Secondary Mental Models: Introducing Conversational Agents in Financial Advisory Service Encounters

Eckhardt, Sven ; Bucher, Andreas ; Kalunder, Madlaina ; Dolata, Mateusz ; Agotai, Doris ; Schwabe, Gerhard

Abstract: When introducing unfamiliar Artificial Intelligence (AI)-based systems, such as conversational agents (CAs), one needs to ensure that users interact with them according to their design. While past research has studied single-user environments, many practical settings involve multiple parties. This study addresses this gap and focuses on financial advisory service encounters and how mental models evolve in multi-party contexts. A multimodal interactive CA is developed and tested in financial consultations with 24 clients. The observations of these consultations and subsequent interviews provide insights into the challenges of using CAs in unfamiliar contexts. The clients have difficulties effectively using the system. This is linked to the institutional setting of financial advisory service encounters and a mismatch between the designer's conceptual model and the client's mental model, which we call secondary mental model.

Originally published at:
Eckhardt, Sven; Bucher, Andreas; Kalunder, Madlaina; Dolata, Mateusz; Agotai, Doris; Schwabe, Gerhard (2023). Secondary Mental Models: Introducing Conversational Agents in Financial Advisory Service Encounters. In: Forty-Fourth International Conference on Information Systems, Hyderabad, India, 9 December 2023 - 13 December 2023. Association for Information Systems, online.

Secondary Mental Models: Introducing Conversational Agents in Financial Advisory Service Encounters

Completed Research Paper

Sven Eckhardt
University of Zurich
Zurich, Switzerland
eckhardt@ifi.uzh.ch

Andreas Bucher
University of Zurich
Zurich, Switzerland
bucher@ifi.uzh.ch

Madlaina Kalunder
FHNW University of Applied Sciences
Northwestern Switzerland
Windisch, Switzerland
madlaina.kalunder@fhnw.ch

Mateusz Dolata
University of Zurich
Zurich, Switzerland
dolata@ifi.uzh.ch

Doris Agotai
FHNW University of Applied Sciences
Northwestern Switzerland
Windisch, Switzerland
doris.agotai@fhnw.ch

Gerhard Schwabe
University of Zurich
Zurich, Switzerland
schwabe@ifi.uzh.ch

Abstract

When introducing unfamiliar Artificial Intelligence (AI)-based systems, such as conversational agents (CAs), one needs to ensure that users interact with them according to their design. While past research has studied single-user environments, many practical settings involve multiple parties. This study addresses this gap and focuses on financial advisory service encounters and how mental models evolve in multi-party contexts. A multimodal interactive CA is developed and tested in financial consultations with 24 clients. The observations of these consultations and subsequent interviews provide insights into the challenges of using CAs in unfamiliar contexts. The clients have difficulties effectively using the system. This is linked to the institutional setting of financial advisory service encounters and a mismatch between the designer's conceptual model and the client's mental model, which we call secondary mental model.

Keywords: Mental Models, Conversational Agent, Artificial Intelligence, Financial Advisory Service Encounters

Introduction

While being fiction a couple of decades ago, AI-based systems have found their way into our private and professional lives. Examples of such systems range from autonomous vehicles (Faisal et al., 2019) to speech-based CAs (Maedche et al., 2019) that control household appliances (Laumer et al., 2019) or serve as medical advisors (Tudor Car et al., 2020). Given recent technological advances, the digital and physical world boundaries are becoming increasingly blurry for AI-based systems, presenting new and multifold possibilities for interacting with them. For instance, recent developments of CAs, like Amazon's Echo Show, explore new input and output modalities by combining speech recognition and generation, motion detection, and touch interaction (Soubutts et al., 2022). Such multimodal approaches, i.e., approaches with multiple means of interaction, such as text, voice, visual, and haptics, open new and rich possibilities to use and interact with these systems. At the same time, they also increase the complexity of systems. Therefore, ensuring that these AI-based systems are used as intended is becoming increasingly challenging for their designers and developers, who are confronted with many design possibilities and considerations.

Generally, the question of how people interact, and appropriate systems compared to the intended use has preoccupied researchers and practitioners for a long time. One of the concepts frequently used in this context is the concept of mental models (Gentner & Stevens, 1983; Johnson-Laird, 1983). Mental models are internal representations of a system within the mind of its users and thus represent their beliefs and thinking about what a system is and how it behaves (Rouse & Morris, 1986). Designers create a system according to the designers' *conceptual model,* and users interact with the system according to the users' *mental models* (Norman, 2013). While the unintended use of information systems can create value, much research aims to align these two models and ensure a system's intended use, especially in human-computer interaction (HCI). A pointed example of the importance of intended use is plane cockpit controls: If a pilot fails to control the aircraft as intended, many lives could be at risk.

As the capabilities of AI-based systems evolve, they are increasingly adopted in multi-party contexts. Examples of that are customer support (Xu et al., 2017), car sales on online platforms (Eckhardt et al., 2022), or financial advisory service encounters (Bucher et al., 2022). Recent literature has investigated the introduction of CAs in financial advisory service encounters (Dolata, Kilic, et al., 2019; Schmid et al., 2022). With the introduction of such CAs, the setting of financial advisory service encounters becomes a triadic relationship between advisor, client, and CA. Prior research of CAs and AI-based systems has primarily focused on the dyadic relationship of human-AI interaction, i.e., one user with one AI-based system. However, many practical settings involve multiple parties with conflicting interests or different levels of knowledge, such as advisory service encounters or market negotiations. It remains an open question how mental models on AI-based systems, such as CAs, are adapted and formed in multi-party settings.

Financial advisory service encounters are a particular form of multi-party interaction. They happen in an institutional setting and are not regularly done by the clients (Schmid et al., 2022). Therefore, as opposed to everyday interactions, they are relatively unfamiliar. The reasons for integrating an AI-based system into financial advisory service encounters are manifold: a loss of trust after the financial crisis of 2008, ever-changing regulations, or higher client expectations have increased the cognitive load on the financial advisor and contributed to an increasingly complex setting. Tools to support financial advisory service encounters have become multimodal to adapt to the advisor's and client's needs (Dolata, Agotai, et al., 2019, 2020). Some studies have already begun investigating the potential of introducing intelligent systems like CAs into financial advisory service encounters (Dolata, Kilic, et al., 2019). Similarly, we believe that a multimodal CA that can keep track of the client's personal details, recognize portfolio adaptations, or explain financial terminology can support the financial advisory service and provide a new immersive experience to the clients. However, this is only if the tool is used to its full potential, which entails a fitting mental model for all participants. Previous studies have primarily been conceptual (Bucher et al., 2022; Schmid et al., 2022), but how advisors and clients interact with such systems remains unexplored.

Our study addresses this gap and explores the formation and influence of mental models of AI-based systems within multi-party settings, particularly in financial advisory service encounters. Herein, we focus on clients during financial consultations. We explore the use of CAs in financial consultations and the formation of mental models for both parties, the client and the advisor. This yields insights into how mental models are created and adapted and how people interact with unfamiliar systems. These findings provide answers to the following research question:

RQ1 *How do clients interact with a multimodal, interactive CA designed to facilitate co-located financial consultations?*

RQ2 *What mental models of the CA do the clients establish during that interaction?*

This study is part of a larger research project involving two universities, two regional banks, and two technology partners in Switzerland. This research project aims to develop a multimodal, interactive CA to support financial advisory service encounters. The prototype of the CA was then tested and evaluated with 12 advisors and 24 clients in financial consultations at the banks' offices. In this study, we analyzed the video recordings of these consultations and subsequent interviews to understand the clients' interactions with the CA. Our findings indicate that interacting with an unfamiliar and multimodal CA leads to several issues. We find that clients have an inhibition to interact with the system, leading the advisor to be their primary interaction source. For one, this is due to the institutional setting of financial advisory service encounters. For the other, this is because clients are only *secondary users* and, thus, cannot adapt their mental models successfully. To conceptualize this, we introduce and discuss *secondary mental models* for specific settings such as financial advisory service encounters and present implications of this phenomenon.

Background and Related Work

IT in Financial Advisory Service Encounters

Financial advisory service encounters are a prime example of the principal-agent problem (Eisenhardt, 1989; Golec, 1992; Schwabe & Nussbaumer, 2009) and happen in an institutional setting (Dolata, Agotai, et al., 2019; Dolata, Kilic, et al., 2019; Schmid et al., 2022). Most often, this happens at a bank or similar office buildings frequented by the advisor. Additionally, financial advisory service encounters have much information asymmetry (Auh et al., 2007; Nussbaumer et al., 2012) and unclear incentives (Jungermann, 1999). Financial advisors exploited the information asymmetry to generate profits until the financial crisis of 2008. Since then, stricter regulatory requirements have ensured that information asymmetries are reduced by educating clients and enabling them to make informed decisions.

Besides regulations, a current stream to mitigate the principal-agent problem and reduce information asymmetry is the introduction of various IT tools. IT tools can engage in information exchange (Kilic et al., 2017) and improve knowledge transfer (Heinrich et al., 2014). Table-top computers can increase perceived transparency (Nussbaumer et al., 2012). Large interactive screens enabled more joyful collaboration (Novak & Schmidt, 2009). Furthermore, interactive systems empower advisors (Boulus-Rødje, 2018; Giesbrecht et al., 2014), reduce cognitive overload (Giesbrecht et al., 2015), enhance the persuasiveness of the advisory service (Dolata & Schwabe, 2018), and increase transparency (Comes & Schwabe, 2016). Additionally, research suggests that IT can enhance the advisory service when it integrates well with the social rituals of an advisory service (Dolata, Steigler, et al., 2019; Heyman & Artman, 2015), but it becomes a problem if it goes against the rituals (Dolata, Schenk, et al., 2020; Nueesch et al., 2016). In some cases, IT has been perceived as dominating interactions and negatively impacting relationship-building (Heinrich, Kilic, Aschoff, et al., 2014) or disturbing the natural flow of conversations (Mørck et al., 2018). Overall, IT tools can support financial advisory service encounters when done cautiously.

Recently, multimodal interactive systems have been shown to enhance overall bank client satisfaction (Dolata, Agotai, et al., 2019, 2020). The next logical step to reduce the cognitive load on the advisor and increase the service quality is the introduction of a CA (Dolata, Kilic, et al., 2019). Some studies look into the design of CAs in financial advisory service encounters (Bucher et al., 2022; Eckhardt et al., 2023; Schmid et al., 2022). Further, it is widely acknowledged that managing AI-based systems such as CAs differs from managing IT tools in the past (Berente et al., 2021). Therefore, more research must be conducted on user behavior in financial advisory service encounters when exposed to CAs.

Multimodal Conversational Agents

CAs are a broadly researched topic. In general, CAs can be divided into two primary modes of communication: text-based and speech-based (Gnewuch et al., 2017). Text-based CAs are chatbots such as ELIZA (Weizenbaum, 1966) or customer support chatbots (Xu et al., 2017). Speech-based chatbots are often called digital assistants (Maedche et al., 2019) or digital agents (Chatterjee et al., 2019), such as Apple's Siri or Amazon's Alexa. We use the general term CA for the multimodal speech-based agent presented in this study. For the integration of AI-based systems like CAs in service encounters, such as financial advisory service encounters, several studies developed archetypes (De Keyser et al., 2019; Ostrom et al., 2019; Poser et al., 2022): AI-based systems can either substitute the service employee or augments them, e.g., by *assisting* or *facilitating* the setting. These archetypes are presented in Figure 1.

Furthermore, over the last years, the rapid development of technology has led to the widespread use of CAs in private households and specific industrial sectors (Hirschberg & Manning, 2015; Nadkarni et al., 2011). The interaction between the user and such systems is perceived as social despite simple communication patterns (Bickmore & Cassell, 2005; Nass et al., 1994). However, designing CAs for a positive user experience is challenging (Diederich et al., 2022). One approach for a positive user experience is multimodal CAs (Kopp et al., 2005; Provoost et al., 2017; Weiss et al., 2015). Multimodality can be defined as „supporting communication with the user through different modalities“ (Nigay & Coutaz, 1993). Examples of modalities include voice, typing, or visual. In summary, modalities can be used *sequentially* or *parallel,* and the adjacent data can be *combined* or *independent* (Nigay & Coutaz, 1993). A *parallel* use of modalities allows the user to use multiple modalities simultaneously, whereas a *sequential* use of modalities forces the user to use one modality after another. *Combined* means that different types of data (from different modalities) can be merged, and *independent* means that all data is considered separate. Besides CAs, multimodality has been researched in many applications (Turk, 2014). Multimodality has increased multitasking performance (Kim & Kim, 2011). Further, guidelines for multimodal user interface design were proposed, such as "*designing for the broadest range of users and contexts of use"* or integrating *"modalities in a manner compatible with user preferences, context, and system functionality"* (Reeves et al., 2004).

Overall, research recognizes the potential of CAs and multimodality. However, generally, users have a bad mental model of these CAs (Luger & Sellen, 2016). This only increases the challenge of creating a positive user experience. While research on CAs is emerging and ongoing (Seeber et al., 2020), current research rarely focuses on mental models of CAs in multi-party settings.

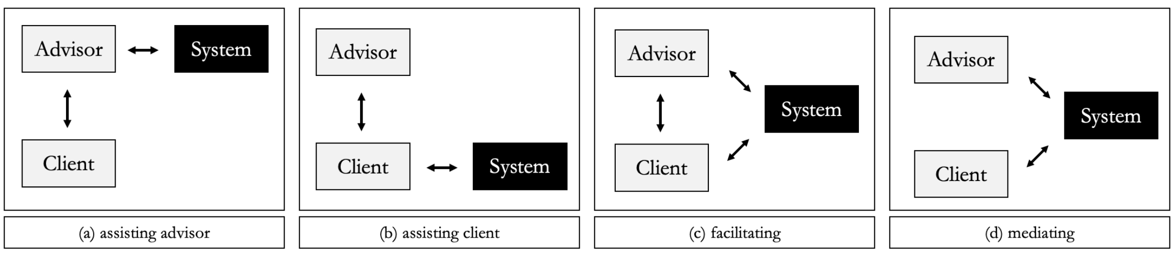


Figure 1. The four possible different archetypes of AI-based systems augmenting service encounters (De Keyser et al., 2019; Ostrom et al., 2019; Poser et al., 2022)

Mental Models

The concept of mental models is rather old and not specific to IT systems (Craik, 1943). The term *mental model* is widespread in the literature and the object of much research (Gentner & Stevens, 1983; Johnson-Laird, 1983). Nonetheless, a clear definition of the "mental model" is surprisingly rare in literature. We adapt the definition of Rouse and Morris (1986) and define the mental model as *"the mechanisms whereby humans can generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states".* Mental models do not have a firm boundary, and people can confuse similar devices with each other (Norman, 1983). People form mental models for any system they interact with, such as AI tools (Cai et al., 2019) or intelligent recommender systems (Kulesza et al., 2012). Further, mental models influence the reliance on AI systems (Nourani et al., 2021), where appropriate reliance is generally seen as desirable (Lee & See, 2004; Schemmer et al., 2023). Literature also discusses the possibility of inferring a user's model of a system directly from their actions (Brooks & Szafir, 2019; Yang et al., 2019), e.g., to improve human-robot interaction. If we transfer learnings about mental models in education and science to systems, we find that users have three possibilities when confronted with a new system (Greca & Moreira, 2000). First, users can try to interpret the system to what they already know, thus generating hybrid mental models. Second, users can learn the system's functionalities by heart—without creating mental models. Third, the user can generate a new mental model from scratch. While the third option might be the most desirable by the system designer, this is also the most far-fetched (Greca & Moreira, 2000). Therefore, when confronted with new systems, users either use existing mental models to build a new one for the new system or try to learn the functionalities by heart.

Figure 2 illustrates the widely acknowledged relationship of designers, users, and systems concerning conceptual and mental models (Norman, 2013). In short, system designers generally have a conceptual model of the system. The conceptual model is the designer's mental model of how users should interact with the system. However, the designers only communicate with the user through the system image, i.e., how the system presents itself. Therefore, aligning the *conceptual model* and *user mental model* is essential.

Without this alignment, the user will have a "wrong mental model" (Norman, 2013). To stress this, practitioners hand out guidebooks on how to ensure alignment with the mental models of users, such as Google's handbook on mental models (Google People + AI Research team, 2019). Aligning the conceptual and user mental models is subject to much research. However, research on mental models is rarely concerned with multi-party settings, such as financial advisory service encounters. The introduction of multimodal CAs in financial advisory service encounters leaves open the question of how users interact with these CAs and adapt their mental models. This is where this study comes in.

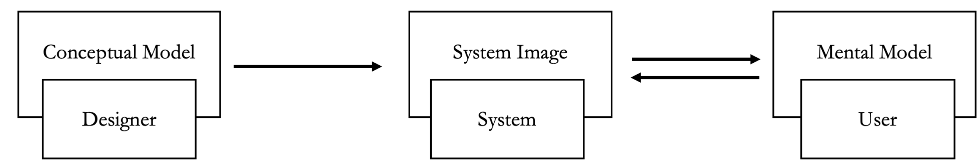


Figure 2. The designer and user models (Norman, 2013)

Methodology

This study is part of a larger research project involving two universities, two regional banks, and two technology partners in Switzerland. The project's goal is to explore and test the application of a CA in financial consultations. In particular, the CA should support the financial advisor and client throughout the financial consultations by automatically detecting clients' personal details, providing further information about financial terminology, recognizing and displaying stock transactions, or actively and continuously managing the portfolio's risk. Overall, the developed system was tested in financial consultations with 12 advisors and 24 clients between April 2022 and May 2022. To answer the introduced research questions, this study is based on qualitative data analysis (Saldaña, 2009), including the analysis of the consultation videos and interviews with the clients conducted after the financial consultation.

System Description

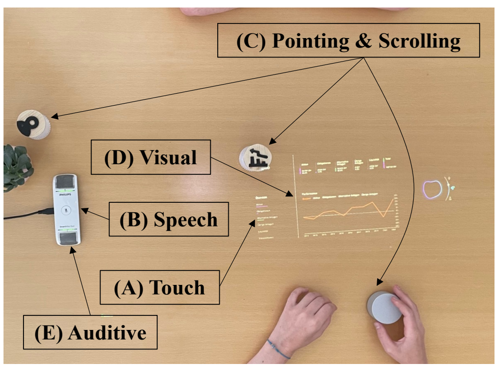
Based on a prior requirements elicitation with financial advisors and banking experts, the project team developed a CA called "MO" as a multimodal CA. The overall conceptual model of the designers of MO is to serve as a digital human-like assistant that can take over tedious tasks in all phases of the financial consultations, such as risk management or giving explanations. It should serve the client and advisor alike. Therein, MO is specifically designed to support the financial consultation in its five key phases: 1) the *Welcome Phase*, 2) the *Introduction Phase*, 3) the *Compliance Recording Phase*, 4) the *Portfolio Management Phase*, and 5) the *Farewell Phase*. In the *Welcome Phase*, MO greets the client and explains the goal of the consultation. In the *Introduction Phase*, MO listens to the spoken word of the client and advisor and takes note of relevant key aspects, relieving the advisor from cognitive intensive notetaking and ensuring full coverage of all relevant key aspects. In the *Compliance Recording Phase*, MO is concerned with notetaking risk-relevant key aspects and explaining technical terms to the client. In the *Portfolio Management* Phase, MO executes transactions (e.g., buying/ selling stocks) by voice command. In the *Farewell Phase*, MO farewells the client and gives the client concluding information about the consultation and the next steps. As displayed and summarized in Figure 3, MO provides various input and output modalities ranging from voice, touch, pointing, and scrolling to visual and auditive output.

MO understands and recognizes the Swiss German language so as not to disturb the natural flow of conversation during financial consultations. Besides allowing for direct commands (e.g., "Hey MO, can you please add ten Novartis stocks to the portfolio"), MO is designed in a "conversational" manner to be able to filter user intents from the conversation without the need for an explicit wake word (B). For example, in the *Introduction Phase*, MO can detect essential information about the client during normal conversation.[1] Further, MO allows for multiuser input by recognizing the spoken word of the advisor and client. However,

[1] Due to technical limitations, speech recognition during the *Introduction Phase* was simulated using an administrative user interface controlled in a Wizard-of-Oz approach (Dahlbäck et al., 1993), as speech often contained localized words, such as names or town names. In all other phases, proprietary natural language processing models were used for speech recognition.

MO not only understands Swiss German, but it is also able to provide auditive feedback in Swiss German (E). Herein, the auditive feedback ranges from short and straightforward utterances like “yes” or “ok, understood” to detailed responses and explanations. For instance, MO greets the client in the *Welcome Phase* and introduces itself and its tasks upon request. Additionally, MO can provide detailed explanations about financial terminology (e.g., bonds, obligations, or stocks) during the *Compliance Recording Phase*. Additionally, while some features require a sequential use of modalities, overall MO allows for parallel use of modalities and can combine the data input on many features. Examples include parallel data input via touch or speech, interactive visualizations by touching, pointing, or scrolling, and speech input that changes the visualization and prompts an auditive output.

Besides voice input and output, the visual representation of MO in the form of a blob is displayed on the table using a projector attached to the ceiling. MO can adapt its visual representation (including facial expressions) depending on the different phases of the advisory service and its specific role. For example, during the *Portfolio Management Phase*, MO monitors the portfolio risk and transforms itself into a risk meter. Additionally, using a Microsoft Kinect camera mounted next to the projector, MO displays information, like the client’s key data or the historical development of the portfolio, on the consultation table next to physical tokens (D). The projected information and visualizations can be easily arranged by moving and turning the physical tokens on the table. Further, a physical scroll dial allows for scrolling within the projection (C). Using the Microsoft Kinect camera, MO can recognize and allow for touch gestures (A). For example, while filling out a risk assessment questionnaire during the *Compliance Recording Phase*, advisors and clients can select their answers using touch input.



	Modality	Implementation in MO
Input	(A) Touch	Camera-based touch recognition
	(B) Speech	Swiss German intent recognition voice input
	(C) Pointing & Scrolling	Physical tokens and scroll dial
Output	(D) Visual	Table-top projection
	(E) Auditive	Swiss German voice output

Figure 3. Overview of the different modalities provided by MO

Data Collection and Analysis

Participants

In total, 24 financial consultations with real advisors and test clients were conducted to gather insights into how advisors and clients interact with MO. Each financial advisor had a half-day training with the system on a separate day and advised two clients over one day. We used thinkLets (Briggs et al., 2003), which were defined and refined in several pre-tests, to train the advisors. These thinkLets were created for all phases of the consultation and covered all aspects of MO. During the training, each advisor conducted at least one consultation as an advisor, one consultation as a customer, and one consultation as a spectator to get the complete picture. Financial advisors have training and tests like these regularly as part of their daily work.

The clients were recruited over the university’s website, social media platforms, and the professional network of the project partners using convenience sampling. The clients did not receive training, such as the advisors did. This was done to reflect the real-world case of a financial consultation, where an advisor would use MO regularly and be able to build expertise, but the client would be unfamiliar with MO. The clients were evenly split into male and female participants and were, on average, 28.7 years old (min: 20 years, max: 49 years). We specifically aimed to recruit not just students or financial experts but a mix of the general public. For their participation, the clients received compensation of an equivalent of $ 100.

Scenario Description and Procedure

The financial consultations were conducted in the offices of the two banking partners. Each advisor advised two clients. Each client had two financial consultations: one with MO and one conventional consultation without MO. Thus, each advisor had four consultations in the course of one day. In this study, we are only interested in consultations with MO. For the financial consultations, the clients received a fictional amount of an equivalent of $ 100,000 to invest. While the consultations were simulations and no real client money was at risk, we told them to treat this simulation as naturally as possible. No further restrictions or premises were made. Albeit not having the exact amount to invest in real life, many clients used this opportunity to gain insights into financial investment.

Further, clients were told that a new system was introduced in the financial consultation. This system is there to support the advisor and client alike. We also trained the advisors to have clients interact with the system to exploit the potential of MO, which all advisors followed. After the financial consultations, semi-structured interviews were conducted with clients. Questions were asked to get statements about the system and financial consultation, not singular system modalities. We also did not ask clients for explicit references to other systems. Clients made these statements on their own. Overall, being at the banking partners' offices and having few premises for the clients, the experience aimed to be as close to reality as possible.

Data Analysis

In this study, clients are abbreviated with *C01* … *C24* and assigned random numbers. On average, the consultations lasted 34.79 min (min: 23 min, max: 53 min). All consultations were filmed from four angles and audio recorded using clip-on microphones. After the consultations, clients were interviewed about their experiences and perceptions, which lasted an average of 60.9 min (min: 43 min, max: 91 min). Interviews were transcribed using intelligent verbatim transcription, and relevant quotes were translated into English.

The analysis of the interviews consisted of multiple rounds of coding with a focus on user interactions with the system. In the first round, one author applied open coding (Saldaña, 2009). This led to the observation that the clients faced various challenges during their interactions with MO and had varying reactions, expectations, and assumptions towards their interaction with such systems in financial consultations. We decided on a second coding round to increase the rigor and consolidate the initial impression. Two authors independently focused on specific aspects of the interactions by selecting the modalities as categories and codes. After the second coding stage, the codes were merged. Disagreements in coding were discussed among the two authors, and inclusion or exclusion into the merged coding was agreed on. The findings and all quotes presented in this study are based on the merged coding set. Afterward, all authors discussed the findings and conclusions present in this study. Further, one author analyzed the videos using BORIS (Friard & Gamba, 2016), a tool for video coding. All parts concerning the interaction with MO were coded and later used to discuss qualitative observations about the interaction with MO among all authors.

Results

The analysis of the interviews and videos yielded insights into clients' preferences and expectations for interacting with a multimodal, interactive CA during financial consultations. After the initial open coding, it became apparent that clients talked about the modalities individually and made individual references for each modality. Thus, we will present our findings for each modality one by one. Additionally, detailed observations from the videos provide further evidence of how clients interacted with such a system and what challenges arose. These observations include challenges using the camera-based touch recognition, limited use of MO's speech recognition capabilities, hesitations regarding the interaction with physical tokens, and positive reactions toward MO's visual and auditive output modalities. The findings are summarized in Table 1. In the following, we provide the observations and quotes from the interviews. Further, we list exemplary comparisons between clients' interaction with MO and other devices or systems.

Not All "Touch Screens" Are the Same

The touch interaction is implemented using a Microsoft Kinect camera mounted to the ceiling. It requires gentle touch movements with a single stretched-out finger for optimal touch recognition. However, many clients faced difficulties with this form of touch interaction. Even being explicitly instructed by the advisors, many clients tried to touch with an open hand, which was not recognizable by the camera. Additionally, some users then subsequently tried to touch harder on the table or tried to touch multiple times very quickly, which had no effect on the camera-based touch recognition. Both are common practices with usual touch interfaces. Similarly to such interfaces, the clients often tried to zoom in with two fingers or scroll, which is not supported by MO. We also observed that clients regularly refrained from touch interactions, even when the screen was pointed toward them until being directly asked by the advisor.

The interviews also supported the observation of problems and limitations with the touch interaction. When asked about their experiences, some clients stated that the touch interaction did not work because of how they tried to touch (e.g., C19 or C09) or mentioned system limitations (e.g., C24). Nonetheless, the clients also described the touch interactions as enjoyable and positive. Therein, despite being hesitant to interact with the system, the clients perceived the touch interactions as empowering and making them more independent (e.g., C04). Many clients also drew comparisons between the touch interactions supported by MO with known touch interfaces like *"tablets"* or *"iPads"* (e.g., C08, C17, C21, C23). More statements from clients supporting these findings can be found in Table 1 in row *(A) Touch*.

Let's Talk Simply

MO understands Swiss German and recognizes information from the conversational context. Nonetheless, it can be observed that many clients naturally started to speak High German with MO and recognized only after a while that the agents could also understand Swiss German. Even as some clients recognized MO's ability to understand the conversational context from the beginning, most clients still interacted with MO in a command-based way, like everyday household CAs, such as Siri or Alexa. That includes using wake words and simple or incomplete sentence structures, such as *"MO, buy Microsoft 100 shares*".

This tendency was also reflected in the interviews. For instance, C15 mentioned that "*it was like a Siri function"*. Additionally, many clients reported inhibitions to speak to MO (e.g., C04). It was also unclear if the clients should talk to MO or if this interaction was reserved for the advisor (e.g., C05). However, even when the advisors asked them to speak to MO, some clients still hesitated, fearing that MO would not understand them. For instance, C18 reported feeling uncomfortable talking to MO during the financial consultations because *"it was like when you visit someone at home, you don't feel like you're at home. You don't do what you want, you hold back. And that was just a little bit that."*. In summary, most clients did not max out MO's advanced speech recognition capabilities by relying on simple command- and wake-word-based interactions, issuing commands in High German, or not talking to the CA. More statements from clients supporting these findings can be found in Table 1 in row *(B) Speech*.

Don't Move It, Scroll It

The pointing and scrolling modality is implemented using two physical tokens. The two physical tokens align and rotate the image on the table to make the displayed information more interactive and movable. Followingly, clients and advisors can move the visualizations closer to themselves for an improved view. Additionally, clients and advisors can use the scroll dial to change their view of the displayed information. However, as observable in the videos, the clients did not move the physical tokens and rarely used the scroll dial. Even though the advisors introduced the functionality of the physical tokens and the scroll dial, almost all clients refrained from interacting with them.

The interviews revealed that despite the explanations from the advisors, many clients did not understand the role and functionality of the tokens. They were unaware that the tokens were used for table projection or scrolling. For instance, C23 mentioned that they *"didn't know that the round stuff was for scrolling"* and that they *"thought that was a doorstop and didn't know that the things, the salt shakers belong on there"*. Nonetheless, those clients who understood the functionality of the physical tokens perceived them as pleasant and recognized their benefits (e.g., C23 or C09). More statements from clients supporting these findings can be found in Table 1 in row *(C) Pointing & Scrolling*.

(A) Touch	*Observation:* Many clients had problems using the input modality touch. As MO relied on the Microsoft Kinect camera on the ceiling to detect touch interactions, a gentle touch movement with a single stretched-out finger was required. Instead, many clients tried to touch the table from above (which the system cannot recognize) or with multiple fingers. Commonly, they also tried zooming or scrolling gestures.
	• *"It didn't work for me sometimes because I had the whole hand [used*]" (C19) • *"Just as I said, a touch with the light didn't always work"* (C09) • *"I have been more independent because I have touched the things and not the consultant"* (C04) • *"The touch didn't work right"* (C23) • *"The surface of the tool was a bit too small for my fingers"* (C24)
	Comparisons: "iPad" (e.g., C17, C21), *"Tablet"* (e.g., C08, C23)
(B) Speech	*Observation:* Some clients used Standard German and were surprised that MO understands Swiss German. Additionally, many clients primarily issued commands and relied on wake words instead of using MO's conversational intent recognition ability.
	• *"I assumed that it does not understand Swiss German well, which means that one would have to speak English or High German"* (C03) • *"I noticed that it's Swiss German [...] usually it's in English or High German"* (C22) • *"And at first I did in High German because I'm so used to it always being High German. But then I realized that you could actually do it in Swiss German"* (C22) • "*So I'm really surprised that MO can speak and understand Swiss German so well"* (C14) • *"But afterward, I didn't do it because of the inhibition, and then he took it over"* (C14) • *"I had inhibitions about addressing MO directly"* (C04) • *"You didn't know exactly, should I talk to [MO], is [the advisor] talking now or I?"* (C05)
	Comparisons: "Siri" (e.g., C12, C20, C23), *"Alexa"* (e.g., C07, C12, C20, C23)
(C) Pointing & Scrolling	*Observation:* All clients refrained from touching the physical tokens. The scroll dial was only used after an explicit request by the advisor.
	• *"I found the wheel to turn the most useful."* (C09) • *"I didn't know that the round stuff was for scrolling."* (C23) • *"What was that little figure that defines the alignment?" (C18)* • *"I didn't understand why there were two tokens."* (C20) • *"I did, never got to turn it myself, but that looked very pleasant."* (C09)
	Comparisons: "Doorstop" (C23), *"Salt shakers"* (e.g., C23, C16, C17, C09), *"3D printed parts"* (e.g., C16), *"Mouse"* (e.g., C14, C02)
(D) Visual	*Observation:* The clients were amazed but, at the same time, distracted by the projection. Many of them stared at the screen for most of the consultation.
	• "It was cool with this projection" (C07) • "This digital tool, which was here on the table, particularly caught my attention." (C10) • "That projection there on the table, that was actually the highlight. For the simple reason that it was like such a surprise effect because I just so fully did not expect something like that." (C18) • "Also the projection, I have never been able to interact in this form before." (C22) • "It is again something different and special if one has that on the table." (C15)
	Comparisons: "Tablet" (C08), "PowerPoint Slides" (C07), "Website" (C09), "Hologram" (C13)
(E) Auditive	*Observation:* Many clients were surprised by MO's ability to give feedback and explain things in Swiss German.
	• *"I noticed that he speaks Swiss German; I wouldn't have thought that."* (C12) • *"It's cool when a tool even speaks Swiss German."* (C19) • *"Do I have to speak in dialect, with the MO? [...] So, what language should I speak now?"* (C01) • *"The voice sounds like Swiss German and sounds like a real voice"* (C16) • *"The fact that it's Swiss German and not robotic German means that it's also natural"* (C18)
	Comparisons: "Siri" (C23, C21, C12, C20, C05), *"Alexa"* (C23, C21, C08, C07, C12, C20), *"Google"* Translate Speech Output (C10)

Table 1. Summary of observations based on video recordings, statements by clients, and comparisons given during the interviews.

Novel Projections Distract

A projector is used for table-top projections. These table-top projections create the visual output of MO. As observed in the videos, many clients were amazed by MO's immersive projections and visualizations. However, being able to see their key data appear in front of them during the *Introduction Phase* or to view the portfolio allocation being displayed on the table, many clients continuously stared at the projection on the table. Hence, clients and advisors could often establish only limited eye contact. Advisors recognized this problem and tried to regain the client's attention, e.g., by using additional questions.

Some clients mentioned the lack of eye contact and stated they were distracted by the projection. Still, the consent amongst the clients was a positive attitude and feedback toward the immersive projection. Many clients state that they have never seen comparable projections or perceived them as unusual. For them, the table-top projections were a surprise they had not encountered before (e.g., C18). Some clients compared the projection with PowerPoint slides, websites, or a hologram. One client (C19) also compared it with a *"large iPad"* and was unsure about the benefits of a projection over a "*touchscreen integrated into the table"*. More statements from clients supporting these findings can be found in Table 1 in row *(D) Visual*.

Familiar Dialects Create a Familiar Environment

MO can give feedback and explanations in Swiss German. Similar to the voice recognition capabilities, MO is designed to speak Swiss German to let the clients and advisors interact with it in their natural language. When hearing MO's answer in Swiss German, many clients showed signs of surprise regarding changes in their facial and bodily expressions. For instance, many clients looked amused and smiled when MO introduced itself at the beginning of the financial consultation or when explaining financial terminology during the *Compliance Recording Phase*. It also became apparent that clients and advisors did not know how to behave during MO's auditive responses and looked around the room aimlessly.

These findings were also reflected in the interviews. Being used to consumer CAs, like Siri or Alexa, many clients did not expect MO to speak Swiss German. Nonetheless, they described MO's Swiss German responses as favorable and pleasant and said they were surprised by these capabilities. Especially in situations where the advisor and client both conversed in Swiss German, the clients appreciated the natural integration of MO's Swiss German responses into the conversation. For instance, C12 mentioned that *"it was good that it was the same language I speak with the consultant. That it was so on the same level. And I also found the language pleasant."*. More statements from clients supporting these findings can be found in Table 1 in row *(E) Auditive*.

Discussion

This study analyzed 24 financial advisory service encounters in the form of financial consultations with a CA and 24 associated interviews with clients. The CA is designed to *facilitate* the financial advisory service encounter by allowing both parties, the advisor and the client, to interact with it (De Keyser et al., 2019; Ostrom et al., 2019; Poser et al., 2022). Further, following recent findings from the literature on financial advisory service encounters (Dolata, Agotai, et al., 2019, 2020), the CA is designed to be multimodal. Overall, we find that even though it is being implemented as a *facilitating* archetype, the parties fall back to an *assisting* archetype of the CA. In the following, we first discuss the institutional setting as one likely reason for the unintended use. Additionally, we identify the problem of mental models as one of the underlying causes for this and call the arising phenomenon *secondary mental models*. Lastly, we end the discussion with the implications of the *secondary mental models*.

Shared System vs. Individual System

Past research has introduced different archetypes of AI-based systems, such as CAs, into service encounters, as presented in Figure 1. The CA in this study is designed to facilitate the service encounter (archetype (c)). However, the findings show that the CA is used divergent from the design to assist the advisor (archetype (a)) mainly. Therefore, the tool follows the situation (e) in Figure 4. In this situation, the CA can facilitate the encounter but is only used to assist one party. Therefore, the CA is used unintended in the multi-party setting. The reasons for that are manifold. In the following, we discuss the institutional setting of financial advisory service encounters as one probable reason.

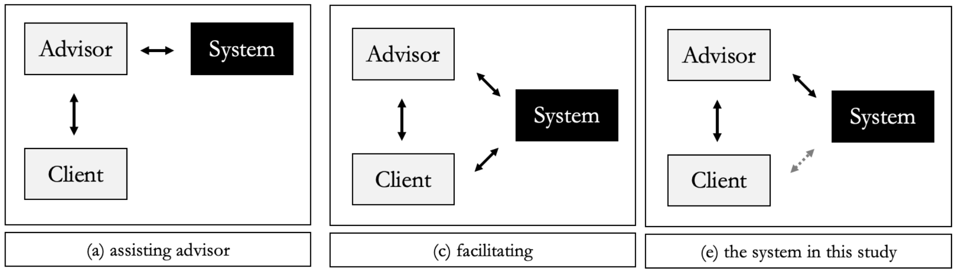


Figure 4. Configurations of AI-based systems in service encounters. Archetypes (a) and (c) are present in the literature (De Keyser et al., 2019; Ostrom et al., 2019; Poser et al., 2022). The system in this study is implemented as an archetype (c) but used as an archetype (a), leading to a situation presented in (e). The solid arrows indicate interaction, and the dashed arrows indicate indirect interaction.

An overt reason for the unintended use of capabilities is the institutional setting of financial advisory service encounters (Dolata, Agotai, et al., 2019; Dolata, Kilic, et al., 2019; Schmid et al., 2022). These settings have existing practices and social norms that people adhere to. As stated in the results, clients feel as if they come to the "advisor's home". When coming to someone's home, one would not simply start interacting with devices and equipment to one's liking. This implies that clients only interact with the system after explicit requests. However, the advisors explicitly requested clients to interact with the system, which, as shown by the results, did not fully work. This is discussed in-depth in the next section.

Further, clients do not have clear incentives to interact with the system. Interacting with a system for the first time often involves a recognizable cognitive effort (Weiss et al., 2015). In our case, the clients need to become acquainted with a complex, multimodal, and interactive CA while interacting with the advisor simultaneously. Several clients would rather have an advisor handle their system interaction. After all, the clients visit financial advisory service encounters to have less cognitive effort in their investment decisions. As stated by several clients, they instead have an advisor to interact with the system than themselves.

These findings have implications for the use of IT in financial advisory service encounters. Interactive systems have been shown to empower advisors (Boulus-Rødje, 2018; Giesbrecht et al., 2014), reduce cognitive overload (Giesbrecht et al., 2015), and enhance the persuasiveness of the advisory service (Dolata & Schwabe, 2018). However, most of this research is conducted on the advisor's side. Our findings show that while IT systems can bring advantages to financial advisory service encounters, they also run the risk of being used in an unintended way. Further, research on CAs proposes that CAs can reduce the cognitive load on the advisor and increase service quality (Dolata, Kilic, et al., 2019). However, if the advisor is the sole user of the system, they not only have to satisfy the client but also interact with the system. Contrary to existing research, this might increase the cognitive load of the advisor. Therefore, AI-based systems not only need to be able to take over unwanted tasks but also integrate well into the existing practices of advisors (Dolata, Steigler, et al., 2019; Heyman & Artman, 2015). Further, studies look into the design of CAs in financial advisory service encounters (Bucher et al., 2022; Schmid et al., 2022). While these studies already contained the client's view on this matter, they could not have clients interact with a live system.

We find that clients do not interact with the systems because of the institutional context. This might come as little surprise when introducing common IT tools, such as computers or tablets. However, AI systems are different, especially the system observed in this study, which the client and advisor can interact with in parallel. This makes the situation more surprising, as the client could interact with the system without hindering the advisor. In the next section, we look at this situation from a different angle and analyze the situation according to the interaction with complex technological systems.

Secondary Mental Models in Financial Advisory Service Encounters

In general, clients rarely interacted with the system, for example, after being explicitly requested by the advisor. Once the clients interacted with the system, many issues occurred. For example, as shown by our results, clients had problems with the touch modality, started speaking in High German instead of Swiss German, or did not know how to use the tokens. The clients are missing a suitable mental model to interact with the system. As indicated by the many references to other systems, users tried to derive a mental model from their existing mental model (Greca & Moreira, 2000). However, given that the interaction with the system was only indirect via the advisor, this did not work.

This indicates that the common notion of the designer's conceptual and user's mental models (Norman, 2013) might not be enough in settings with multiple users. Based on the findings in the preceding section, we find that financial advisory service encounters consist of two users with different roles. The advisor is the *primary user* who primarily interacts with the system. The client is the *secondary user*. The client primarily interacts with the advisor and rarely with the system. Most of the time, the client interacts with the system transitively, i.e., with the advisor, who interacts with the system. This is illustrated in Figure 5. This extension is introduced and explained in the following.

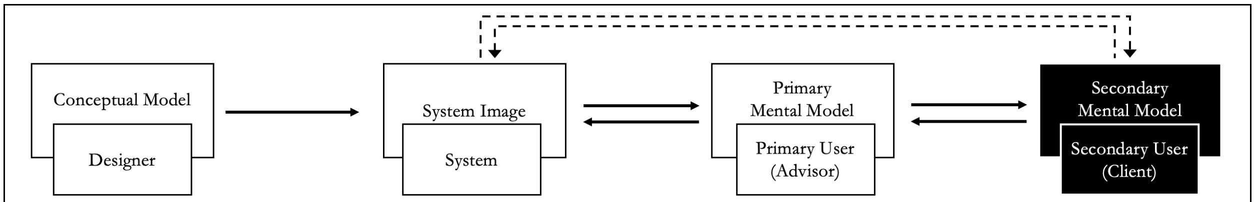


Figure 5. The interplay of different mental models in financial advisory service encounters. The white rectangles are based on Norman (2013), and the black rectangles are introduced in this study. The solid arrows indicate the actual interaction, and the dashed arrows indicate indirect interaction.

While designers have their mental models when developing systems, they cannot directly communicate with the end users. Therefore, the designers build the systems according to their conceptual model and must ensure that the end users interact with the systems according to them. That is done by the system presenting itself correctly, i.e., the system image. Practitioner guides like Google's handbook on mental models (Google People + AI Research team, 2019) can guide designers in creating a system image that transports their model to the end users. However, in financial advisory service encounters, we also need to consider the mental model of the *secondary user*, which we call the *secondary mental model*. As the clients primarily interact with the advisor and not with the system, their mental model can only be built by observing the interaction between the advisor and the system. This implies that the conceptual model needs to be transported via the system image and the *primary mental model* to the *secondary mental model*. On each step of this "mental model propagation", different issues arise. HCI research is primarily concerned with the system image. However, we propose an extension to the original framework of Norman (2013) for collaborative settings like financial advisory service encounters.

We can explain many of the clients' behaviors with the extension of a *secondary mental model*. One example from our finding is the usage of wake words. Most advisors preferred to use wake words and direct addressing of the system. However, the system can also extract information and intents from context. Nonetheless, as most clients were never exposed to this possibility, these clients never could adapt their mental model according to that. Therefore, the *secondary mental model* of the clients strongly relies on the mental model of the advisors, leading the clients to not know about the possibility of not using a wake word. Another observation is the missing mental model for the physical tokens used for pointing. A likely reason for that is that the advisors used this token exactly once at the beginning of the service encounters and rarely moved them around. Even though this modality arguably is relatively intuitive to use, as there was no interaction with it, the clients could not build a mental model for it. This is contrary to the token used for scrolling. Scrolling is used quite often throughout the service encounters. As clients can observe the advisor regularly using this function, the client's *secondary mental model* was able to arise for that functionality. Another example and a more interesting case is the touch modality. The advisors regularly and successfully used the touch modality. However, clients seemed to have a problem using the touch modality, even after being able to observe the advisor. One possible reason is that the clients were not attentive enough to the advisor's interactions. However, this contradicts the fact that even after being told by the advisor how to use this modality, many clients still had issues with it. This indicates that the *secondary mental model* might be weaker than the *primary mental model*. Changing known interaction patterns might be hard to achieve with the *secondary mental model*.

The system in this study had several features that can be used in parallel and combined (Nigay & Coutaz, 1993). However, the clients compared many unimodal systems, devices, or physical objects to our system, such as Siri, Alexa, or salt shakers. This indicates that the clients could not build one mental model for the system. Instead, they had many mental models for subparts of the system. They primarily focused on the unimodal interaction sequentially. Many of the mental models of the clients were hindering some of the interplays. As presented in the results, one client compared the projection to a "PowerPoint" presentation. Thus, touching the PowerPoint presentation does not seem intuitive. Also, some users did not have a mental model for some modalities. For example, several users had no mental models for the tokens. Thus, these users did not interact with the tokens in the first place. Additionally, many people referred to speech input and output systems like Siri or Alexa and used wake words. However, for the system's full potential, the clients would need a holistic mental model instead of many single mental models. Someone who thinks of the system as "Siri" + "large iPad" + "Mouse" + "PowerPoint" will never use the full potential of the system.

Possible reasons for the difference in perception are that the designer's conceptual model of a digital human-like assistant was too complex to propagate to the *secondary mental model* or clients did not have the time to adapt their mental models. First, the conceptual model was that of a digital human-like assistant that can take over tedious tasks, such as risk management, explanations, etc. Clients could only build their mental model by observing the system and the advisor's interaction. This conceptual model might be too complex to propagate to the *secondary mental model*. It seems that without being able to build a fitting mental model, clients fell back to their existing simple mental models and interacted in an unimodal way. Second, the clients only interacted once and could not take the time to adapt their mental model. The setting of a financial advisory service does not allow for an extensive tryout. Therefore, the system image alone might not propagate the conceptual model to the clients. While the advisors regularly conduct financial consultations, financial consultations are not an everyday chore for the clients. Thus, they do not have the chance to adapt their mental model over time but rather have it as a one-off situation. More interactions with the system would likely make it easier for the clients to create a mental model of the system. However, multiple interactions are not feasible in settings like financial advisory service encounters.

To conclude, we propose an extension to the designer conceptual and user mental models framework (Norman, 2013) for specific settings like financial advisory service encounters, which we call the *secondary mental model*. Additionally, the discussion on inferring a user's mental model of a system directly from their actions (Brooks & Szafir, 2019; Yang et al., 2019) implies the existence of a *secondary mental model*, as presented in this work. However, these discussions are limited to the single-user setting. Thanks to this extension, we can explain the behavior of *secondary users* in multi-party settings involving AI-based systems. In the following section, some additional challenges with multimodality are discussed.

Implications of Secondary Mental Models

The phenomenon of *secondary mental models* has implications beyond financial advisory service encounters. *Secondary mental models* will likely arise if there are asymmetries between the users. Most notable is the role asymmetry of primary and *secondary users*. This difference in roles creates knowledge and possible information asymmetry about the interaction with the system, i.e., one user has more experience interacting with the system. Many people may have been in a similar situation before. When one person tries to explain how a system works to another, the second person will only form a mental model of what the first person showed and explained. Thus, the second person will have a *secondary mental model*.

Further, the context of the interaction plays a role. In the case of financial advisory service encounters, clients often resisted interacting with the system directly. Something similar could happen in doctor-patient settings, where patients might resist interacting directly with the system. Another example where the context might play a role is used car sales, where AI systems can make price estimations (Eckhardt et al., 2022). While the used car dealer might have interacted with the system numerous times, a private seller might only interact with it once. The used car dealer can try to use this *primary mental model* to persuade the private sellers from a deviation from the AI price estimation.

The above example already hints towards the influence of mental models on appropriate reliance, i.e., the desired behavior of users (Lee & See, 2004; Schemmer et al., 2023). Research shows that existing mental models influence the reliance on AI systems in single-user settings (Nourani et al., 2021). However, there is more to consider in multi-party settings. Different parties might have different mental models, which might result in a difference in reliance. Specifically, there is potential that only one party relies on the system—leading to the notion of *partial reliance*. In our example of financial consultations, if the advisor relies on the system and the client does not, the client will not rely on information and assistance from the system, ultimately questioning the quality of the overall consultation. If the client relies on the system and the advisor does not, the advisor will not use the system's full potential; thus, the system does not yield the expected positive effect on the consultation. We need to ensure fitting mental models for an appropriate reliance of both parties. If only one party relies on the system, this has the potential to create a failed consultation. However, since the *secondary mental model* is based on the *primary mental model*, the reliance of the *primary user* might influence the reliance of the *secondary user*. For example, if the advisor does not rely on the system, the client will likely not rely on the system, and vice versa. This is also subject to the social influence of the interaction between client and advisor and the context of financial advisory service encounters. Overall, the concept of *secondary mental models* offers a new behavioral perspective to study AI reliance in multi-party contexts.

The findings in this study also point towards design aspects that similar systems should consider. First, developers of AI systems in multi-party settings should consider the *secondary mental models* of the *secondary users*. That is, to create a system image that also transports to the *secondary users* by making it easy to grasp and more straightforward than system images of single-user AI systems. Second, developers should not change several aspects of existing technology at once. This comes down to the discussion of innovation and improvement. While technology is often innovation-driven, complex AI-based systems introduced in multi-party settings should aim to improve existing and known technology. Finally, developers should consider the roles of primary and *secondary users* in multi-party settings and aim to design the system explicitly for these roles. The system should be able to differentiate between the users and have different interactions with these users. This includes, for example, more explanations for the *secondary user* or addressing which user should interact with the system at what time. This study hints at some design aspects that should be further investigated in follow-up studies.

In summary, *secondary mental models* are a new lens to explore multi-party settings. They occur in many settings and can influence the reliance on AI systems. Therefore, developers need to consider several design aspects to transport the conceptual model to the mental model of the *secondary user*.

Conclusion

This study introduces a multimodal, interactive CA called MO in financial advisory service encounters. While the literature indicates that this is the next logical step to improve service quality and reduce the cognitive load on the advisor, our findings indicate that this is not straightforward. To answer ***RQ1***, we find that MO can facilitate the setting for co-located financial advisory service encounters but is only used to *assist* the advisor. This creates a setting where the advisor is the *primary user,* and the client is the *secondary user*. Additionally, to answer ***RQ2***, we introduce the concept of a *secondary mental model* as an extension to the existing designer conceptual and user mental models framework for advisory settings. Following this framework, we can explain our observations in the results and see the problem of multimodal systems in financial advisory service encounters.

This study comes with some limitations. First, our study is exploratory. We discussed several possible confounders for *secondary mental models*, such as language or dialects, technical interface challenges, or rituals of financial advisory service encounters. However, the discussion on these might be extended or more possible confounders explored. Further, we focus on financial advisory service encounters as one form of multi-party settings. While our findings might be transferable to other settings, such as a doctor's appointment, this cannot concludingly be done in this study. Finally, we observe only one system with one design configuration. Nonetheless, many of our findings can be generalized.

In the future, we propose to explore the *secondary mental model* more. For one, unintended use might not always be undesirable but also create value. Therefore, future research should study the challenges of *secondary mental models* and the opportunities that these differences in mental models can create. Further, future research could observe this phenomenon in other settings, such as car sales platforms or health treatments. While in financial advisory service encounters, a wrong *secondary mental model* might lead to issues in the interaction of clients with the system, in domains like health treatment, a wrong mental model by the patient could directly impact the patient's life. Further, while we remained in the synchronous setting, future research can be conducted on asynchronous settings.

References

Auh, S., Bell, S. J., McLeod, C. S., & Shih, E. (2007). Co-production and customer loyalty in financial services. *Journal of Retailing*, *83*(3), 359–370.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). *Managing artificial intelligence*. 18.

Bickmore, T., & Cassell, J. (2005). Social Dialogue with Embodied Conversational Agents.

Boulus-Rødje, N. (2018). In Search for the Perfect Pathway: Supporting Knowledge Work of Welfare Workers. *Computer Supported Cooperative Work (CSCW)*, *27*(3), 841–874.

Briggs, R. O., De Vreede, G.-J., & Nunamaker Jr, J. F. (2003). Collaboration engineering with ThinkLets to pursue sustained success with group support systems. *Journal of Management Information Systems*, *19*(4), 31–64.

Brooks, C., & Szafir, D. (2019). *Building Second-Order Mental Models for Human-Robot Interaction* (arXiv:1909.06508). arXiv. http://arxiv.org/abs/1909.06508

Bucher, A., Staehelin, D., Dolata, M., & Schwabe, G. (2022). Form Follows Function: Designing For Tensions Of Conversational Agents In Service Encounters. *ECIS 2022 Research Papers*.

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 104:1-104:24.

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., & Agrawal, P. (2019). Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior*, *93*, 309–317.

Comes, T., & Schwabe, G. (2016). From Fuzzy Exploration to Transparent Advice: Insights into Mobile Advisory Services. *Hawaii International Conference on System Sciences*.

Craik, K. J. W. (1943). *The nature of explanation* (pp. viii, 123). University Press, Macmillan.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. *Knowledge-Based Systems*, *6*(4), 258–266.

De Keyser, A., Köcher, S., Alkire (née Nasr), L., Verbeeck, C., & Kandampully, J. (2019). Frontline Service Technology infusion: Conceptual archetypes and future research directions. *Journal of Service Management*, *30*(1), 156–183.

Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, *23*(1), 96–138.

Dolata, M., Agotai, D., Schubiger, S., & Schwabe, G. (2019). Pen-and-paper Rituals in Service Interaction: Combining High-touch and High-tech in Financial Advisory Encounters. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24.

Dolata, M., Agotai, D., Schubiger, S., & Schwabe, G. (2020). Advisory Service Support that Works: Enhancing Service Quality with a Mixed-reality System. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 120:1-120:22.

Dolata, M., Kilic, M., & Schwabe, G. (2019, January 8). When a computer speaks institutional talk: Exploring challenges and potentials of virtual assistants in face-to-face advisory services.

Dolata, M., Schenk, B., Fuhrer, J., Marti, A., & Schwabe, G. (2020). When the system does not fit: Coping strategies of employment consultants. *Computer Supported Cooperative Work*, *29*(6), Article 6.

Dolata, M., & Schwabe, G. (2018). Don't be afraid! Persuasive Practices in the Wild. *Computer Supported Cooperative Work (CSCW)*.

Dolata, M., Steigler, S., Nüesch, F., Schock, U., Agotai, D., Schubiger, S., Kilic, M., & Schwabe, G. (2019). Welcome, computer! How do participants introduce a collaborative application during face-to-face interaction? *Proc. Intl. Conf. on Human-Computer Interaction - INTERACT*.

Eckhardt, S., Knaeble, M., Bucher, A., Staehelin, D., Dolata, M., Agotai, D., & Schwabe, G. (2023). *"Garbage In, Garbage Out": Mitigating Human Biases in Data Entry by Means of Artificial Intelligence*. Proc. Intl. Conf. on Human-Computer Interaction - INTERACT.

Eckhardt, S., Sprenkamp, K., Zavolokina, L., Bauer, I., & Schwabe, G. (2022). Can Artificial Intelligence Help Used-Car Dealers Survive in a Data-Driven Used-Car Market? 115–127.

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, *14*(1), 57–74.

Faisal, A., Kamruzzaman, M., Yigitcanlar, T., & Currie, G. (2019). Understanding autonomous vehicles. *Journal of Transport and Land Use*, *12*(1), 45–72.

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Psychology Press.

Giesbrecht, T., Comes, T., & Schwabe, G. (2015). Back in Sight, Back in Mind: Picture-Centric Support for Mobile Counseling Sessions. *Proc. Conf. Computer Supported Cooperative Work*.

Giesbrecht, T., Schenk, B., & Schwabe, G. (2014, June 11). Learning with facilitation affordances: The case of citizens' advice services. *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*.

Gnewuch, U., Morana, S., & Mädche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *ICIS 2017 Proceedings*.

Golec, J. H. (1992). Empirical Tests of a Principal-Agent Model of the Investor-Investment Advisor Relationship. *Journal of Financial and Quantitative Analysis*, *27*(1), 81–95.

Google People + AI Research team. (2019). *People + AI Guidebook*. https://design.google/ai-guidebook

Greca, I. M., & Moreira, M. A. (2000). Mental models, conceptual models, and modelling. *International Journal of Science Education*, *22*(1), 1–11.

Heinrich, P., Kilic, M., Aschoff, F.-R., & Schwabe, G. (2014). Enabling relationship building in tabletop-supported advisory settings. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 171–183.

Heinrich, P., Kilic, M., & Schwabe, G. (2014). Microworlds as the locus of consumer education in financial advisory services. 18.

Heyman, S., & Artman, H. (2015). Computer Support for Financial Advisors and Their Clients: Co-creating an Investment Plan. *Proc. Conf. Computer Supported Cooperative Work and Social Computing*, 1313–1323.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Johnson-Laird, P. N. (1983). Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press.

Jungermann, H. (1999, February 1). *Advice Giving and Taking*.

Kilic, M., Dolata, M., & Schwabe, G. (2017). *Why do you ask all those questions? Supporting client profiling in financial service encounters*. Hawaii International Conference on System Sciences.

Kim, G., & Kim, H. C. (2011). Designing of multimodal feedback for enhanced multitasking performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3113–3122.

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent Virtual Agents* (pp. 329–343). Springer.

Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012). Tell me more? The effects of mental model soundness on personalizing an intelligent agent. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1–10.

Laumer, S., Gubler, F. T., Racheva, A., & Maier, C. (2019). *Use cases for conversational agents: An interview-based study*.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. Human Factors, 46(1), 50-80.

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based Digital Assistants: Opportunities, Threats, and Research Perspectives. *Business & Information Systems Engineering*, *61*(4), 535–544.

Mørck, P., Langhoff, T. O., Christophersen, M., Møller, A. K., & Bjørn, P. (2018). Variations in Oncology Consultations: How Dictation allows Variations to be Documented in Standardized Ways. *Computer Supported Cooperative Work (CSCW)*.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94, 72–78.

Nigay, L., & Coutaz, J. (1993). A design space for multimodal systems: Concurrent processing and data fusion. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 172–178.

Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

Norman, D. A. (1983). Some Observations on Mental Models. In *Mental Models*. Psychology Press.

Nourani, M., Roy, C., Block, J. E., Honeycutt, D. R., Rahman, T., Ragan, E., & Gogate, V. (2021). Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. *26th International Conference on Intelligent User Interfaces*, 340–350.

Novak, J., & Schmidt, S. (2009). When Joy Matters: The Importance of Hedonic Stimulation in Collocated Collaboration with Large-Displays. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2009* (Vol. 5727, pp. 618–629). Springer Berlin Heidelberg.

Nueesch, R., Zerndt, T., Alt, R., & Ferretti, R. G. (2016). Tablets Penetrate the Customer Advisory Process: A Case from a Swiss Private Bank. *BLED Proc.*, 18.

Nussbaumer, P., Matter, I., & Schwabe, G. (2012). "Enforced" vs. "Casual" Transparency—Findings from IT-Supported Financial Advisory Encounters. *ACM Trans. Management Information Systems*, *3*(2).

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences. In *Handbook of Service Science, Volume II* (pp. 77–103). Springer International Publishing.

Poser, M., Wiethof, C., & Bittner, E. (2022, May 19). Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions.

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *Journal of Medical Internet Research*, *19*(5), e6553.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J.-C., McTear, M., Raman, T., Stanney, K. M., Su, H., & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM*, *47*(1), 57–59.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*, 349–363.

Saldaña, J. (2009). *The coding manual for qualitative researchers* (pp. xi, 223). Sage Publications Ltd.

Schemmer, M., Kuehl, N., Benz, C., Bartos, A., & Satzger, G. (2023). Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 410–422.

Schmid, D., Staehelin, D., Bucher, A., Dolata, M., & Schwabe, G. (2022). Does Social Presence Increase Perceived Competence? Evaluating Conversational Agents in Advice Giving Through a Video-Based Survey. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1–22.

Schwabe, G., & Nussbaumer, P. (2009, June 10). *Why IT is not being used for financial advisory*. 17th European Conference on Information Systems (ECIS 2009), Verona.

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. (2020). Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, *30*, 1–18.

Soubutts, E., Ayobi, A., Eardley, R., McNaney, R., Cater, K., & O'Kane, A. A. (2022). Amazon Echo Show as a Multimodal Human-to-Human Care Support Tool within Self-Isolating Older UK Households. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–31.

Tudor Car, L., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y.-L., & Atun, R. (2020). Conversational agents in health care: Scoping review and conceptual analysis. *Journal of Medical Internet Research*, *22*(8).

Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, *36*, 189–195.

Weiss, B., Wechsung, I., Kühnel, C., & Möller, S. (2015). Evaluating embodied conversational agents in multimodal interfaces. *Computational Cognitive Science*, *1*(1), 6.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). *A new chatbot for customer service on social media*. *2017-May*, 3506–3510. Scopus.

Yang, T., Meng, Z., Hao, J., Zhang, C., Zheng, Y., & Zheng, Z. (2019). *Towards Efficient Detection and Optimal Response against Sophisticated Opponents* (arXiv:1809.04240). arXiv.