable from being a real person. Indeed, the focus groups called into question whether a personal embodied agent should be visually representative of the Source, or if it should rather present itself as a separate Delegate entity.

Overall, the focus group sessions shed light on key considerations regarding Dittos. However, we recognized that their opinions were underpinned by many assumptions about its capabilities and performance. Since it can be particularly difficult for users to imagine an experience [33], we therefore saw that it would be beneficial to conduct an empirical investigation in which people

could have a hands-on experience interacting with a Ditto to gain deeper and more authentic insights. This led us to a second investigative phase, with a particular focus on contrasting the merits of a personal embodied agent that takes on an identity that resembles the Source (a Ditto), compared to an agent that presents itself as an entirely separate and distinct personality serving on behalf of the Source (a Delegate).

5 Investigation Two: Wizard of Oz Experiment

Motivated by the findings from the focus groups, we designed a Wizard of Oz (WOz) user study to gain empirical insight into how personalized embodied agents may impact meeting dynamics.

5.1 Method and Rationale

The study received approval from our organization's institutional review board (IRB). Although we were in the process of building a version of an agent with ChatGPT, we chose to use a WOz user study methodology to get early reactions to the concept that could help shape its development. Our WOz implementation was based on the use of avatars in a video conferencing application. In considering this approach, we recognized that it would raise unique challenges regarding establishing a convincing *simulation of interacting with an autonomous agent* distinct from a common experience of interacting with someone who is using an avatar during a meeting. The study protocol and technical setup would need to be designed with care to successfully evoke a plausible interaction with a generative-AI powered autonomous agent.

Our rationale for using a WOz approach was governed by weighing some key benefits and drawbacks. Three key advantages convinced us to use this approach. Firstly, it meant we could get early feedback on the concept while we were still in the process of developing it. Secondly, we could elicit people's responses towards an "ideal" form of a Ditto or Delegate, decoupled from temporary present-day technical limitations on ChatGPT's performance with regards to latency and accuracy. Thirdly, the approach minimizes risks for study participants. Allowing the person represented by the autonomous agent to actually assist in simulating their agent helps ensure that anything communicated in the context of the study aligns with their wishes and does not leak sensitive information. Our main concern with this approach was carrying out the illusion so that the Collaborators did not conclude that it was simply a person puppeting an avatar. We detail the measures taken to guard against this in Sections 5.1.2, 5.1.3, and 5.1.4.

5.1.1 Participant Recruitment: Ten teams comprising three to four work colleagues (n = 39, 18 male, 21 female) were recruited via internal email lists at a technology company. For the study, group members needed to be familiar with one another, since we wanted to explore how an agent that resembles a specific person affects the meeting experience. Eight participants were 18-24 years old, thirteen were 25-34, seven were 35-44, eight were 45-54, two were 55-64 years of age, and one preferred not to specify. Their roles varied, including software engineer, product/program manager, researcher, business administrator, media/video producer, UX researcher, scientist, and designer.

One participant per group (4 male, 6 female) was recruited to be the **Source** and engaged in the meeting as the Ditto/Delegate. This person recruited the other group participants outside their direct reporting line to avoid the possibility of undue coercion. Sources participated **in-person** for approximately 1.5 hours and were each compensated with a $75 USD gift card. The remaining participants were **Collaborators** (14 male, 15 female) who attended the meetings **remotely** over video conferencing call. They were each given a $50 USD gift card for 1 hour of remote participation. Each team member provided their consent via an online form prior to participating. Participants will be referred to through their session number and letter indicating if they were the Source (A) or a Collaborator (B-D).

5.1.2 ***Technical Setup and Materials:*** The study was organized into two parallel streams: *in-person* and *remote*. One to two researchers handled the in-person stream, and another researcher handled the remote stream.

The in-person study stream for Sources was held in an experiment room with a dedicated desktop computer, two monitors, mouse, keyboard, webcam, speakers, and headset with a microphone (see Figure 1c). A video conferencing tool, Microsoft Teams[3], was installed on the computer with the Avatars plugin[4], in addition to VoiceMod[5], a real-time voice changer. For the Delegate condition, a single gender-neutral avatar was designed and used for this condition across all study sessions. For the Ditto condition, prior to each user study session, a researcher manually crafted a digital avatar to resemble the Source-participant (based on a submitted profile photo), using the native avatar editor. The decision to use an avatar as the basis of the visual representation was motivated by findings from the focus groups. Avatars would satisfy the desire for a more cartoon-like representation, rather than a photo-realistic representation (e.g., a deepfake). However, to make it visually distinct from a standard teleoperated avatar, both Ditto and Delegate avatars were rendered in black and white (see Figure 1a and b).

Collaborators attended the study session stream remotely from their own computers using a standard version of Microsoft Teams. Two sets of consent forms, surveys and interview questions were prepared with respect to the two distinct participant roles (see Appendix). The consent forms and surveys were distributed online via Microsoft Forms. Video and audio recordings were captured during the experiment and interview sessions using Microsoft Teams' in-built recording features.

5.1.3 ***Remote Study Procedure:*** A researcher greeted the Collaborators in the video meeting room as a host at the scheduled time and provided link to the online survey with pause points. The researcher briefed them about the concept of Dittos and Delegates and the overall study procedure. As part of the WOz procedure, the Collaborators were informed that given advancements in AI, they would experience speaking and interacting with working prototypes of both types of autonomous meeting agents who would serve as proxies for their absent colleague, in two distinct back-to-back meetings. To enhance the necessary deception, they were told that this colleague would simultaneously be meeting with a researcher in-person to observe the meetings and provide feedback on their agents' performances. After answering questions raised by the participants, the researcher confirmed with them that their meeting software was configured consistently (i.e., Gallery mode with Fit-to-Frame activated, so that each attendee would be in full view), and cued the other researcher to introduce the first agent into the meeting. The order of the conditions was counterbalanced across the study sessions. When the agent entered the meeting, it introduced itself. The host then described the meeting topic. Two topics were prepared, one for each meeting, which were presented in consistent order:

- **Group Travel Plan:** Create a proposal for a group trip to celebrate a major team accomplishment. Decide on the location, lodging, and main activities—assuming a generous budget.
- **Holiday Party Plan:** Create a proposal for an in-person holiday party for your team. Decide on the venue, food and beverages, and entertainment—assuming a generous budget.

After clarifying the topic, the researcher proceeded to turn off their camera and mute themself, allowing the participants to discuss the topic amongst themselves. The host notified them when one minute remained. After a total of five minutes, the host researcher unmuted and reappeared in the meeting room, and asked for someone to verbally summarize their proposal. At this point, the agent left the meeting, and the Collaborators were asked to complete the corresponding survey

---

section. Participants repeated this meeting process with the next discussion topic using the second agent type. After both meetings, Collaborators were requested to complete the remainder of the survey and participate in a group exit interview with the host. At the end of the session, the host checked whether the Collaborators believed they were interacting with an agent or whether they thought it was a human (i.e., manipulation check) before debriefing them about the deception, and thanked them for participating.

*5.1.4   **In-Person Study Procedure:*** A researcher greeted the participant in the dedicated experiment room 20 minutes before the scheduled virtual meeting time and provided a walkthrough during which the participant could ask questions. The walkthrough detailed the concept of Dittos and Delegates, the technical setup, the study procedure, as well as the study tasks. As part of this, the participant was given time to see and familiarize themselves with their pre-prepared personalized Ditto-avatar. Seated in front of the webcam, the participant was asked to experiment with moving their head and face to observe how the avatar mirrored their head movement and facial expressions, but not their larger body movements nor hand gestures. The researcher explained how they would work alongside the participant to "puppet" the different agents. Finally, participants were asked to contemplate the meeting topics in advance and note their ideas and preferences, as well as start their survey (pausing at the given breakpoints). The participant could then relax until an agent was cued to join the meeting, either as a Ditto or a Delegate depending on the condition:

- **Ditto Condition:** The personalized Ditto avatar was shown on a plain white virtual background with the text "[First Name's] Ditto." The voice changer was used to modulate the participant's pitch downwards to level 40 (50 was neutral), so that the Ditto's voice was recognizable but not identical to the Source's voice to contribute to the illusion of being an AI agent. The Source sat in the main seat in front of the webcam to puppet the Ditto avatar, while the researcher sat beside them, outside of the view of the webcam. When the Ditto entered the meeting room and was prompted to introduce itself, the Source read the following statement out loud *"Hi, I'm [First Name]'s Ditto. I'm representing [pronoun] interests because [pronoun] couldn't be at the meeting today."* During the meeting, Sources were instructed to respond much as they normally would. This included speaking in first-person and relying on their existing knowledge. However, they were asked to speak in full, discrete sentences (rather than in sentence fragments punctuated by filler-words) and to refrain from laughter as best they could. This was to support the illusion of an AI agent, which would not be expected to speak with these characteristics at this point in time. To support this, the researcher kept the microphone muted unless the participant indicated with a small gesture that they wished to speak.
- **Delegate Condition:** The Delegate avatar was shown on a plain white virtual background with the text "[First Name's] Delegate." The voice changer was used to modulate the pitch downwards to level 35 (to convert the researcher's voice to a gender-neutral pitch). The researcher sat in the main seat in front of the webcam to puppet the Delegate avatar. The Source sat adjacent to the researcher facing them in order to be able to give cues for when and how to respond (see Figure 1). When the Delegate entered the meeting room and was prompted to introduce itself, the researcher read aloud the following statement: *"Hi, I'm [First Name]'s Delegate. I'm representing [pronoun] interests because [pronoun] couldn't be at the meeting today."* During the meeting, the researcher referred to notes the Source made during the walkthrough as well as any additional prompts or comments (with the mic muted) from the Source to puppet the Delegate in the meeting.

The Source was prompted to complete survey questions after each meeting, and was asked to complete the remaining survey questions after both conditions were complete. Following this,

Fig. 1. Wizard of Oz (WOz) Study Setup: a) Screenshot of a participant's **Ditto** and its corresponding background. b) Screenshot of a participant's **Delegate** and its corresponding background. c) Photo depicting the **in-person WOz study setup** in the Delegate condition. A Source-participant sits next to the experimenter and provides their input while the mic is muted. The experimenter puppets the Delegate during a meeting with the Collaborators over a video call. In the Ditto condition, the Source-participant sits in front of the webcam to puppet the Ditto. The Collaborators are each attending the meeting from a separate remote location.

the Source participated in a semi-structured exit interview with the researcher, separate from the remote Collaborator participants. At the conclusion, they were thanked for their involvement in the study and were informed that their colleagues were debriefed about the procedure.

5.2 Data Collection and Analysis

Surveys and semi-structured interviews were used for data collection (see Appendix B). The surveys comprised demographics questions, 7-point Likert questions (1 = Strongly Disagree to 7 = Strongly Agree) spanning participants' preferences, perceptions, and feelings towards the meetings and the respective agents (e.g., satisfaction, trust, social presence, etc.), and open comment sections. While the majority of questions were custom, those pertaining to social presence were based on the Networked Minds Social Presence inventory [2]. For the Collaborators, the interviews probed thoughts or feelings evoked by each agent, how the agent's knowledge and behaviors corresponded with the Source's, and their impression of the social dynamics in the meetings. For the Sources, the interviews questions probed how they felt their Ditto and Delegate agents were able to engage in the meetings, how they were treated, preferences they had between the two different conditions, situations they would consider using the agents, opportunities for improvement, and any other comments they had about the concept and experience. In both cases, the interview questions formed a structure for the interview, but follow-up questions were asked to gain clarity and elicit more detailed explanations of points raised by the participants.

To compare results of the Likert questions for the within-subjects condition (Delegate vs. Ditto), we used Wilcoxon signed-rank tests. We applied a continuity correction to account for the discrete nature of our data, as well as the small sample size. As a measure of effect size, we used Cliff's Delta given its suitability for ordinal data and because it does not assume a normal distribution. In reporting these results, we present the Wilcoxon test statistic (W) and p-value (p), and the Cliff's Delta (denoted as d). To compare results of the Likert questions to the hypothetical neutral median, we used one-sample Wilcoxon signed-rank tests, and again applied a continuity correction. As a measure of effect size, we computed the Rank-biserial correlation between observations and the

hypothetical median. In reporting our results, we present the Wilcoxon test statistic (W) and p-value (p), and the Rank-biserial correlation (denoted as r). Open-ended survey responses and group interviews were analyzed through a reflexive thematic analysis approach [4, 5]. The researcher who consistently hosted the remote Collaborators and the researcher who consistently hosted the in-person Sources led the analyses of their respective survey and interview data. They followed an inductive coding approach by manually reviewing and coding the interview transcriptions and written survey responses, as well as conceptualizing themes. Over the course of three meetings, these researchers as well as two others on the team reviewed, discussed, and refined the themes across both streams in an iterative manner to arrive at the final set presented in this paper.

6 WOz Study Results

In this section, we report results from the WOz study, including findings distilled from the survey and interviews with participants regarding their perceptions of the personalized embodied meeting agents, the interaction dynamics of the meetings they had, and considerations for how such agents could impact their relationship dynamics in the future.

6.1 Personalized and Embodied Agents in Meetings

Participants in our study rated the meeting experience with agents positively across several measures, demonstrating that users are open to having personalized embodied agents represent themselves, or their colleagues in meetings. Survey results revealed that 80% of Source participants and 77% of Collaborators agreed with the statement *"I was satisfied with the overall meeting experience"* regardless of whether the meeting included a Ditto or a Delegate. This resulted in an overall median of 6 (7-point scale from strongly disagree to strongly agree), which was significantly above the neutral median value (Mdn = 4) ($W = 102.0$, $p < 0.001$, $r = 0.87$).

Collaborators also indicated that *"It was beneficial to have the [Delegate, Ditto] in the meeting, INSTEAD of having my colleague skip the meeting altogether"*, with the majority of participants agreeing with this statement (76%). These user ratings (Mdn = 6) were significantly above the neutral median value (Mdn = 4) ($W = 115.5$, $p < 0.001$, $r = 0.82$). Similarly, Collaborators disagreed with the statement that *"It would have been better if the [Delegate, Ditto] had not been there"*, with the majority of participants disagreeing with this statement (67%). These user ratings (Mdn = 2) were significantly below the neutral median value (Mdn = 4) ($W = 95.0$, $p < 0.001$, $r = -0.72$). Participants also commented that having a meeting agent was better than their colleague sending text bullets: *"It did remind us to think about her opinion rather than have some list somewhere to reference async"* (2D).

*6.1.1 **Ditto versus Delegate**.* We examined whether there were differences in users' ratings towards agents in meetings depending on whether it was a Delegate or a Ditto in the meeting. Collaborators' overall satisfaction with the meeting was significantly higher in the Ditto condition ($W = 47.5$, $p = .029$, $d = .295$), but no significant difference was found for the Source participants ($W = 4.0$, $p = .850$, $d = .010$).

Source participants were asked about their preference to send either a Ditto, a Delegate, or neither to a meeting they needed to miss, while Collaborators were asked whether they would want a missing colleague to attend using a Ditto, a Delegate or neither. The majority of Source participants (7/10, 70%) favored sending a Ditto to a meeting they could not attend (Ditto: 7, Delegate: 1, Neither: 2), and the majority of Collaborators preferred having a Ditto in a meeting (22/29, 76%) in situations where their colleague could not attend, (Ditto: 22, Delegate: 5, Neither: 2).

6.2 Interaction Dynamics

We also analyzed the study results to better understand what effects personalized embodied agents might have on feelings of presence, trust, and social dynamics among people in meeting scenarios.

6.2.1 Presence. Several questions in the survey probed whether the Dittos and Delegates created a sense of presence of the Source amongst the meeting participants. While both agents were able to establish a sense of presence, the Ditto's visual and vocal resemblance to the Source helped create stronger feelings of presence than the Delegate. For the statement *"I had the feeling that the real person (my colleague) was attending the meeting when their [Ditto, Delegate] was there"*, Collaborators ratings were significantly higher in the Ditto (Mdn = 5) condition compared to Delegate (Mdn = 2) condition ($W = 29.0$, $p < .001$, $d = .572$). Although similarity to the Source was viewed positively by Collaborators, most of them (20/27[6], 74%) indicated that the agent should still be visually distinct from a live person puppeting an avatar, so that they would know whether they are interacting with a live person or an agent.

We also wanted to understand how much the embodied agent triggered thoughts of the missing colleague. Collaborators agreed with the statement *"Having my colleague's [Ditto, Delegate] in the meeting helped me think of them,"* with Dittos (Mdn = 7) being rated significantly higher than Delegates (Mdn = 6), ($W = 9.0$, $p = .003$, $d = .350$).

Generally, both Dittos and Delegates facilitated high degrees of social presence, according to questions from the Networked Minds Social Presence scale [2]. Collaborators agreed that both types of agents paid *close attention to them* (Ditto: $Mdn = 6$, Delegate: $Mdn = 5$), were *not easily distracted* (Ditto/Delegate $Mdn = 2$, reverse scale), that *[Sources'] thoughts were clear to them* (Ditto/Delegate: $Mdn = 6$), and that the Ditto/Delegate *behavior was in direct response to the group's actions* (Ditto/Delegate: $Mdn = 6$). In sum, these measures indicate that meeting agents are capable of facilitating presence in virtual meetings. This was also evident in the participants' comments: *"even though it's a Ditto, I feel like [Person] was there [herself]"* (2C). *"I think both helped remind me of [Person] and her thoughts and what would she think?... Like her presence was still a good reminder"* (10D).

Interviews with Collaborators highlighted that vocal-resemblance to the Source may be a key factor in evoking presence: *"I think it's nice having a Ditto that sounds so similar to the real person, it gives a 'they are actually in the meeting' feel, which would make others more comfortable to listen to feedback from the Ditto"* (5B). 9C expressed *"... this isn't [Source], but it sounds like her, where I think that alone makes you think... she's here..."* (9C).

6.2.2 Trust. For both the Ditto and Delegate conditions, Collaborators felt that they *"trusted what the [Ditto, Delegate] had to say"* (Ditto Mdn = 6, Delegate Mdn = 6), however, trust ratings were significantly higher when interacting with Dittos as opposed to Delegates ($W = 30.5$, $p = .016$, $d = .276$). This was also echoed in the interview feedback from the participants. 8C stated, *"it's like I could trust the responses [from the Ditto] a little bit more, even though it was the same information."* 5B reasoned that their trust was influenced by general familiarity, as well as vocal resemblance: *"because the Ditto one sounds like [Source], that's definitely something I was more receptive to rather than the Delegate [...and...] because we don't know [Source]'s Delegate, it makes it more difficult to trust [it]."* Similarly, Sources felt their colleagues trusted what their agents had to say (Ditto Mdn = 6.5, Delegate Mdn = 6), however, there was no significant different in the Ditto versus Delegate ratings ($W = 9.0$, $p = .430$, $d = -.170$).

Trust is a complex issue, and there are many factors which could impact users' feelings of trust. In this study, the conversation topics were relatively benign, and as such, there was low-risk in

---

[6]One group did not answer this question.

trusting (or not trusting) what the agents were saying. Despite the positive trust ratings, during the interviews, some participants made comments highlighting concerns related to trust. For example, 6A worried that people could be overly suspicious: *"...[with an AI model], they would just be like, oh, that's just pulling from the ether - like they're not actually wanting that."* Collaborator 10D found herself questioning the validity of the Ditto's statements: *"I was always like, hey, is this actually what [10A] wants? Or is this like an extrapolation of what you know the AI thinks she wants? Did the Marie Antoinette idea just actually come from like [10A]? [...] Or is that like something that came out of thin air?"* Unfamiliarity with the Delegate may also have contributed to trust concerns. For instance, 1B asked, *"Why is this random person stepping in for our team member to try and build an activity that they're not going to participate in?"* Similarly, 9C stated *"it almost felt like a stranger was in the meeting, and the comments threw me off enough it derailed my thoughts,"* and 5B said *"...because we don't know [Source]'s Delegate, it makes it more difficult to trust that Delegate."*

The meetings Sources would consider using agents was predicated on their level of trust in their agent's performance. Sources commented that they would most commonly consider using an agent in meetings where the agent would primarily listen, gather and summarize information (3A, 5A, 8A), participate in structured meetings with clear agendas, objectives, and topics (3A, 4A, 6A), or in meetings where the agent could stick to a script (4A, 6A, 7A, 8A, 9A, 10A). Others believed it should stick to facts (2A) and should neither be used for persuading nor arguing with others (3A, 4A). Generally, agents were seen to be more suited for less important, low-stakes meetings (1A, 10A).

6.2.3 **Decision Influence:** Although some Source participants expressed concern that others *"may not take [Delegates' statements] as seriously"* (3A) and that if it came to a decision point *"the group probably would not consider AI as an equal vote... probably be like a .5 vote..."* (6A), survey responses contradicted this notion, and both Sources and Collaborators felt that the opinions of the agents were taken into account in the final decisions. Collaborators agreed to the statement *"I felt that the views/opinions of my colleague represented by the [Ditto, Delegate] were considered in the final group decision"* (Ditto Mdn = 6, Delegate Mdn = 6). Additionally, Collaborators' ratings for Dittos were significantly higher than for Delegates ($W = 11.0$, $p = .003$, $d = .407$, "medium"), which suggests they felt that Dittos had a stronger influence on the group's decision. Sources also strongly agreed with this notion (Ditto Mdn = 7, Delegate Mdn = 6.5), but there was no significant difference in their ratings for Dittos versus Delegates ($W = 4.5$, $p = 1.00$, $d = .080$).

Interview comments from several Collaborators supported the idea that agents had a strong influence on the group's decision, and in some cases, participants felt that agents' opinions were even weighed more heavily than the people in the room. However, they were divided about which representation had a stronger influence. 2D felt that *"the Delegate's opinion seemed very strong, and thus had greater weight...."*, whereas 10D said *"I do wonder if the Dittos' opinions are almost taken more strongly than everyone else's, since everyone is sort of trying to accommodate the Ditto's ideas."*

6.2.4 **Conversational Flow**. Many participants were sensitive to factors that affected the naturalness of a conversation, including delays, the use of first-person or third-person language, and the agent's engagement model (e.g., if it waits to be called upon or interjects). In fact, speech delays were the main concern for participants who preferred neither agent (6D, 7D). According to 7D, *"Once people learn there is a delay to respond, we wait and give it time. This can potentially slow down a meeting...",* and the inability to quickly bounce ideas off of the group members led 6D to believe that refraining from the use of agents would *"allow more efficient communication."* The use of first-person and third-person language was also a contributing factor. Several Collaborators disliked the Delegate's use of third-person language. 9C found it *"awkward"* and commented that it *"felt more like some random stranger that knows nothing about [Source] is in the room..."* In some cases, it eroded feelings of trust: *"Ditto is more believable whereas Delegate has a bit too formal language*

and difficult to convey concise information in a live meeting," (6A). The difference in perspective likely contributed to the Collaborators indicating that Dittos, more than Delegates, helped make the Sources' thoughts *"clear"* ($W = 39.0$, $p = .011$, $d = .276'$), and appeared to pay closer attention to them ($W = 46.5$, $p = .025$, $d = .295$). However, some participants preferred the indirection of the Delegate and that is was more transparent about being an AI agent : *"I just liked how it [Delegate] was framing itself as an assistant rather than impersonating the person."* (7B).

Some participants noticed distinct conversation patterns resulting from treating agents differently than humans. 9A remarked that, *"...when the agents were talking, they were all quiet to listen to the agent, which doesn't necessarily happen in real life..."* 7A's comments resonated with this: *"I think there's an element of where '**there's something talking to us**' as opposed to another person in the conversation."* 3A also felt as though the Collaborators did not acknowledge the Delegate as they would a regular person: *"...it's almost like they didn't even consider providing input back to the Delegate. It was like someone just tossed in this idea, and they were reacting to that idea. But not like a person had provided it."*

6.3 Relationship Dynamics

The WOz study experience evoked thoughts and considerations regarding new social customs and practices that would need to develop to accomodate the use of personalized embodied agents.

*6.3.1 **Social Etiquette and Protocols**.* Some participants expressed concern that sending an agent to a meeting instead of attending themselves (6A, 7A, 8A), could be perceived as insulting. 6A likened it to a movie scene where Iron Man sends his empty suit to Spiderman, saying *"...if it were important, I would be there in person."* Others considered matters of reciprocity. 8A stated, *"If one person uses an agent, why not do the same?"* Similarly, 7A questioned *"Why be at a meeting if the rest of the participants are Dittos?"* These comments indicate that a Source's presence would be valued more than their agent. Interestingly, power hierarchies might also impact this perception. According to 6A, a person of higher rank may be open to interacting with an agent, especially if it could result in time-savings: *"It would be a risk to have managers sending Dittos to one on ones with their subordinates, but I'm sure it would be hugely helpful for my manager to get a like twice a week – like what's [this person] doing?"*

*6.3.2 **Managing Reputational Risk and Accountability**.* A common concern for Sources was that agents' inappropriate actions could damage their reputation. According to 2A, there is a risk: *"...if the Ditto does not do a good job representing me, it may leave the wrong impression in others working with me."* 7A shared a similar concern: *"I guess I'm a bit wary of the idea of it thinking for itself... "* Some Collaborators similarly identified the risk of negative transfer (6B, 6D, 8D). 6B stated, *"I never knew this side of this person... there is this danger, especially in big meetings, of people just thinking that this is the personality of the person."* In line with this, 8D commented *"if it was someone I didn't know. Like, I think it might affect how I feel about them."* In the worst case, an agent could trigger anger at the real person. For instance, despite some frustration from the meeting, 6D caught herself thinking: *'I wouldn't want to go to the person later and be like man, you were a d*** in that meeting, like you did not listen to me!"*

Concerns regarding accountability were also raised by Sources and Collaborators. Some Sources preferred the Delegate's visual representation for this reason. According to 4A, *"I'd rather a Delegate represent me... because there's still a little bit of distrust in the information that was provided... I'd rather have someone who speaks in third-person, otherwise, their action would hold accountable to me, right?"* 9A echoed a similar sentiment, stating she would prefer the Delegate go on her behalf for the *"separation"* aspect of it.

Several participants also desired follow-ups or escalation mechanisms (3B, 5C, 6B, 7B, 9A, 10A) as strategies to manage accountability, particularly in cases of uncertainty. Source 9A recommended that if Collaborators were to disagree with an agent, the agent should *"make a note,"* and offer to *"circle back"* when the Source is available. 3B stated *"even after making this decision, I need to double check with a person - I'm not sure that's something he agreed."* Sources were also interested in follow-ups to understand how their agent represented them. As 10A noted, *"I would like to receive a post-meeting report/recording so I could double check that my Ditto represented me accurately."*

6.3.3 ***Managing Social Impressions****.* In interviews, some Sources expressed a desire to have control over their agents, particularly with respect to their personality and behaviors (4A, 6A, 8A, 10A). 10A aptly pointed out: *"So I don't think you have a one size fits all right? I don't think I can have a Ditto that can go to every meeting."* 10A went on to say that her behavior around executives would be different than her behavior among close colleagues. 8A's comments resonated with managing their agent's behavior: *"I'd want to be able to customize how much it talks versus doesn't talk... so more the functionality of like it going to the meeting... "* 6A also envisaged having some control over their agent, and elaborated on how he could imagine making such changes:

> *"...in theory, like if you could make your own character... I'd just put charisma all the way up to max, and being able to control that meeting to meeting might be nice. Like, if I'm meeting one on one with the person I work with all the time, my affect changes compared to like an all hands meeting. So if I could prep the Ditto just saying like - don't be funny - this meeting is not the place, and whether that is me having a one on one with the Ditto or sliding a bar with Sims-like personality bar before the meeting, I think all of that would be helpful. If I were to train an ideal model, I would say before every meeting I would get a prompt saying - like - how do you want me to act?"* (6A)

While 4A saw possible upsides to tuning an agent, she was disconcerted by the idea that people might prefer her agent more than her, *"... if another colleague, who might be more senior than me, respected my agent a little bit more than me, then I'd feel a little bit offended, right?!"*

6.3.4 ***Difficulty Arguing with Agents****.* Some Collaborators' perceived the agents to be insistent and unyielding in their views, yet had difficulty 'arguing' with them. 2D stated, *"the delegate was very much like this is where I wanna go... not even considering how we could do some combination of the idea."* 6C complained that speaking to the Ditto was *"like talking to a young child who's not considerate of other's perspectives."* At the same time, several participants expressed reluctance towards 'arguing' with the agents. *"Arguing with... a bot feels really, umm like banging my head against the wall right?... I don't wanna start a conversation with a a bot, so... fine, let's just go with... I don't wanna argue with it, AI. "* (5C). In another group, 10B said: *" I felt almost like it'd be harder to push back, even though I would want to."* 3A also mentioned *"It's hard to argue with a robot."*

What makes these comments striking is that the Collaborators were actually participating in a WOz experience. Their reactions, impressions, and opinions were in response to the Sources' authentic responses as they helped to puppet the agents, as opposed to predefined scripts. It is possible that people's preconceived notions of the capabilities of AI agents and large language models may have influenced how meeting agents' statements and behaviors were interpreted. Furthermore, these expectations may have important downstream implications when striving for a balance in power between conversants in hybrid human-agent meetings.

6.3.5 ***Debrief****.* Finally, the Collaborators were told that the Ditto and Delegate were actually being controlled by their colleague. When asked whether they suspected this during the study, most reported being convinced by the WOz experience. Only two of 39 thought they were actually interacting with humans instead of agents, and seven voiced some suspicion about human involvement.

7   Discussion

Results from this research help us understand the main opportunities, risks, and effects personalized embodied agents could have in work meetings. Overall, our investigation provides early evidence that Dittos could be beneficial to represent users when they are unable to be present. Nevertheless, there are still many factors that need to be carefully considered to successfully realize this vision.

7.1   Dittos as a Novel Form of AI-MC to Provide New Opportunities for Collaborative Interactions

We proposed a **novel concept of leveraging generative AI to enable personalized embodied meeting agents** that serve as interactive proxies for people who are otherwise unavailable to attend. Both the focus group and WOz investigations highlighted that people are excited about the **opportunities such meeting agents can afford in the future of collaborative work**. While earlier systems have used the capture and replay of recordings to overcome availability constraints between team members [13, 31, 46], our work contributes a new technique. Instead of Collaborators consuming pre-recorded materials and otherwise interacting with the missing colleagues asynchronously, they are able to synchronously converse with an agent that dynamically generates and relays content based on its knowledge of its Source. This work is also distinct from prior literature, in that it explored the concept of using agents as proxies that can **advocate for specific people** (i.e., oneself or a work colleague), rather than agents with generic personalities (e.g., Alexa, Siri, Otter.ai) or agents that resemble public figures (e.g., Meta's celebrity-based AI chatbot characters [28]).

   As the use of AI in Computer Mediated Communication (CMC) rapidly expands, it can be useful to other researchers and practitioners in this domain to situate our work with respect to a common framework, such as the AI-Mediated Communication **(AI-MC) framework** outlined by Hancock et al. [16]. Dittos have a sender *role orientation*, with the *optimization goal* to reliably represent the Sources's perspective in meetings where the Source is unable to attend. Dittos are multimodal and use audio and video *media types*, with a very high *magnitude* of AI-involvement, as well high levels of AI-*autonomy*. While outlining the design space, Hancock et al. raised several questions pertaining to AI-MC, touching upon issues such as self-presentation, impression formation and trust, transparency, and misrepresentation. Our work contributes **empirical insight** into the impact these agents can have in meetings, and highlights that Dittos in particular show promise in evoking feelings of **presence** and **trust**, as well as informing **decision making**. While it has previously been known that agents that emit social cues and have anthropomorphic features foster feelings of trust [39, 40, 43], our findings highlight that closer resemblance to a particular colleague can amplify this phenomenon in team meetings. Additionally, our work reveals in what **scenarios** people would like to use these agents, and how others react and respond to an AI-agent that fulfills a mediary role in communication between working team members in meeting contexts. In both the focus group and user study sessions, we discovered that people can appreciate the potential utility of Dittos to cover for them in meetings they would otherwise be unable to attend themselves. Nonetheless, this work is a very **early stage exploration** into a novel concept. As such, people's attitudes as captured in this work may evolve to be different in the future, as people gain more exposure and build more familiarity with this technology over time.

7.2   Ditto Interaction Design Requirements

Our explorations unveiled several pertinent design considerations that must be carefully addressed before Dittos would be largely adopted. For example, while the AI-MC framework [16] states that disclosure of the use of AI may not always be necessary, our empirical investigation found that in

the case of using Dittos in meetings, **transparency** is critical. In both the focus groups and the WOz study, many participants believed that it should be visually clear that Dittos are AI-agents rather than teleoperated avatars, with participants suggesting a cartoon-like appearance rather than a highly realistic depiction (e.g., deepfake). Based on the dimensions outlined by Garau [14], Mansour et al. [25], Collaborators value *truthfulness* over high levels of *realism* for the use of *anthropomorphic* agents in work meetings. This result also aligns with earlier findings [19, 38] that full realism is undesirable with regards to avatar-like representations for work.

Another critical design requirement is that the Ditto should not disrupt the **pace of a conversation**. People in our study were very sensitive to **delayed responses** from agents. Slowness to respond tended to halt conversation and erode Collaborators' confidence and trust in the agents. Collaborators were more likely to dismiss what the agent had to say, if the point it made connected to a conversation topic that the group had already gotten past. In some cases, it could also exacerbate perceptions of the agents being stubborn or unyielding, if the delayed comments do not properly acknowledge a prior comment. Participants also preferred the first-person language of the Ditto over the **third-person language** of the Delegate, primarily because the latter added unnecessary bulk to the conversation and slowed things down. Participants' sensitivity to delays corroborates previous research in human robot interaction (HRI) and human agent interaction (HAI), where latency has been found to negatively impact users' perception of the experience and their appraisal of the robot or agent [3, 32, 45].

Although the WOz investigation did not span interactions preceding or following meetings, participants' commentary brought to light the importance of **follow-up and escalation mechanisms** as part of a full end-to-end infrastructure. In cases where the Ditto is unsure or unable to comment on a topic, it could suggest that it will escalate the issue to the Source, and that the Source will get back to the Collaborators, which we did not explore in our study. When Collaborators have doubts about what a Ditto is saying, or are uncertain about whether the meeting outcomes will be relayed to the Source, a method to request a follow-up with the Source is desirable. In other words, such mechanisms should operate **bidirectionally** and will be vital for ensuring accurate and reliable communication. Without sufficient means to ensure that the agents can be trusted, people would be inclined to confine the use of Dittos to **highly passive or scripted use cases**, such as attending all-hands meetings or delivering pre-prepared presentations. Hesitation around agent autonomy, and a default preference for more predefined behaviors parallels findings from a survey conducted by Schiaffino and Amandi [41] on the use of interface agents in mixed-initiative interaction contexts. However, our findings suggest that user preferences may shift; through an iterative process of trial and confirmation, both Sources and Collaborators may come to trust Dittos over time and allow them more agency. We are curious to see if, over time, people find value in Ditto participation in meetings compared to the alternatives (e.g., delaying the meeting until everyone is available, not having a meeting at all and handling it asynchronously instead).

7.3 Design Considerations to Manage Reputational Risk and Accountability

Our results also revealed that a core factor that will affect the social acceptance of Dittos is its impact on the Source's **reputation** and **accountability**. In both the focus groups and WOz, participants' generally believed that the act of sending a Ditto to a meeting rather than attending oneself could be considered rude and could damage their reputation if the circumstances do not necessitate its use (e.g., illness, timezone differences). Participants also worried that inappropriate agent behavior during a meeting could have negative repercussions on their professional image. This concern echoed throughout our focus groups and the user study, and was also a prominent speculated fear in the investigation by Lee et al. [22]. However, ensuring that the Source's reputation stays intact is a multifaceted challenge, that overlaps with a number of **key design tensions**.

One core design tension revolves around the Ditto's degree of **speech autonomy**. Sources expressed a strong desire for agents to stick to a script, as it would help ensure that it does not say anything inappropriate or inaccurate. However, this is at odds with Collaborators' desired user experience. In the WOz study, Collaborators did not like it when the agent's opinions did not seem to adapt to the group's conversation. In fact, Collaborators went so far as to reason that the agents seemed frustratingly unable to adapt beyond their programmed script. This was a particularly curious reaction, given that the agents' statements were generated by the human Sources themselves. It is possible that the Collaborators' reactions may have been strongly driven by their projections of having been told that they were interacting with an AI agent, along with experiencing some communication latency in the conversation, even though they were actually interacting with a live human. This reaction merits more research in that current user reactions to interacting with agents may be strongly shaped by what they are expecting as much as how the agent actually behaves.

Another design tension revolved around **visual and vocal resemblance**. Despite some notable pushback during the focus group sessions against using agents that visually resembled the Source, in our user study, Collaborators preferred Dittos over Delegates. Visual and vocal resemblance also amplified the sense of **presence** of the Source in the meeting and was a strong cue to Collaborators to think of them. This finding corroborates prior research on synthesized voices that those resembling one's own voice [21] or the voices of friends of family [9] can boost feelings of presence. It also had the effect of inspiring higher levels of Collaborator **trust** in the agent. While the majority of Sources also expressed preference for the Ditto representation, it is worth noting that a fraction preferred the Delegate representation instead. If the agent assumes another identity than the Source's, it can help distance the agent's statements and behavior from the Source. This lends itself towards "plausible deniability" for Sources, mitigating reputation and accountability risks.

As part of reputational management, some Sources expressed a desire for **personality controls** that could enable them to adapt the agent's demeanor to suit their goals for the meeting at hand. For example, Sources may want the agent to be more charismatic or assertive in some meetings to be more effective in negotiations. At the same time, a participant worried that it could lead to her Collaborators preferring her Ditto over herself, mirroring concerns of "a better version of me" raised in the investigation by Lee et al. [22] and "prioritizing the machine teammate" as highlighted by Seeber et al. [43] in a panel discussion on the use of technology-based autonomous agents. **Follow-up and escalation mechanisms** may also be useful for managing reputation and accountability, by allowing Sources to monitor and ensure that the agent behaves appropriately and performs accurately.

7.4 Navigating Existing AI Biases

Our findings suggested that the overall perception of Dittos is influenced by people's **pre-conceived notions and expectations around AI and agents**. For example, participants tended not to view **AI agents as equal conversation partners to humans**, which shaped their thoughts around **social etiquette and protocols**. Many participants across both the focus groups and WOz expressed that it would be insulting to receive a Ditto at the meeting instead of the Source, unless the Source has clearly cited a sanctioned reason ahead of time, such as a timezone conflict or illness. Our user study also revealed that trust is impacted by people's awareness of AI-shortcomings, such as **hallucination**. In this study, Collaborators were susceptible to discounting an agent's statements if they did not align with their expectations, even though they were actually coming from the Source themselves. In one instance, when a Source spoke of a Marie Antoinette themed event, Collaborators presumed this was a case of the AI-hallucinating rather than being representative of the Source's true thoughts and opinions, unexpected as they were. Furthermore, in tasking

participant groups in the WOz study to engage in divergent and convergent thinking together, we were able to gather insight into how people may treat AI agents in a decision-making scenario. For example, we learned that **people are reluctant to argue with AI, under a belief that it is futile to change its opinions**. While overreliance [6, 37] is a well known issue in AI-research and literature, our work cautions that people may also be easily swayed by AI in group conversation settings due to an unwillingness to try and convince an AI of another opinion. This is a large risk worthy of further consideration and attention in future research on using AI in group conversations.

7.5   Limitations

Although we tried to recruit participants with a broad array of backgrounds, our participants came from within a single large technology company. As such, our sample likely captured an "early-adopter" population who had some positive bias towards using novel technologies. Additionally, the number of participants recruited was relatively small. In future studies, it would be helpful to recruit a larger sample of participants, from beyond this work's recruited sample, to uncover a broader range of insights and gain more statistical power in quantitative analyses.

As noted in Section 5.1, we followed a WOz study methodology to get early reactions to the concept to shape their development. More research is needed to observe people interacting with actual autonomous Dittos (and other agents) once they have been implemented.

Given the uniqueness of the Ditto and Delegate concepts, it is probable that there was a strong novelty effect in the WOz user study, particularly since the experiment spanned a short period of time. Additionally, the study scenario was based on a single meeting type among close-knit colleagues, and was limited to involving only one agent rather than multiple. It also did not cover practices before and after meetings, nor did it span multiple meetings sessions. However, the ability to seamlessly share and receive information with one's meeting agent in a kind of feedback loop will be a critical part of the overall system infrastructure. Hence, in future work, it would be fruitful to extend this research to cover pre- and post-meeting interactions and protocols, different types of meetings, different sets of collaborators, including other agents, and varied meeting circumstances. Conducting a longitudinal study, or a case study of deployments in real meetings would also be beneficial. Beyond uncovering more about the practices people may develop around the use of this technology, this may help reveal its potential downstream effects on workplace meeting cultures, team productivity levels, and workers' wellbeing. For instance, there is some evidence to suggest that reducing the frequency in which people need to attend meetings can lessen their perceived fatigue and workload [24].

Our study demonstrates some of the potential for Dittos to help shape the future of work. Having confirmed the opportunity, it would be helpful to more extensively investigate associated design and technical implementation challenges. Besides finding ways to navigate the design tensions noted in Section 7.3, this includes minimizing response latency, mitigating hallucinations, and creating pipelines to collect appropriate and reliable user information. Building effective safeguards to protect against sensitive user information from being leaked is also of great importance. Additionally, more exploration would be useful to uncover the ethical implications of how Dittos could introduce problematic interactions, or even be exploited to introduce social harm in the future of collaborative work. Furthermore, it is important to recognize that the AI landscape is rapidly evolving, triggering diverse and changing perspectives. The success or failure of Dittos in the future will be impacted by users' perspectives of agents, as well as the underlying AI technology, as this sector continues to evolve and becomes more ubiquitous.

8 Conclusion

In this work, we proposed and explored the novel concept of a Ditto—a personalized, embodied, and mimetic agent that would encapsulate a person's knowledge, appearance, and behavior and serve as their proxy in a meeting that they would otherwise be unable to attend. Our initial focus group sessions revealed areas of opportunity and concern, and highlighted divergent opinions about whether such agents should resemble Sources as Dittos, or represent them as distinct Delegate entities.

In a follow-up Wizard of Oz user study, we simulated the experience of meeting with a Ditto and Delegate for multiple working teams to gain more authentic insights. We learned that Dittos may invoke stronger feelings of presence and trust, but that it is critical to develop methods to manage concerns around reputational risk and accountability. This investigation also highlighted that there are design tensions that arise from differing needs and perspectives between Sources and Collaborators as well as pre-existing notions about interacting with AI agents which work to shape the experience beyond its actual design. Looking ahead, this work suggests that personalized, embodied and mimetic agents may play a helpful role in the future of collaborative work, but that several design and social aspects need to be addressed to maximize the utility of Dittos as proxies in collaborative online meetings.

Acknowledgments

References

[1] Jeremy N Bailenson. 2012. Doppelgangers-a new form of self? *Psychologist* 25, 1 (2012), 36–38.
[2] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*. 1–9.
[3] Halim-Antoine Boukaram, Micheline Ziadee, and Majd F Sakr. 2021. Mitigating the Effects of Delayed Virtual Agent Response Time Using Conversational Fillers. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. 130–138.
[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[5] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
[6] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.
[7] Justine Cassell. 2000. Embodied conversational interface agents. *Commun. ACM* 43, 4 (2000), 70–78.
[8] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001), 67–67.
[9] Sam WT Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. 2021. Kinvoices: using voices of friends and family in voice interfaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
[10] Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afil Farhan, Birthe Nesset, Meriam Moujahid, Tanvi Dinkar, et al. 2023. FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions. *arXiv preprint arXiv:2308.15214* (2023).
[11] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[12] Evan DeFilippis, Stephen Michael Impink, Madison Singell, Jeffrey T Polzer, and Raffaella Sadun. 2020. *Collaborating during coronavirus: The impact of COVID-19 on the nature of work*. Technical Report. National Bureau of Economic Research.

[13] Andreas Rene Fender and Christian Holz. 2022. Causality-preserving asynchronous reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[14] Maia Garau. 2003. *The impact of avatar fidelity on social interaction in virtual environments*. University of London, University College London (United Kingdom).

[15] Victoria Groom, Clifford Nass, Tina Chen, Alexia Nielsen, James K Scarborough, and Erica Robles. 2009. Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies* 67, 10 (2009), 842–849.

[16] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication* 25, 1 (2020), 89–100.

[17] Yuji Hatada, Shigeo Yoshida, Takuji Narumi, and Michitaka Hirose. 2019. Double shellf: What psychological effects can be caused through interaction with a doppelganger?. In *Proceedings of the 10th Augmented Human International Conference 2019*. 1–8.

[18] Manuel Flurin Hendry, Norbert Kottmann, Martin Fröhlich, Florian Bruggisser, Marco Quandt, Stella Speziali, Valentin Huber, and Chris Salter. 2023. Are you talking to me? a case study in emotional human-machine interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 19. 417–424.

[19] Kori M Inkpen and Mara Sedlins. 2011. Me and my avatar: exploring users' comfort with avatars for workplace communication. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 383–386.

[20] Robert Johansen. 1988. *Groupware: Computer support for business teams*. The Free Press.

[21] Kwan-Min Lee and Clifford Nass. 2005. Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology* 7, 1 (2005), 31–45.

[22] Patrick Yung Kang Lee, Ning F. Ma, Ig-Jae Kim, and Dongwook Yoon. 2023. Speculating on Risks of AI Clones to Selfhood and Relationships: Doppelganger-phobia, Identity Fragmentation, and Living Memories. *Proceedings of the ACM Human-Computer Interaction* 7, CSCW1 (2023), 1–28. https://doi.org/10.1145/3579524

[23] Gale Lucas, Evan Szablowski, Jonathan Gratch, Andrew Feng, Tiffany Huang, Jill Boberg, and Ari Shapiro. 2016. The effect of operating a virtual doppelganger in a 3D simulation. In *Proceedings of the 9th International Conference on Motion in Games*. 167–174.

[24] Alexandra Luong and Steven G Rogelberg. 2005. Meetings and more meetings: The relationship between meeting load and the daily well-being of employees. *Group Dynamics: Theory, Research, and Practice* 9, 1 (2005), 58.

[25] Samah Mansour, Mostafa El-Said, Carolyn Rude-Parkins, and Jagadeesh Nandigam. 2006. The interactive effect of avatar visual fidelity and behavioral fidelity in the collaborative virtual reality environment on the perception of social interaction. *WSEAS Transactions on Communications* 5, 8 (2006), 1501–1509.

[26] Moira McGregor and John C Tang. 2017. More to meetings: challenges in using speech-based technology to support meetings. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 2208–2220.

[27] Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, Solon Barocas, and Ashton Anderson. 2022. Mimetic Models: Ethical Implications of AI that Acts Like You. In *5th AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. 479–490. https://doi.org/10.1145/3514094.3534177

[28] Meta 2023. *Introducing New AI Experiences Across Our Family of Apps and Devices*. Meta. https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/

[29] Katashi Nagao and Akikazu Takeuchi. 1994. Social Interaction: Multimodal Conversation with Social Agents. *Proceedings of the AAAI-94* (1994), 22–28.

[30] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[31] Mamoun Nawahdah and Tomoo Inoue. 2013. Virtually dining together in time-shifted environment: KIZUNA design. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 779–788.

[32] Andreea Niculescu, Betsy Van Dijk, Anton Nijholt, Dilip Kumar Limbu, Swee Lan See, and Alvin Hong Yee Wong. 2010. Socializing with Olivia, the youngest robot receptionist outside the lab. In *Social Robotics: Second International Conference on Social Robotics, ICSR 2010, Singapore, November 23-24, 2010. Proceedings 2*. Springer, 50–62.

[33] Nielsen Norman Group 2022. *Interviewing Users*. Nielsen Norman Group. https://www.nngroup.com/articles/interviewing-users/

[34] OpenAI 2022. *Introducing ChatGPT*. OpenAI. https://openai.com/index/chatgpt/

[35] OpenAI 2023. *Introducing GPTs*. OpenAI. https://openai.com/index/introducing-gpts/

[36] Ye Pan and Anthony Steed. 2016. A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. *Frontiers in Robotics and AI* 3 (2016), 12.

[37] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI Literature Review. *Microsoft Research* (2022).

[38] Vrushank Phadnis, Kristin Moore, and Mar Gonzalez Franco. 2023. Avatars in Work Meetings: Correlation Between Photorealism and Appeal. *arXiv preprint arXiv:2304.01405* (2023).

[39] Lingyun Qiu and Izak Benbasat. 2005. An investigation into the effects of text-to-speech voice and 3D avatars on the perception of presence and flow of live help in electronic commerce. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 4 (2005), 329–355.

[40] Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of management information systems* 25, 4 (2009), 145–182.

[41] Silvia Schiaffino and Analıa Amandi. 2004. User–interface agent interaction: personalization issues. *International Journal of Human-Computer Studies* 60, 1 (2004), 129–148.

[42] Ben Schniederman and Michael Muller. 2023. On AI Anthropomorphism. https://medium.com/human-centered-ai/on-ai-anthropomorphism-abff4cecc5ae

[43] Isabella Seeber, Lena Waizenegger, Stefan Seidel, Stefan Morana, Izak Benbasat, and Paul Benjamin Lowry. 2020. Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research* 30, 1 (2020), 1–18.

[44] N Sadat Shami, Li-Te Cheng, Steven Rohall, Andrew Sempere, and John Patterson. 2010. Avatars meet meetings: Design issues in integrating avatars in distributed corporate meetings. In *Proceedings of the 2010 ACM International Conference on Supporting Group Work*. 35–44.

[45] Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2008. How quickly should communication robots respond?. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. 153–160.

[46] John Tang, Jennifer Marlow, Aaron Hoff, Asta Roseway, Kori Inkpen, Chen Zhao, and Xiang Cao. 2012. Time travel proxy: using lightweight video recordings to create asynchronous, interactive meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3111–3120.

[47] John C Tang, Chen Zhao, Xiang Cao, and Kori Inkpen. 2011. Your time zone or mine? A study of globally time zone-shifted collaboration. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 235–244.

[48] Marty J Wolf, K Miller, and Frances S Grodzinsky. 2017. Why we should have seen that coming: comments on Microsoft's tay" experiment," and wider implications. *Acm Sigcas Computers and Society* 47, 3 (2017), 54–64.

[49] Rui Zhang, Wen Duan, Christopher Flathmann, Nathan McNeese, Guo Freeman, and Alyssa Williams. 2023. Investigating AI Teammate Communication Strategies and Their Impact in Human-AI Teams for Effective Teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

A Appendix - Focus Group Materials

A.1 Sending A Ditto: Imagine you could SEND your Ditto on your behalf to a meeting...

- What kind of meetings would you send your Ditto to?
- What kind of meetings would you NOT send your Ditto to?
- What kinds of factors might impact whether you choose to send a Ditto or not?
- What kinds of things would you want it to be able to say/do?
- What kinds of things would you NOT want it to be able to say/do?
- What pros do you see with being able to send a Ditto when you have to miss a meeting?
- What cons do you see with being able to send a Ditto when you have to miss a meeting?

A.2 Speaking with Someone Else's Ditto: Imagine you could talk to someone else's Ditto...

- What would you want to be able to ask or tell someone else's Ditto?
- When would you prefer to speak to the real person and not their Ditto? Or voice versa
- What kind of pros can you imagine, regarding speaking to someone else's Ditto?
- What kinds of cons can you image, regarding speaking to someone else's Ditto?

A.3 General Thoughts and Concerns

- What are some general thoughts or concerns you have about Dittos?

- Any comments about what it should look and sound like?

B   Appendix - Wizard of Oz Study Materials

B.1   Collaborator - Survey Questions

B.1.1   Demographic Questions.

- What is your Participant ID#?
- Please select your age range. [18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75 and above, I prefer not to answer]
- How would you describe your gender? [Male, Female, Non-binary, Prefer not to say]
- What is your current job title (occupation)?

Pause #1! Please wait here until the experimenter tells you to proceed.

B.1.2   Condition Questions. * Note – The term Ditto was replaced with Delegate for the Delegate condition.

Please indicate how you feel about the statements below. [Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree]

- I was satisfied with the overall meeting experience.
- I trusted what the agent had to say.
- I felt that the views/opinions of my colleague represented by the agent were considered in the final group decision.

Given your experience of your meeting with the Ditto Agent, please indicate how you feel about the statements below. [Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree]

- It was beneficial to have the Ditto in the meeting, instead of having my colleague skip the meeting altogether.
- It would have been better if the Ditto had not been there.
- Having my colleague's Ditto in the meeting helped me think of them.
- I had the feeling that the real person (my colleague) was attending the meeting when their Ditto was there.

Given your experience of your meeting with the Ditto Agent, please indicate how you feel about the statements below. [Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree]

- The Ditto paid close attention to us.
- The Ditto was easily distracted from us when other things were going on.
- The owner of the Ditto's thoughts were clear to me.

Any additional thoughts or comments about this experience?
Pause # 2! Please wait here until the experimenter tells you to proceed.

B.1.3   Exit Questions.

- The next time your colleague cannot attend a meeting, what would you prefer?
- Have my colleague's Ditto agent attend the meeting on their behalf.
- Have my colleague's Delegate agent attend the meeting on their behalf.
- Have neither agent attend the meeting on my colleague's behalf.

Why do you have this preference?
Thanks! You have reached the end of this survey. Please submit this form. (Please let the experimenter know, and they will tell you the next steps!)

B.2 Collaborator - Semi-Structured Interview Questions

(1) What did you think about the two conditions? What would you like to improve?
(2) To what extent did each agent (Ditto/Delegate) evoke thoughts for you about their owner?
(3) To what extent did what each agent (Ditto/Delegate) said match what you felt the real person would say?
(4) To what extent did the agent (Ditto/Delegate) make you consider the opinions, views, or values of the real person?
(5) How did they feel about the turn-taking with the agent (Ditto/Delegate)?
(6) Do you feel like the Ditto should always be visually discernible?
(7) Were there any points in time when it seemed that the agent was/wasn't a real person?
(8) To what extent did the agent (Ditto/Delegate) influence what you said, as if the owner was there?
(9) How much thought did you give to the real person's opinion across the conditions? Or how much cognitive effort did you invest into thinking about the real person's opinion?
(10) Were there any points in time when the agent (Ditto/Delegate) made you think it was the real person?

B.3 Source - Survey Questions

B.3.1 Demographic Questions.

- What is your Participant ID#?
- Please select your age range. [18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75 and above, I prefer not to answer]
- How would you describe your gender? [Male, Female, Non-binary, Prefer not to say]
- What is your current job title (occupation)?

Pause #1! Please wait here until the experimenter tells you to proceed.

B.3.2 Condition Questions. * Note – The term Ditto was replaced with Delegate for the Delegate condition.

Please indicate how you feel about the statements below. [Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree]

- I was satisfied with the overall meeting experience.
- I felt my colleagues trusted what I had to say.
- I felt that my views/opinions were considered in the final group decision.

Given your experience embodying the Ditto Agent, please indicate how you feel about the statements below. [Strongly Disagree, Disagree, Slightly Disagree, Neutral, Slightly Agree, Agree, Strongly Agree]

- My colleagues paid close attention to me.

Any additional thoughts or comments about this experience?
Pause #2: Thanks! Please wait until the experimenter asks you to proceed.

B.3.3 Exit Questions. The next time you cannot attend a team meeting, what would you prefer to do?

- Send my Ditto agent to attend the meeting on my behalf.
- Send my Delegate agent to attend the meeting on my behalf.
- Do not send either agent to attend the meeting on my behalf.

Why do you have that preference (selected above)?

Thanks! You have reached the end of this survey. Please submit this form. (Please let the experimenter know, and they will tell you the next steps!)

B.4 Source – Semi-Structured Interview Questions

(1) Did you feel there were differences in the way your colleagues treated your Ditto and Delegate agents?
(2) Did you feel there were differences between the Ditto and Delegate conditions in your ability to express your thoughts, opinions and preferences?
(3) Did you feel there were differences in the way your colleagues reacted to what your Ditto and Delegate agents have to say?
(4) Based on your experience observing how your colleagues interacted with your Ditto and Delegate agents, do you have a preference for how you would like to be represented by an agent in a meeting?
(5) Are there any improvements that could help the agent more effectively represent you in a meeting?
(6) What situations would you feel comfortable sending an autonomous agent to represent you in a meeting?