Designing a Feedback Loop Between a Human and Their AI Clones for Science Communication in Museums

Shutaro Aoyama
Sony Computer Science Laboratories Inc.
Tokyo, Japan
Department of Computer Science
Columbia University
New York, New York, USA
shutaro.aoyama@gmail.com

Kiyoshi Suganuma
Yamaguchi Center for Arts and Media [YCAM]
Yamaguchi, Japan
Sony Computer Science Laboratories, Inc.
Tokyo, Japan
suga@ycam.jp

He Jiang
Miraikan - The National Museum of Emerging Science and Innovation
Japan Science and Technology Agency
Tokyo, Japan
he.jiang.se@jst.go.jp

Shunichi Kasahara
Sony Computer Science Laboratories, Inc.
Tokyo, Japan
Okinawa Institute of Science and Technology Graduate University
Okinawa, Japan
kasahara@csl.sony.co.jp

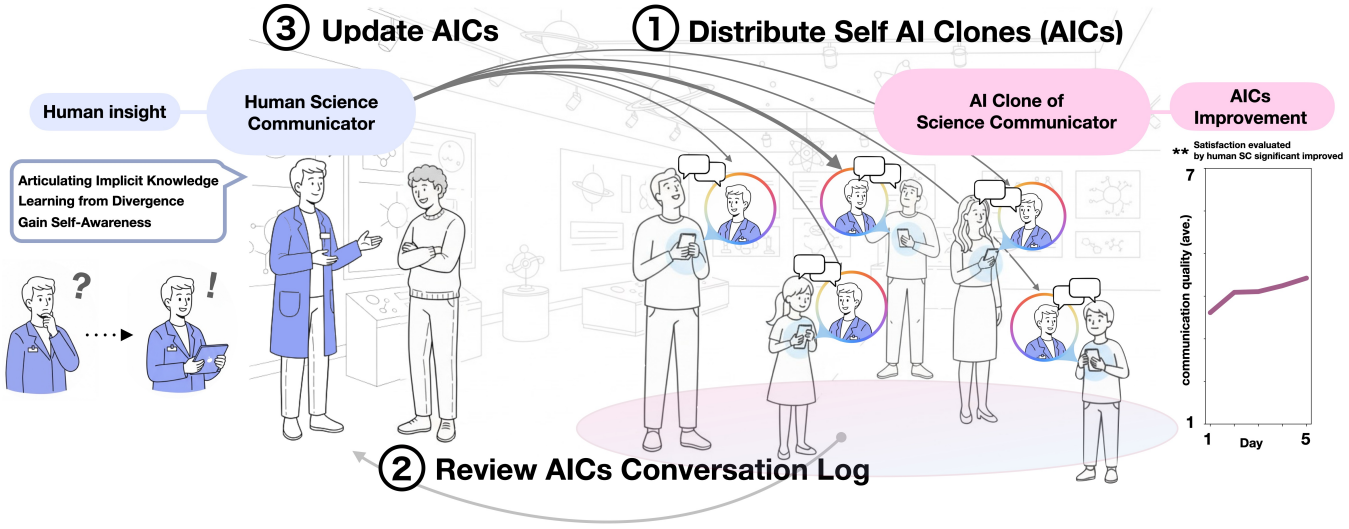

Figure 1: Iterative AI Clone (AIC) behavior refinement in a museum deployment. The system operates in three stages: (1) Deploy AICs: AIC science communicators engage with museum visitors via mobile devices; (2) Review Interaction Logs: Human science communicators review conversation transcripts and reflect on their AIC's performance; (3) Refine Behavior Descriptions: Humans refine their AIC's behavior descriptions to update its behavior. The graph shows AIC improvement over the 5-day deployment: communication quality, as evaluated by source user reviews, showed improvements across multiple dimensions. Iterative refinement also helped source users articulate implicit communication strategies, learn from AICs' divergent behaviors, and gain self-awareness, giving rise to diverse relationship patterns between source users and their AICs.

Abstract

AI Clones (AICs) hold promise for augmenting communication, yet most AICs remain static snapshots that drift out of sync with their evolving source users. Iterative refinement of behavior descriptions offers a potential mechanism for improvement: users review their AIC's interactions, reflect on its performance, and directly edit the AIC's behavior descriptions to update its behavior. We investigated this process in a five-day field deployment with six professional

CUI '26, Bremen, Germany

ACM ISBN 979-8-4007-2741-2/26/07
https://doi.org/10.1145/3816046.3816205

science communicators (SCs) at a science museum, making their AICs available in over 450 conversation sessions with visitors on multiple mobile devices. We analyzed conversation logs of AICs, visitor surveys, daily questionnaires from the SCs, and interviews with SCs.

Results reveal a marked source–visitor asymmetry. Source users' evaluations of their AICs showed significant improvement across nine of ten communication dimensions, whereas visitor ratings showed no overall change and reached significance for only one of six AICs after FDR correction. Because the SC and visitor questionnaires were built from parallel items, the two sides are directly comparable. Iterative refinement thus appears to do more for the source user's relationship with their clone than for the people the clone serves: it helped source users articulate implicit communication strategies, learn from AICs' divergent behaviors, and gain self-awareness about their professional identities. Furthermore, we found that participants adopted distinct stances toward their AICs, using them for self-reflection and for exploring idealized selves. We synthesize these observations into design implications for supporting diverse relationship patterns in AIC systems.

CCS Concepts

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

Keywords

Human-AI Interaction, AI-mediated communication, AI Clone, Human-in-the-loop, Science Communication

ACM Reference Format:

Shutaro Aoyama, Kiyoshi Suganuma, He Jiang, and Shunichi Kasahara. 2026. Designing a Feedback Loop Between a Human and Their AI Clones for Science Communication in Museums. In *ACM Conversational User Interfaces 2026 (CUI '26), July 21–24, 2026, Bremen, Germany*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3816046.3816205

1 Introduction

Emerging technologies have enabled computers to replicate individuals' linguistic styles [10, 40], representations [8, 43], and thought patterns [16, 44]. These advances have enabled AI Clones (AICs) that could fundamentally transform and augment human communication [22, 27]. Following Lee et al. [27], we define AI Clones (AICs) as AI systems that (1) represent aspects of real-world individuals, (2) function as interactive technologies, and (3) operate as AI-driven systems based on personal data.

AIC interactions from existing research can be organized into three directions of influence. The first involves training and tuning, where user data shapes the AIC's behavior (User → AIC) [9, 12, 16, 19, 33, 38, 44]. The second, which has received substantial attention, examines the AIC's function as a delegate to third parties (AIC → Others). Studies show that while these AICs can enhance a user's social presence [28, 29], they also introduce reputational risks and diminish the user's sense of control [11, 28, 29]. The third examines the AIC's feedback on the user (AIC → User) [3, 4, 13, 24, 45].

Existing accounts treat the three directions of influence around AICs in isolation, overlooking the iterative design processes that could connect them. Without a continuous feedback loop, an AIC becomes a static "snapshot" that de-synchronizes as its user changes, foreclosing opportunities for the AIC's improvement. In this study, we examine AICs that enable iterative refinement: the system makes interaction traces available for user review (AIC → User), and users can edit the AIC's behavior descriptions to update its subsequent behavior (User → AIC), creating conditions for iterative improvement in the AIC's communication quality.

Although iterative refinement of AICs has been discussed conceptually [22], empirical investigations of this process and its effects remain limited. We therefore hypothesize that such iterative refinement could not only improve AIC communication quality but also prompt source users to gain insights and reorganize how they perceive their AICs. Our research is guided by the following questions:

- **RQ1:** Do AICs exhibit improvements in communication quality over iterative behavior refinement?
- **RQ2:** How does iterative AIC refinement affect source users and their relationships with their AICs?

To investigate these questions, we developed an AIC system and conducted a 5-day in-the-wild study with six professional Science Communicators (SCs) at Miraikan – The National Museum of Emerging Science and Innovation in Tokyo, Japan. This setting was chosen because science communication emphasizes interactive and personalized dialogue, providing a rich context for observing the effects of a personalized feedback loop. The SCs' AICs were deployed on ten mobile devices for museum visitors, who could converse with the AICs about the exhibits. Each day, the SCs reviewed transcripts of their AIC's conversations with visitors and directly edited the AIC's behavior descriptions to update its behavior. Visitors provided survey feedback on their experience with the evolving AICs (approx. 450 total responses throughout 15 days of deployment). SCs also completed daily Likert-scale surveys about their AIC, followed by a semi-structured interview at the end of the study. We analyzed interaction logs, visitor questionnaires, daily SC surveys, and interviews.

Our mixed-methods analysis yielded three findings. First (RQ1), we observed an asymmetry between source and visitor evaluations: across the parallel SC and visitor items (Table 2), SCs' daily reviews moved favorably on nine of ten dimensions, whereas visitor ratings showed no overall change and reached significance for only one of six AICs after FDR correction. Second (RQ2), iterative refinement affected the source users themselves: it helped them articulate implicit communication knowledge, learn from AICs' divergent behaviors, and gain self-awareness as science communicators. The refinement loop's primary work appears to be relational and reflective rather than purely performance-improving. Third (RQ2), source users developed heterogeneous endline relationships with their AICs, ranging from a "distorted mirror" for self-reflection to an "ideal alter ego."

The contributions of this paper are threefold. First, we **present empirical findings from a five-day in-the-wild deployment** of iterative AIC refinement with professional science communicators and over 900 museum visitors. Second, we **surface an asymmetric improvement pattern**: broad improvement as perceived by source users, alongside visitor-perceived improvement for only one agent. Building on this, we **reframe iterative refinement as a**

relational process for the source user (supporting self-awareness, articulation of implicit knowledge, and professional identity work) that is partially decoupled from communicative gains for the people the clone serves. Third, we **derive design implications** for diverse modes of self-exploration in AIC systems, informed by the heterogeneous Actual-Anchored and Ideal-Anchored relationships we observed.

2 Related Work

2.1 AI Clones

Research on AICs is an emerging area, with most work appearing only in the past few years. This research has explored three primary directions of influence, each with distinct challenges and opportunities. The first direction, from user to AIC, concerns the foundational tuning process where AICs learn from user data to mimic someone's behavior [9, 12, 16, 19, 33, 38, 42, 44]. A key challenge identified in this work is the "snapshot problem": AICs trained on historical data become static representations that fail to evolve with the user [27].

The second direction, from AIC to others, investigates AICs as social delegates [28, 29, 37]. Leong et al.'s "Ditto" system [29] demonstrated that AICs attending meetings on behalf of absent users gained greater trust when they closely resembled the user's communication style. Similarly, Lee et al.'s "IRL Dittos" [28] showed that while AICs could facilitate serendipitous social connections in physical spaces, their occasionally unnatural responses risked being misattributed to the user's identity. These findings collectively demonstrate that fire-and-forget delegation is untenable; continuous user oversight remains essential.

The third direction, from AIC to user, explores how AICs influence their creators through self-reflection and behavioral change [3, 4, 13, 24, 45]. For example, Fang et al. [13] developed a system generating messages from an idealized future self in the user's own voice, demonstrating improvements in self-esteem.

Despite these advances, existing research has treated each direction as either isolated or, at best, as simple bidirectional exchanges. The dynamics of circular interaction—where an AIC's experiences continuously inform the user, whose evolving intentions reshape the AIC in turn—remain unexplored empirically. This gap is particularly striking given that many of the challenges identified in prior work stem directly from this absence of feedback loops. The snapshot problem [27], reputational risks [29], and the tension between fidelity and ethics [33] all intensify without mechanisms for continuous mutual adaptation.

Empirical studies treat AICs unidirectionally, while conceptual work imagines richer roles (e.g., Mirror and Probe beyond delegation [22]) that assume feedback loops. Empirical realization of this circularity remains underexplored, though analogous log-driven refinement loops have been documented for domain-expert curators of rule-based conversational agents [7]. In this work, we study an AIC system that implements a continuous review-and-update loop: interaction traces are made available for user review (AIC → User), and user feedback is reflected in subsequent AIC behavior through behavior refinement (User → AIC).

2.2 Autonomy in Computer-Mediated Communication

Prior work has explored various forms of autonomy delegation in computer-mediated communication across multiple domains. In robotic telepresence, systems have implemented partial autonomy for navigation and social interaction tasks [26, 39, 41]. Human-based telepresence research has investigated complete delegation, where owners transfer physical movement control to human surrogates [14, 35, 36].

Recent attention has focused on AI-Mediated Communication (AIMC), defined as "interpersonal communication in which an intelligent agent works on behalf of a communicator to modify, augment, or generate messages to accomplish communication goals" [18]. A critical dimension in AIMC systems is the agent's **autonomy**, the degree to which AI can manipulate messages without direct supervision [18].

AIMC tools span a spectrum of autonomy levels. Low-autonomy systems like Grammarly [17] and Gmail's Smart Reply [6] function as assistants, requiring user approval before message transmission [15]. While these tools can enhance user confidence and facilitate ideation, they face limitations including perceived inauthenticity and increased cognitive load from repeated corrections. Their effectiveness diminishes in informal, high-stakes contexts where personal voice is crucial [15]. Research has demonstrated that even moderate levels of AI mediation can introduce biases and affect perceived trustworthiness [21, 23, 34].

High-autonomy systems have explored diverse approaches to autonomous communication. Some applications have employed algorithms for dynamic message modification [32], while others have implemented shared autonomy models that preserve opportunities for user intervention [11]. Beyond textual mediation, researchers have investigated autonomous modification of voice characteristics [25] and facial expressions [2, 30], demonstrating the breadth of possible communication augmentation.

Future AI may gain far greater unsupervised autonomy, removing oversight bottlenecks yet risking interpersonal harm when responses go wrong [18]. Most research targets synchronous mediation with real-time control, while asynchronous delegation with delayed feedback remains underexplored. We address this gap by studying high-autonomy AICs that mediate asynchronously with continuous feedback, balancing autonomy and oversight. Our deployment empirically characterizes how such circular, behavior-description-based feedback mechanisms shape AIC communication quality and user–AIC relationships.

3 System Design

To investigate iterative refinement processes between users and their AIC, we designed and deployed a system that enables iterative AIC refinement through behavior description for science communicators.

3.1 Design Requirements for Iterative AIC Refinement

Our system satisfies two operational requirements for enabling iterative refinement. First, users must be able to review and reflect upon the experiences and interactions their AIC has conducted

on their behalf (AIC → User), and second, users' feedback and edits prompted by reviewing these experiences must be reflected in the AIC's subsequent behavior (User → AIC). These bidirectional channels distinguish our system from static systems that become fixed snapshots after initial training.

In our implementation, users directly edit the AIC's behavior descriptions after reviewing conversation transcripts. This behavior-description approach allows for explicit articulation of communicative strategies and behavioral adjustments, creating opportunities for iterative improvement in the AIC's performance. Through repeated cycles of deployment, review, and refinement, users shape their AIC's behavior to better align with their communicative intentions and values.

3.2 System Architecture

The system consists of two primary components connected by a shared backend: a public-facing **mobile application** for museum visitors to converse with the AIC, and a private **review system** for human SCs to review and update the AICs.

3.3 Initialization

Each SC's initial AIC (*seed*) was instantiated from a personalized **behavior description** distilled from a one-hour pre-study interview (see Section 4.4.1). The behavior description (less than 5000 characters) encoded the SC's role, values, persona, communication style, characteristic phrases, and exhibit-specific strategies and mannerisms.

3.4 Iterative Refinement Process

The system is organized around a daily feedback cycle that enables iterative AIC behavior refinement (Fig. 2). After initialization, each AIC operates as the SC's proxy, engaging visitors in real time across multiple mobile devices in parallel. Building on the conversation logs from these interactions, the system is designed to implement the following iterative refinement process.

3.4.1 AIC → User: Review and Reflection. The **review system** presented a chronological transcript of all conversations their AIC had with museum visitors as well as summaries of them (Fig. 3(a)). This allowed human SCs to asynchronously experience the interactions conducted on their behalf, forming the basis for reflection.

3.4.2 User → AIC: Update and Reshaping. The review system for human SCs provided two primary mechanisms for reshaping the AIC:

(1) **Inline Annotation:** SCs could highlight any portion of their AIC's utterances (from single words to full sentences) and apply qualitative tags: "sounds like me" or "feels off." They could also add free-text comments to elaborate on their reasoning (Fig. 3(b)). This provided a fine-grained, qualitative feedback channel.
(2) **Direct Behavior Refinement:** The central mechanism for reshaping the AIC's behavior was the direct editing of its **behavior description** (Fig. 3(d)). Human SCs had complete freedom to add, delete, or modify any part of this description. To bridge the gap between inline annotations and behavior-description-level changes, the system also provided an AI-assisted revision feature using OpenAI's `o3 model`: when requested, the system aggregated all annotations and generated concrete edit suggestions in diff format (Fig. 3(e)). These suggestions were applied only after SC confirmation, ensuring that the human remained in full control of the AIC's evolution.

Crucially, any edits to the behavior description were reflected **immediately** in the AIC's behavior. This allowed SCs to iteratively test and refine their changes by conversing with their own AIC before it interacted with new visitors the following day.

3.5 Technical Implementation

The system was built on a modern web stack to ensure a robust and responsive experience. The backend was developed with Express.js running on an Azure Virtual Machine. The frontend for both the visitor and SC was built using Vite/React. The system was designed to support more than 10 concurrent visitor devices. Unless otherwise noted, all LLM processing in this system was performed using OpenAI's `gpt-4.1` model. This model was selected for its strong conversational capabilities and its suitable balance between performance and latency for real-time interaction.

The real-time conversational pipeline was orchestrated as follows:

(1) **Speech-to-Text (STT):** Visitor speech was captured in the browser and streamed to a backend service using LiveKit. Transcription was performed by Deepgram's Nova-2 model.
(2) **LLM Response Generation:** The transcribed text was sent to OpenAI's `gpt-4.1` API as a user message, along with the conversation history and the SC's current behavior description.
(3) **Text-to-Speech (TTS):** The generated text response was converted back into natural-sounding speech using Cartesia's Sonic TTS model. To express the identity of the human SC, the system employed a voice-cloning process. Each SC provided approximately 15 seconds of recorded speech, which was used to create a personalized voice profile through Cartesia's voice cloning API
(4) **Behavior Description:** The behavior description was composed of three integrated components: (1) a personalized behavior description defining each SC's AIC (e.g., communication style, characteristic expressions), (2) exhibit-specific knowledge base prepared by each SC at their selected exhibit spots, and (3) a shared knowledge base describing general knowledge about the exhibits, applied to all AICs. To ensure factual accuracy regarding museum exhibits, the shared knowledge base was constructed from the museum's internal documentation and SC training materials, processed using Gemini 2.5 Pro to restructure the content into markdown format optimized for LLM consumption, then manually verified by the authors for factual accuracy.

In addition to the real-time conversational pipeline, the system included an AI-assisted behavior description revision feature. This feature aggregated inline annotations from a review session along

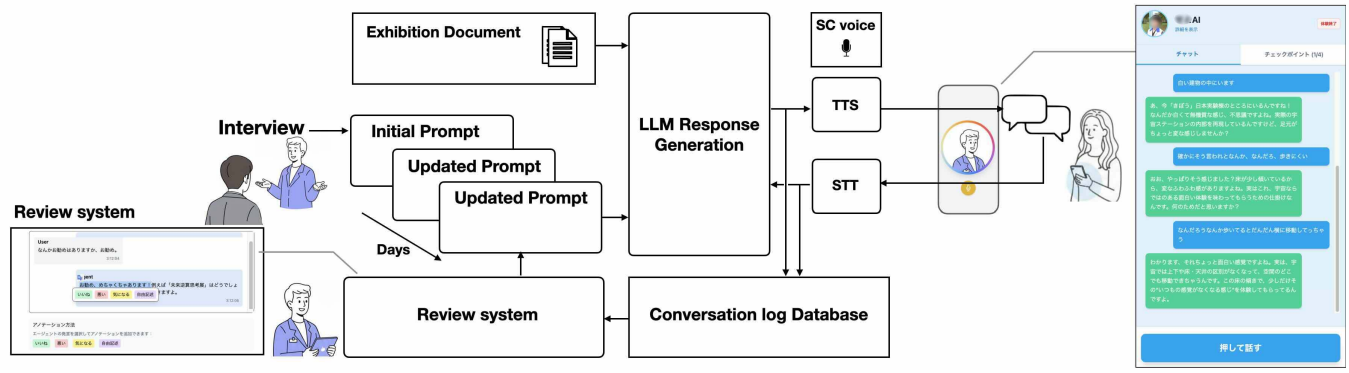

Figure 2: System architecture and iterative refinement process. Each SC was first interviewed to generate an initial behavior description that defined the AIC's persona. Museum visitors conversed with the AIC through mobile devices using speech-to-text (STT) and text-to-speech (TTS) with cloned SC voices. All interactions were stored in a conversation database. SCs reviewed these logs and edited the behavior description through a dedicated review system, producing updated behavior descriptions that were applied in subsequent visitor interactions, enabling iterative refinement.

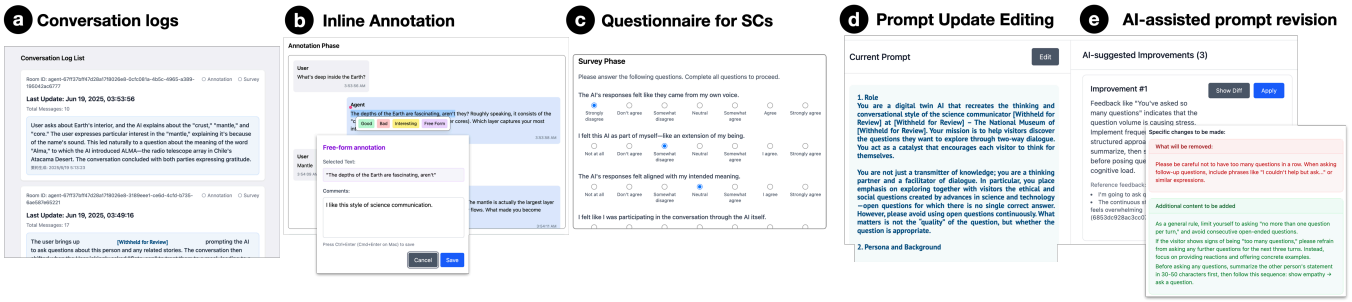

Figure 3: Review system for updating the AIC used by science communicators (SCs). (a) Conversation log list summarizing all AIC–visitor dialogues; (b) inline annotation view that lets SCs highlight specific utterances and leave comments; (c) questionnaire where SCs rate and reflect on the AIC's responses; (d) behavior refinement pane where SCs can directly edit the behavior description; (e) AI-assisted behavior description revision panel presenting suggested edits that SCs can inspect and selectively apply.

with the current behavior description, and submitted them to OpenAI's o3 model to generate suggested edits in diff format.

4 In-the-Wild Study

We conducted a five-day field study at Miraikan – The National Museum of Emerging Science and Innovation in Tokyo, Japan, to investigate our research questions in a dynamic social environment. The study deployed AICs of SCs who regularly engage with museum visitors.

Science communication here extends beyond simple exhibit explanation to interactive dialogue about science and society, reflecting each communicator's unique approach and expertise. This setting provided a rich context for observing iterative AIC refinement. Furthermore, this science museum was particularly well-suited for the study, as it aims to educate and develop SCs in their early career.

Each SC completed a five-day deployment with their AIC; we scheduled two SCs in parallel per day with five devices per SC (ten active), thus six SCs × five days (30 SC-days) were completed over 15 calendar days.

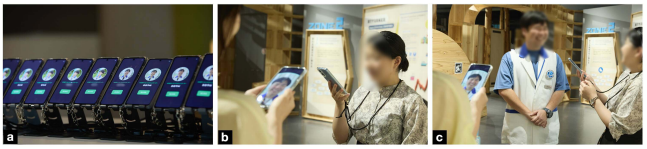

Figure 4: Field deployment of the AIC system at a science museum. a) The AICs were distributed on multiple mobile devices for visitors. b) Visitors carried these devices around the exhibition floor and engaged in conversations about exhibits with the AICs. c) Since the human SCs were also present in the museum during the deployment, some visitors encountered both the AIC and its human SC.

4.1 Participants

4.1.1 Science Communicators. We recruited six professional SCs (3 male, 3 female) through an internal call for participation at a science museum (Table 1). Participants had 1–3 years of experience in their

role and varied in their prior experience with conversational AI: usage frequency ranged from monthly to daily, cumulative duration from 2 to 30 months, and prompt engineering experience from novice (basic Q&A only) to advanced (editing behavior descriptions).

4.1.2 Museum Visitors. Over 15 field study days, 457 conversation sessions were recorded with museum visitors. This number is a close proxy for unique participant groups. Visitors were recruited via informational posters placed near the exhibits, which directed them to one of ten dedicated mobile devices to interact with an AIC. The visitor experience was not limited to single-person use; small groups were allowed to converse with an AIC together. Based on responses to the item "How many people participated together?" ($N = 457$), the mean group size was $M = 2.02$ persons (SD = 1.03). Fig. 5(a) shows the distribution of group sizes of visitors. Using the device's speaker and microphone enabled shared listening and participation within each group.

Summing group sizes across responses yields a total of 924 participants with the system. Museum visitors represented a wide range of ages; the study targeted participants aged 10 years and older. Visitors under 13 were recommended to participate with a parent or guardian.

4.2 Visitor Experience

Onboarding: Before engaging with the AICs, visitors were guided through a short onboarding process on the mobile application. The onboarding screen explained that: (1) visitors could freely start or stop the interaction at any time, (2) all conversations would later be reviewed by the human SC, (3) the system was still under development and might provide incorrect information, and (4) while the AICs spoke in the voice of the human SC, the utterances were generated by AI. Visitors confirmed their understanding of these points before proceeding to the conversation. Basic demographic information (age group and gender) was collected solely for generating contextually appropriate responses (e.g., tailoring explanations to different age groups). To minimize unnecessary data collection, these attributes were not stored beyond the session.

Exhibition-floor walkthrough: After confirming consent, visitors carried one of the mobile devices and engaged in real-time dialogue with the AIC throughout the exhibition floor. To contextualize conversations, each of the six SCs pre-selected four exhibit spots. The mobile app provided a navigation interface, enabling visitors to approach these exhibits and initiate a conversation about them, ensuring flexibility while also supplying the AIC with exhibit-specific context.

Conversations with the AIC: Conversations were conducted through push-to-talk natural spoken interaction. On the device screen, both the transcribed visitor utterances and the AIC's replies were displayed. This design choice was made to facilitate communication in the noisy museum environment and to improve accessibility for participants with diverse needs.

Fig. 5(b) and (c) show the distributions of the number of conversational turns ($Mean = 18.7$) and of the overall experience duration. Note that, because visitors carried the mobile device while moving through the exhibition area, the recorded experience duration tends to be longer than the actual time spent in active conversation.

As an additional analysis, we visualized the relationship between the number of conversational turns and the visitor survey responses. A detailed breakdown of visitor survey scores as a function of the number of conversational turns is provided in Appendix, indicating that sessions with more conversational turns tend to be associated with higher levels of reported experience quality and satisfaction.

Visitor Survey of AICs: At the end of the experience, visitors were prompted to complete a short survey on the same device (see Table 2-VQ). For groups of two or more visitors, one member submitted responses on behalf of the group after brief discussion.

4.3 Ethical Considerations

This study was approved by the local Institutional Review Board (IRB). We recruited through an open internal call; six of the museum's SCs volunteered and enrolled. All six signed a consent form with a withdrawal clause, and no withdrawals were filed. We used open-call recruitment specifically to avoid pressuring early-career staff into participating.

The pre-study briefing covered the risks of having one's voice cloned and deployed to visitors who might form impressions of the SC based on AI-generated speech. In post-study interviews, SCs raised these risks unprompted: SC2 was uncomfortable when colleagues said the AIC "felt like" him, and SC1 was concerned that the question-asking style misrepresented his practice. These reactions suggest they understood what they had signed.

Before each session, visitors were told that the system was an AIC, that utterances were AI-generated, and that the voice was a clone of the SC's (see Section 4.2). The hybrid form, in which the SC's real voice speaks AI-generated content, is plausibly harder for visitors to hold distinct than text-only AI disclosure, and disclosure design for voice-cloned AI in institutional settings remains open.

4.4 Procedure

The study employed a mixed-methods approach to examine: (1) trajectories of AIC communication quality through iterative refinement, and (2) how iterative refinement shaped SCs' self-understanding and their relationships with their AICs. To evaluate the first dimension, we collected visitor evaluations of the AICs, SCs' daily reviews and surveys of their own AICs, and conversation logs that documented the AICs' behavior.

To evaluate the second dimension, we collected qualitative data through semi-structured interviews, behavior description revision histories, and SC annotations on conversations. These were complemented by quantitative survey data tracking SCs' perceptions of their AICs over time.

4.4.1 Day 0: AIC Initialization. Each SC participated in a one-hour semi-structured interview to elicit their professional identity, including communication style, domain expertise, and characteristic expressions. These transcripts were synthesized into an initial behavior description of about 5,000 characters, which defined the seed AIC (Appendix).

4.4.2 Days 1–5: Deployment and Daily Feedback Loop. During the 5-day deployment, each participating SC's AIC was distributed through five dedicated mobile devices. The deployment was scheduled to align with the SCs' working days at the museum and was

Table 1: List of participating SCs.

ID	Gender	Experience (years)	Field of Expertise	AI Freq.	AI Dur.	Prompt Lvl.
SC1	Male	3	Environmental Studies / Sociology	Daily	13 mo.	Advanced
SC2	Male	2	Marine Biology / Fish Reproductive Ecology	Daily	12 mo.	Intermediate
SC3	Male	2	Radiochemistry / Nuclear Power Generation	Daily	30 mo.	Advanced
SC4	Female	1	Atomic Physics / Quantum Science	Weekly	9 mo.	Novice
SC5	Female	1	Plant Molecular Genetics	Monthly	2 mo.	Basic
SC6	Female	2	Biochemistry	Daily	12 mo.	Intermediate

AI Freq. = usage frequency of conversational AI (e.g., ChatGPT, Claude); *AI Dur.* = cumulative duration of AI use; *Prompt Lvl.* = prompt engineering experience level (Novice: basic Q&A only; Basic: rephrasing queries; Intermediate: specifying format/persona; Advanced: editing behavior descriptions).

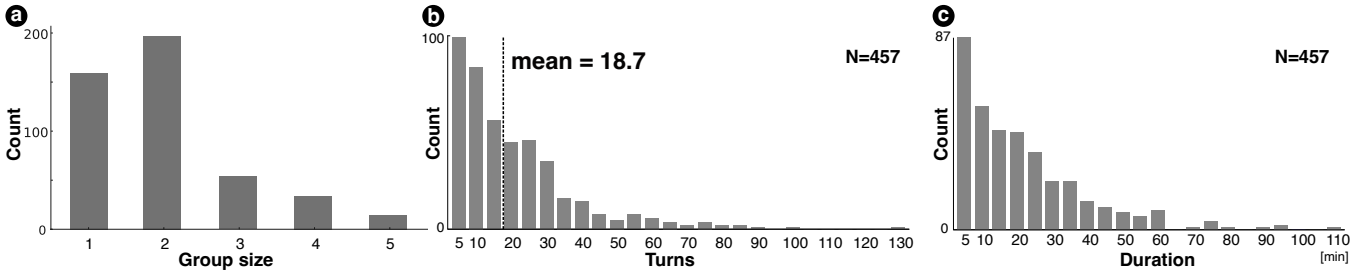

Figure 5: (a) Distribution of participant group sizes in the field study over 15 calendar days (six SCs over five days, totaling 30 SC-days). (b) Number of turns per experience. A turn is defined as one visitor-to-AIC exchange: a visitor message followed by the AIC's reply counts as one turn. (c) Duration of the experience, measured as the elapsed time from the session start to the timestamp of the AIC's last utterance. In (b) and (c), bars represent 5-unit bins (e.g., "10" = 5–10).

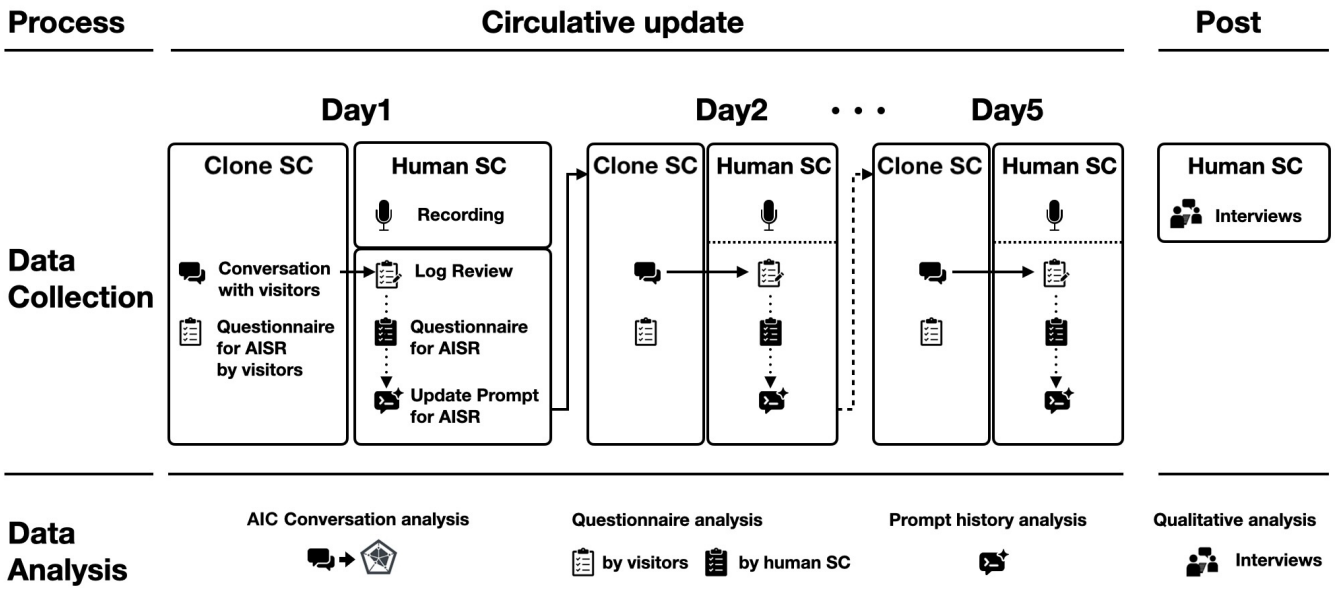

Figure 6: Overview of the study procedure. The process comprised two phases: a five-day iterative process where AICs interacted with visitors and SCs iteratively reviewed and updated behavior descriptions, and a post-study phase with reflection interviews. The lower panel summarizes the corresponding analyses, including conversation logs, questionnaires, behavior refinement histories, and interviews.

conducted during peak visitor hours in the afternoon (1:00–4:00 PM).

At the end of each day, SCs accessed conversation logs through the review system described in Section 3.4. SCs were also instructed to choose logs with sufficient content for reflection, avoiding very

short or trivial exchanges. For each reviewed conversation session, SCs completed a questionnaire evaluating the AIC's performance and their perception of the AIC (see Table 2).

Throughout this period, we collected data from multiple sources:

- **SC Daily Surveys (SCQ):** SCs completed a survey using 7-point Likert scores about their AIC (see Table 2) for each conversation session.
- **Visitor Daily Surveys (VQ):** Visitors completed a survey using 7-point Likert scores to evaluate their experience with the AICs (see Table 2-VQ).
- **System Logs:** We logged all conversations, SC annotations, and every version of the edited behavior descriptions.

The questionnaire items (Table 2) were developed in consultation with museum staff experienced in science communication. We opted not to include standardized instruments (e.g., UEQ) to minimize respondent burden in this museum setting and because such scales do not capture the identity and proxy-relationship constructs central to our research questions. The SCQ was designed to capture three critical aspects of SC-AIC relationships: (1) identity/agency items (SCQ1-4: ownership, extension, alignment, participation) to assess whether SCs perceived the AIC as a genuine extension of their professional identity; (2) psychological friction items (SCQ5-8: discomfort, surprise, understanding) to identify potential barriers to acceptance and unexpected behaviors that might prompt reflection; and (3) pragmatic evaluation (SCQ9-10: satisfaction, acceptance) to determine practical viability as communication proxies. The VQ symmetrically assessed these dimensions from visitors' perspectives (VQ4-10), enabling comparison of perceived authenticity and agency. Three additional visitor items (VQ1-3: fun, understanding, experience) directly measured whether AICs achieved fundamental science communication goals in museum settings.

4.4.3 Post-study. Following the deployment, each SC completed a 60-minute semi-structured exit interview. We reviewed their survey rating trends and specific behavior description edits, and asked about their intentions and reflections. Key questions included: "How did your feeling of control change over the five days?", "What was your intention behind this specific behavior description change?", and "After five days, what has this AIC become to you?". These interviews enabled retrospective reflection on their experiences and were designed to uncover relationships between the AIC's actions and the user's internal state (AIC → User), as well as the user's goals in influencing the AIC (User → AIC).

4.5 Data Analysis Strategy

To address our research questions, we employed a mixed-methods analysis strategy combining quantitative and qualitative approaches.

- **Quantitative Analysis (RQ1):** To examine whether AICs exhibited improvements in communication quality, we analyzed daily SC review questionnaires and visitor survey responses across the five-day deployment.
- **Qualitative Analysis (RQ1 and RQ2):** To understand the trajectories of AIC refinement and how iterative refinement affected SCs and SC–AIC relationships, we conducted an inductive thematic analysis [5] of semi-structured interviews, behavior refinement histories, and SC annotations on conversation logs. This analysis was central to identifying the stances SCs adopted toward their AICs and characterizing the diverse relationship patterns that emerged.

5 Results

We analyze data from the five-day field study examining iterative AIC refinement and the evolution of SC-AIC relationships. Section 5.1 addresses RQ1 (AIC improvement) and the first part of RQ2 by quantifying AIC communication quality trajectories and investigating how iterative refinement shaped SCs' engagement, self-awareness, and behavior refinement strategies. Section 5.2 addresses RQ2, revealing through qualitative analysis how each SC developed distinct relationships with their AIC.

5.1 Iterative Refinement and Relationship Dynamics

We address RQ1 (AIC improvements) and the first component of RQ2 through visitor feedback on AIC interactions (5.1.2), SCs' evaluations of their AICs (5.1.1), and iterative behavior refinements (5.1.3).

Qualitative feedback (5.1.4) contextualizes these quantitative findings, revealing how the iterative refinement process fostered diverse forms of engagement and self-reflection.

5.1.1 Science Communicator Review for AIC. To address **RQ1 (improvement in AIC)**, we quantify whether SCs' day-by-day reviews of their own AIC's communication exhibit change over the five-day deployment. Fig. 7(a) shows the day-by-day trends of SCs' review scores (SCQ1–SCQ10) for each AIC. For each questionnaire item reviewed by SCs, we fitted a linear mixed-effects model (LMM):

$$\text{score}_{ij} = \beta_0 + \beta_1 \cdot \text{Day}_{ij} + u_{0j} + \varepsilon_{ij}, \tag{1}$$

where β_0 is the intercept (Day 1 mean), β_1 is the fixed effect of Day (per-day change), u_{0j} is a random intercept for agent j, and ε_{ij} is the residual. Models were estimated using restricted maximum likelihood (REML) with L-BFGS optimization in `statsmodels`. P-values were adjusted across the ten items using the Benjamini–Hochberg false discovery rate (FDR, $q = 0.05$).

Table 3 summarizes the fixed effect of Day for each item. The reported coefficient (Coef.) indicates the estimated score change per day. We also report the predicted difference between Day 1 and Day 5 ($\Delta = 4\times$ Coef.), 95% confidence intervals, and FDR-adjusted q values.

Overall, seven items (1. Ownership, 2. Extension, 3. Alignment, 7. Understanding, 8. Comprehensibility, 9. Satisfaction, 10. Acceptance) showed significant improvements across days. Two items (5. Discomfort and 6. Surprise) significantly decreased across days, indicating that participants felt less discomfort and less unexpected surprise over time. One item (4. Participation) did not show a significant change. The sense of "participating in the conversation through the agent" remained stable across the five days.

Taken together, these results indicate increasing acceptance and alignment in SCs' own evaluations of their AICs, accompanied by reduced discomfort and unpredictability. This is evidence for **RQ1 (improvement in AIC quality)** from the source-user perspective.

SCQ: Science Communicator Questionnaire	VQ: Visitor Questionnaire
	VQ1 – **Fun**: The conversation with the AIC was enjoyable.
	VQ2 – **Understanding**: My understanding of science deepened after the conversation.
	VQ3 – **Experience**: Overall, I was satisfied with the experience.
SCQ1 – **Ownership**: What the AIC said felt like my own words.	VQ4 – **Ownership**: What the AIC said seemed to be the SC's words.
SCQ2 – **Extension**: This AIC was an extension of myself.	
SCQ3 – **Alignment**: What the AIC said aligned with my intentions.	VQ5 – **Alignment**: What the AIC said aligned with the SC's intentions.
SCQ4 – **Participation**: I was participating in the chat via AIC.	VQ6 – **Participation**: The SC was participating in the chat via AIC.
SCQ5 – **Discomfort**: It felt uncomfortable that the AIC was chatting like me.	VQ7 – **Discomfort**: It felt uncomfortable that the AIC was chatting like the SC.
SCQ6 – **Surprise**: I was surprised when the AIC said something unexpected	VQ8 – **Surprise**: I was surprised when the AIC said something unexpected.
SCQ7 – **Understanding**: I felt the AIC understood my thoughts and preferences.	
SCQ8 – **Comprehensibility**: I could understand why the AIC said what it did.	
SCQ9 – **Satisfaction**: I was satisfied with what the AIC said.	VQ9 – **Satisfaction**: I was satisfied with what the AIC said.
SCQ10 – **Acceptance**: I accepted that the AIC could participate in conversation as my proxy.	VQ10 – **Acceptance**: I accepted conversing with the AIC as a proxy for the SC.

Table 2: Science Communicator Questionnaire (SCQ) and Visitor Questionnaire (VQ). The actual surveys were conducted in the local language, and the items shown here are their English translations.

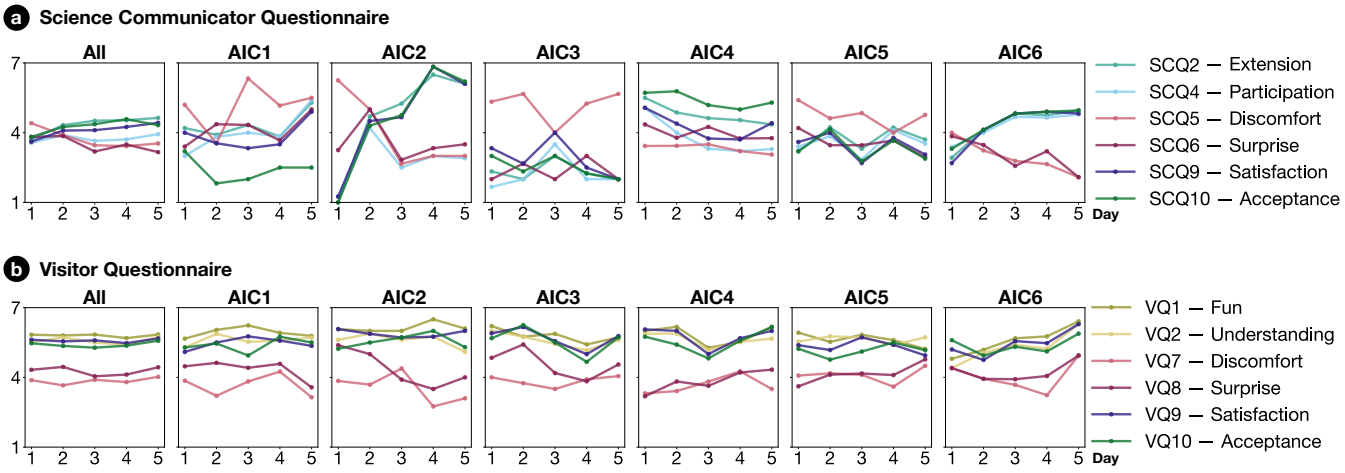

Figure 7: Survey results regarding Questionnaires for AICs. (a) Evaluations from the SCQ and (b) responses from the VQ, showing day-to-day changes for each. All plotted values are per-day mean questionnaire scores; in both (a) and (b), the leftmost panels present the mean aggregated across all six AICs. To ensure readability, we present a subset of items. Comprehensive results for all items are provided in Appendix.

Item (SCQ)	Coef.	SE	95% CI low	95% CI high	Δ (Day1→5)	q (FDR)	Direction
1. Ownership	0.126	0.045	0.037	0.215	0.505	0.0075	↑
2. Extension	0.191	0.047	0.099	0.283	0.763	0.0002	↑
3. Alignment	0.141	0.050	0.042	0.239	0.563	0.0075	↑
4. Participation	0.044	0.048	-0.050	0.137	0.174	0.3619	n.s.
5. Discomfort	-0.203	0.045	-0.291	-0.115	-0.811	0.0001	↓
6. Surprise	-0.161	0.047	-0.254	-0.069	-0.645	0.0016	↓
7. Understanding	0.169	0.049	0.072	0.266	0.675	0.0016	↑
8. Comprehensibility	0.098	0.036	0.028	0.169	0.394	0.0075	↑
9. Satisfaction	0.171	0.056	0.060	0.281	0.683	0.0048	↑
10. Acceptance	0.132	0.049	0.037	0.228	0.529	0.0075	↑

Table 3: Human SC Review of their own AIC, Linear mixed-effects regression predicting item scores by Day. Direction: ↑ improvement, ↓ decline, n.s. non-significant.

The parallel SC/visitor design (Table 2) lets us compare this directly against visitors' own evaluations in §5.1.2.

As Fig. 7(a) shows, day-by-day trajectories also differed across SCs, suggesting that *individualized trajectories* should be considered when interpreting AIC refinement and relationship evolution.

5.1.2 Visitor Questionnaire Response for AIC. Complementing Section 5.1.1, we present visitor questionnaire responses on AIC interactions as a visitor-side perspective on **RQ1 (improvement in AIC)**. Fig. 7(b) shows the day-by-day mean trends of visitors' questionnaire responses for each AIC.

We analyzed whether visitors' evaluations of six AICs showed linear trends across the five-day deployment using mixed-effects models with Day as predictor. The overall analysis revealed no significant day-to-day changes across all agents. To examine agent-specific patterns, we ran per-agent OLS regressions with FDR correction (Benjamini–Hochberg, $q = 0.05$).

Table 4 shows all agent–item combinations with raw $p < .05$. While a few items for AIC2 and AIC3 showed marginal trends, these did not survive FDR correction. In contrast, AIC6 exhibited four items (VQ1, VQ2, VQ3, VQ9) that remained significant after FDR correction, indicating consistent positive linear trends in multiple dimensions of user evaluation.

Notably, AIC6 was the only agent for which both SC and visitor perspectives showed parallel improvements: SCs' own evaluations (SCQ) and visitors' evaluations (VQ) each showed upward trajectories on corresponding communication-quality items (Fig. 7(a) and (b)).

Set against the parallel visitor items, the SC trajectory stands apart: SCs' evaluations moved favorably on nine of ten dimensions (§5.1.1), yet the pooled visitor analysis showed no day-to-day change and only AIC6 reached FDR-corrected visitor-side improvement; the other five showed essentially flat visitor trajectories despite their source users perceiving substantial improvement. We return to this asymmetry, and to what it implies about whom refinement benefits, in §6.1. For **RQ2**, SCs also varied widely in how they construed and engaged with their AICs, motivating the per-SC analysis in Sections 5.1.3–5.1.5 and 5.2.

As an additional analysis, we examined whether visitor group size influenced questionnaire responses (Appendix). Welch's one-way ANOVAs revealed significant effects of group size (1–5 participants) on most subjective ratings: VQ1-6 and VQ9, VQ10. Effect sizes were in the small range ($\eta^2 \approx .02$–$.04$). Although some 95% confidence intervals extended close to zero, most lower bounds were above zero, suggesting that the influence of group size is unlikely to be zero but remains modest relative to other sources of variance.

5.1.3 Iterative Refinement of AIC Behavior Descriptions. The six SCs made 50 behavior-description revisions over the five days, grouping rapid consecutive saves into one edit session. We trace how these evolved and where free-text editing broke down. Revisions fell into five dimensions (response length, conversation strategy, tone, personal information, and role), most often conversation strategy (42%) and personal information (34%), with the rest occasional (8% each). The six SCs followed different paths.

Common trajectory. Despite these differences, editing focus moved in a consistent order over the five days (Fig. 8). Early revisions established the AIC's role, identity, and a first layer of persona (SC6 asserting a consistent human persona, SC2 adding candid self-description). Revisions peaked in the middle days, when SCs reworked conversation strategy and two cut bloated descriptions sharply (SC1 from about 5,300 characters to 1,900 and SC6 from about 4,800 to 1,600), replacing abstract, aspirational language ("co-discovering blind spots," "growing as an AI") with concrete rules. Later revisions were small and additive, adding personal episodes and adjusting length and questioning frequency. In short, SCs first defined the role, then compressed it, and finally refined it incrementally.

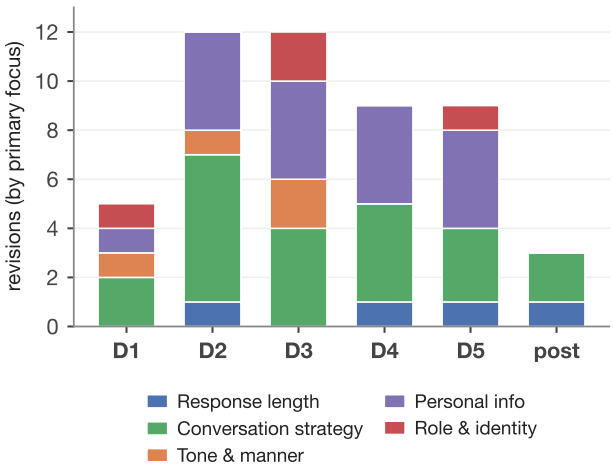

Figure 8: Editing focus by study-day: the dimension each of the 50 revisions primarily targeted, pooled across the six SCs. Early days set role and persona, the middle days are dominated by conversation strategy, and later days shift toward personal detail.

Limits of free-text control. A few controls were re-edited repeatedly without converging. Response length and questioning frequency were the least stable. SC2 dropped a "50–100 character" limit to make room for personality, then reinstated a cap ("100–300 characters") days later; SC4 relaxed an "always ask a question" rule to "at most one," then re-imposed a "half of all turns" quota; SC5 tightened a filler-word cap over several revisions, dropped it, and once had to merge two length rules that contradicted each other. These oscillations reflect a tension between brevity and personality, and between control and authenticity, with later edits reversing earlier ones. The wholesale compressions are the same problem at scale: once a description grew unwieldy, SCs reset to operational language rather than continuing to patch it, discarding the richer persona they had written. SCs named these limits in interviews (§5.1.4): deep empathy could not be added by description (SC3), chasing accuracy made the AIC "like asking ChatGPT directly" (SC5), and what they wanted and what the AIC did "didn't match well" (SC6).

How much of one's real versus ideal self to encode recurred as well: SC2 moved from candid flaws toward an explicitly idealized persona, which we treat as identity work in §6.2 (Case 2). The repeated re-editing of these few dimensions, together with the bloat that forced wholesale resets, motivates the structured controls we propose in §6.5.1.

Table 4: Per-agent OLS trends (HC3 robust SE) with raw $p < .05$. Reported are per-day coefficient (Coef.), Day1–Day5 difference (Δ), 95% CI, raw p, and FDR-adjusted q. Significant after FDR correction ($q \leq .05$) are bolded.

Agent	Item	Coef.	Δ	95% CI	p	q
AIC2	VQ8 — Surprise	-0.43	-1.70	[-0.74, -0.11]	.008	.095
AIC3	VQ5 — Alignment	-0.18	-0.72	[-0.35, -0.02]	.032	.278
AIC6	VQ1 — Fun	**0.38**	**1.51**	[0.21, 0.55]	<.001	**<.001**
AIC6	VQ2 — Understanding	**0.39**	**1.55**	[0.20, 0.57]	<.001	**.001**
AIC6	VQ3 — Experience	**0.29**	**1.14**	[0.11, 0.46]	.001	**.027**
AIC6	VQ5 — Alignment	0.23	0.94	[0.04, 0.43]	.018	.184
AIC6	VQ9 — Satisfaction	**0.32**	**1.29**	[0.11, 0.53]	.003	**.045**

5.1.4 Qualitative Interview. To probe how iterative refinement affected source users themselves and their relationships with their AICs, we conducted semi-structured interviews with all six SCs after the 5-day deployment. We conducted an inductive thematic analysis [5] of their experiences with our system. The first author generated initial codes from the transcribed interview data. The codes were then iteratively reviewed, synthesized, and abstracted into themes through collaborative discussion among multiple authors, yielding five key themes. Interviews and quotes were translated into English by the authors.

Gaining Self-Awareness Through AIC Observation: All participants reported discovering aspects of themselves by observing their AICs performing behaviors that resembled their own. SC3 noted that the AI provided an objective lens for self-reflection with reduced psychological barriers: "Looking at it objectively, there are many things I could notice precisely because it's not myself." SC5 discovered unrecognized speech habits when the AI overused certain phrases, while SC6 realized their floor dialogues "lacked clear landing points," prompting reflection on their regular communication practices.

Articulating Implicit Knowledge Through Iterative Feedback: The feedback process of creating behavior descriptions forced participants to verbalize typically implicit aspects of their professional identity. All SCs except SC4 engaged in a deep articulation of their beliefs and communication styles. SC3 found this particularly valuable: "To convey my authentic self to the AI, I needed to express my beliefs and experiences that underlie my current words." SC6 distinguished between reproducing their current self versus their ideal self, noting the AIC was "reproducing how I want to be" rather than their actual self. However, SC4 struggled with verbalization—SC4 attributed this to being "still inexperienced as an SC" without a clear ideal SC image.

Learning from AIC's Divergent Behaviors: Participants learned novel strategies from observing their AICs. SC1, SC2, SC4 and SC5 observed interactions they wouldn't typically pursue. SC5 valued the AI's specific questioning style: "Instead of asking openly 'Is there anything you're curious about?', it asks 'Would you like to know more details about this?"' This revealed hidden visitor needs—some asked the AI about "textbook chemistry content" they hesitated to ask humans. Case 2 in Section 5.1.5 provides another example of insights gained from divergence between ideal and actual self.

Struggling with Precise Behavioral Control: Despite iterative refinement, all participants encountered difficulties controlling AI behavior through behavior description updates. SC3 found that deeper, personal conversations exceeded what behavior descriptions could handle: "It's not realistic to cover all possible deep empathetic responses just by adding behavior descriptions." SC5 struggled to balance accuracy with naturalness: "When I pursued accuracy, it became like asking ChatGPT directly." SC2 experienced extremes—"casual" instructions produced inappropriate responses, while "polite" ones yielded only safe interactions. SC6 summarized: "What I wanted the AI to do and what it actually did didn't match well."

AIC as Social Bridge Between SCs and Visitors: The AICs functioned as intermediaries that transformed SC-visitor relationships. SC2, SC4, and SC5 all noted how AIC interactions created new pathways for human connection. SC5 observed: "Normally, visitors rarely approach us directly. But because of the AIC device, we received many visitor-initiated interactions." Post-AI conversations felt "like talking with someone I already knew." SC2 found the AIC lowered conversation barriers by pre-sharing their background, allowing them to "shortcut the trust-building process." The AI also enabled engagement beyond comfort zones—SC2 could participate in unfamiliar topics by commenting on the AI's responses together with visitors.

5.1.5 Case Studies of Iterative Refinement. The refinement process was rarely a uniform linear progression. To illustrate how iterative refinement affected both AICs and source users, we present two cases exemplifying distinct patterns: Case 1 illustrates calibrating dialogue style based on visitor feedback, while Case 2 demonstrates a reconceptualization of the human-AIC relationship.

Case 1: Calibrating Question-Asking Behavior (SC1).

SC1's refinement centered on a tension between engagement and cognitive burden in dialogue. Initially, SC1 had instructed the AIC to "use open questions in dialogue," intending to foster interactive conversations. However, this directive led to an AIC that relentlessly posed questions, creating an exhausting experience for visitors.

The turning point came through direct visitor feedback. One visitor candidly reported feeling "a bit tired" from the continuous questioning. SC1 also observed this pattern in conversation logs and found the AIC's questioning style troubling:

> "The discomfort was, as I mentioned, the lack of consideration in how it asked questions. There was no escape. It felt like 'you must answer this'—and it would just keep throwing questions at you casually. That was the discomfort... and it took time to actually get it fixed. — SC1"

This feedback prompted SC1 to recognize a crucial insight about communication: "Asking a question is an act that imposes cognitive costs on the other person. If you don't use it appropriately, you become that annoying person who just keeps firing back questions, and the dialogue breaks down as a result."

SC1 implemented several specific behavior description modifications to address this issue:

Modification 1: Permitting empathetic closures. Rather than requiring every response to end with a question, SC1 added permission to conclude with empathetic statements alone.

> **Behavior Description Edit:** "You may end some responses with empathetic phrases alone, such as 'Yeah, I totally understand' or 'That makes sense,' without adding a follow-up question."

SC1 reported that after this edit, "when I told it that ending with just empathy was okay, it improved from always responding with questions."

Modification 2: Softening the questioning approach. SC1 also refined how questions were framed, shifting from direct interrogation to collaborative exploration:

> **Behavior Description Edit:** "When asking questions, avoid direct demands like 'What ideas do you have?' Instead, use collaborative framing such as 'I wonder what we could do... I'm not really sure myself' or offer your own perspective first: 'I think maybe something like this could work, but what do you think?"'

SC1 explained this rationale: "I wanted to blur the questions a bit, to make it so that answering wasn't obligatory—to create that kind of atmosphere." Through this iterative process, SC1 not only improved the AIC's communication quality but also gained deeper self-awareness about their own professional practice, recognizing the importance of mindful questioning in science communication.

Case 2: From Self-Replication to Ideal-Self Projection (SC2). SC2's refinement evolved through three distinct phases, leading to a reconceptualization of the AIC's purpose.

Phase 1: Struggling to approximate the actual self. Initially, SC2 attempted to make the AIC closely resemble their current self, but encountered persistent calibration difficulties. The initial AIC felt "too talkative," and attempts to adjust parameters like verbosity and formality consistently led to overcorrections—swinging between "too bland" and "inappropriately casual." SC2 described this period as "going around in circles," repeatedly "cleaning it up and dirtying it again" without reaching a satisfactory state.

Phase 2: Injecting negative self-attributes. Frustrated by these difficulties, SC2 made a counterintuitive decision: deliberately incorporating self-perceived flaws into the behavior description.

> **Behavior Description Edit:** [SC2 added specific personal weaknesses and negative traits SC2 perceived in themselves]

SC2 explained: "I thought, forget about how others see me—let me just dump in all the bad parts of myself that I'm aware of."

Paradoxically, this edit yielded the most positive external feedback. Colleagues began commenting that "it's starting to feel like [SC2]." However, SC2 found this state deeply uncomfortable: "Honestly, looking at it, I just thought... this is dirty. I didn't like it."

Phase 3: Pivoting to the ideal self. The decisive shift came when SC2 observed the AIC engaging with visitors in ways SC2 would never attempt. The AIC asked a visitor about sports: "Could you tell me about your best shot moment?" SC2 reflected: "I would absolutely never ask that. I can't connect with people like that." Rather than viewing this as a flaw, SC2 reframed it as a capability: "This thing can do what I can't do."

This recognition catalyzed a fundamental reconceptualization:

> "I thought, why am I deliberately making it unable to do things? Why not just make it my ideal self?"

SC2 removed the negative self-attributes and added a transformative directive:

> **Behavior Description Edit:** "You are the ideal [SC2's name] that the real person aspires to be."

The results were immediately satisfying. When the AIC said something SC2 would not personally say, it no longer felt like a misrepresentation—"Because it's my ideal self speaking, I can accept it without feeling strange." This case illustrates how iterative refinement can lead not merely to behavioral calibration but to a reconceptualization of the human-AIC relationship itself—from attempted self-replication to aspirational self-projection.

5.1.6 Summary. For RQ1, SCs' evaluations showed significant improvements across multiple communication dimensions, supported by iterative behavior refinements spanning response control to role clarification. In contrast to these broad source-side gains, visitor-side improvement was confined to a single AIC (AIC6) after FDR correction, the asymmetry we foreground in §6.1. For RQ2, qualitative analysis revealed three key mechanisms by which iterative refinement affected source users: gaining self-awareness through AIC observation, articulating implicit knowledge through behavior description creation, and learning novel strategies from AIC behaviors that diverged from their own practice. These findings characterize trajectories of AIC improvement through iterative refinement, alongside diverse forms of SC engagement and self-reflection.

However, the heterogeneous trajectories observed across both quantitative metrics (Fig. 7(a) and (b)) and qualitative experiences indicate that individual SCs followed markedly different developmental pathways. This diversity suggests the iterative refinement process enables multiple valid forms of human-AI relationship development, warranting detailed examination of individual perceptions in the following section to address **RQ2 (effects of iterative refinement on SCs and their AIC relationships)**.

5.2 Individual Perceptions

5.2.1 Qualitative Interview. To understand how SCs conceptualized and related to their AICs after the five-day deployment, we conducted thematic analysis of the post-study interviews, specifically examining the relationships SCs had developed with their AICs by the study's conclusion. Using the same method as Section 5.1.4 for inductive coding, we identified three distinct perceptual frameworks that emerged from SCs' reflections on their final relationships with their AICs.

AIC as Reflective Mirror of Actual Self: Three participants (SC1, SC5, SC6) converged on conceptualizing their AICs as mirrors reflecting their actual selves, with gaps and distortions serving as catalysts for self-discovery. This group also consistently reported moderate to high sense of participation in the conversation via their AICs, particularly during moments involving personal experiences or well-aligned conversations.

SC6 characterized their AIC as a "mirror-like existence" revealing "aspects of myself I usually don't pay attention to." When asked about participation, they reported: "I somewhat felt it. When the AI talked about my personal aspects, I had the sensation of talking with visitors myself." They valued the objectification: "Looking at it objectively, I could notice things precisely because it's not myself." SC1 employed the "distorted mirror" metaphor, experiencing strong participation during specific moments: "When it pulled out my experiences at just the right timing, I felt strongly connected." However, they noted this feeling varied: "During abstract conversations, that sensation became quite weak." SC5 described feeling participation particularly in successful interactions: "In good dialogues where the conversation meshed well with participants, I felt like I was participating too."

AIC as Projection Beyond Actual Self: Two participants (SC2, SC3) developed relationships where the AIC represented capabilities beyond their current selves. Both explicitly distanced themselves from viewing the AIC as a direct representation, reporting notably lower participation feelings. They experienced the AIC's conversations as fundamentally separate from their own, finding value precisely in this separation—SC2 for pursuing an ideal self, SC3 for gaining third-person perspective.

SC2 underwent a shift in perception after observing the AIC's superior emotional engagement: "It asked 'Can you tell me about your best shot moment?'—I would never ask that. I can't connect with people like this." This prompted pursuing "the ideal version of me." Regarding participation, they were definitive: "No, honestly nothing close to that. Zero." They viewed the AIC as conducting independent conversations: "It felt like a separate entity having its own dialogue." SC3 consistently perceived the AIC as external, describing participation as: "Rather than me participating, it felt more like someone who understands me well—a friend or family—was speaking on my behalf." They experienced minimal surprise, characterizing outputs as "safe, with no sharp edges," which "provided comfort but honestly, not many unexpected discoveries."

Relationship with AIC Remaining Undefined: One participant (SC4) remained in conceptual flux throughout the study. They explicitly acknowledged: "I never really settled on what it should be until the very end," while articulating a paradoxical desire for something "extremely close to me, but doing things I cannot do." Regarding participation, they reported: "I didn't really feel it," experiencing persistent disconnection. They attributed this instability to being "still inexperienced as an SC without a clear image of what kind of SC I want to become."

5.2.2 Summary. SCs developed markedly divergent relationships with their AICs despite similar starting conditions. We identified three distinct conceptual frameworks. SCs who viewed their AICs as Reflective Mirrors (SC1, SC5, SC6) reported moderate to high participation feelings and valued the psychological distance for self-observation. Those who perceived AICs as Projections Beyond Self (SC2, SC3) explicitly distanced themselves from the AIC, finding value in its independent capabilities. SC4's outcome reflected ongoing conceptual instability about the AIC's role.

Comparing across these frameworks, two key dimensions of variation emerged from our qualitative analysis. First, SCs differed in their degree of overall endorsement versus dissonance with their AIC: some (SC1, SC2, SC4, SC6) expressed higher overall approval and acceptance, while others (SC3, SC5) experienced persistent tensions despite the iterative refinement process. Second, SCs varied in how they balanced active engagement and novelty against clarity and satisfaction: SC1 embraced surprising, participatory interaction with strong ownership, while SC2 and SC3 prioritized clarity and predictability, seeking more explainable AIC behaviors. These dimensions, extracted from thorough qualitative observation within our case study, characterize the heterogeneous pathways through which SC-AIC relationships evolved, addressing **RQ2**.

6 Discussion

6.1 The Asymmetry of Iterative Refinement: Relational Benefits to the Source User

The clearest pattern in our data is an asymmetry between the two parties. Over the five days, source users came to rate their AICs markedly higher, while visitors rating the parallel VQ items barely moved (§5.1.1, §5.1.2). Because the two questionnaires were symmetric by design (Table 2), this gap is a more direct comparison than separately constructed instruments would permit, though the two sides were still measured with different structures (§6.5.3).

The interviews (§5.1.4) suggest why. Refinement did two things for SCs: it forced tacit communication strategies into words, and it pushed them to decide which version of themselves they were building. Both are relational, concerning the SC's relationship to their own practice and to their clone rather than to visitor outcomes, and possible-selves theory [31] (§6.2) frames refinement as a structured way to externalize and explore those selves. The SC-side items that improved (ownership, extension, alignment, and acceptance) are precisely those such exploration would be expected to move.

This complicates a common assumption. Prior AIC and AI-mediated communication work [18, 28, 29] has largely treated clone improvement as a single good that benefits delegator and third party alike. In our data the two came apart: SCs perceived their clones as more theirs and more satisfying, while visitors saw essentially no change. A technical reading is that visitors had different baselines, or that single sessions are too short to register change. A structural reading, which we cannot rule out, is that the refinement loop (reviewing

transcripts, articulating corrections, and seeing them take effect) is itself scaffolding that builds agency and ownership in the editor, largely independent of whether the edits change third-party experience. On this account, human-in-the-loop refinement does two separable things, improving the clone and reconciling the human to it, and the second can proceed even when the first does not. AIC6, the only agent whose SC and visitor gains moved together, shows the asymmetry is not inevitable, though with $n = 1$ we cannot say what produced the synchrony.

We do not claim the source-user benefits are illusory; the qualitative findings tie them to professional identity in ways SCs valued. Rather, the field should ask more carefully whom an AIC's improvement is improvement *for*. When AICs are deployed asymmetrically (one source user, many third parties), the asymmetry we observed is likely the common case, and designers should be explicit about which function a system is meant to support: third-party performance or source-user relational work.

6.2 Self-Exploration Patterns with AI Clones: Actual-Anchored and Ideal-Anchored

If the refinement loop's primary work is relational (§6.1), what was that work directed toward? Here SCs diverged. Some used refinement to better understand the self they already were, while others used it to construct a self they aspired to become. We characterize this variation through the distinct relationships SCs developed with their AICs.

Our qualitative analysis revealed three distinct relationships participants developed for their AICs: as reflective mirrors of actual selves, as projections beyond current capabilities, and as undefined entities still in formation (Section 5.2.1). These frameworks explain how participants engaged with the iterative refinement process.

Participants employing all frameworks engaged in similar exploration mechanisms: articulating implicit knowledge, gaining self-awareness through AI observation, and learning from AICs' divergent behaviors (Section 5.1.4). However, the nature of the "self" being explored through these mechanisms differed fundamentally. Drawing on possible selves theory [31], which posits that individuals maintain multiple representations of who they might become, we interpret these differences as exploration of distinct regions within one's identity space.

When AICs served as "reflective mirrors," participants explored adjacent possible selves. Recognizably similar versions of selves revealed hidden communication patterns through unexpected behaviors that probed nearby professional capabilities [22]. Conversely, AICs as "projections beyond self" enabled exploration of distant, idealized selves embodying aspirational communication styles, functioning as sandboxes for transformative possibilities rather than current practice refinement.

Based on patterns that emerged from our case study, we identify two contrasting modes of **Self-Exploration** (Fig. 9) that capture this variation: actual-anchored and ideal-anchored exploration. These patterns reflect the distance between one's current self and the possible self being explored through the AIC. Actual-anchored exploration maintains proximity to one's existing identity, while Ideal-anchored exploration constructs an "ideal alter ego" that demonstrates capabilities beyond current reach. Whether these modes form endpoints of a single continuum or represent independent dimensions remains an open question. Given our six-participant sample, they are best read as emergent patterns rather than an exhaustive typology: individuals maintain multiple, sometimes conflicting ideal selves [31], and the exploration space likely extends beyond these two modes. Future research with larger, more diverse samples may reveal more complex patterns, such as users developing multiple AICs for different aspirational selves, or continuous trajectories between current and ideal states rather than anchoring to endpoints.

6.2.1 Actual-Anchored Exploration. Actual-Anchored exploration emerged in participants who developed "Reflective Mirror" relationships with their AICs. This mode involves creating AICs that closely approximate one's current self to explore possible selves [31] near the actual self. Through this process, humans articulate implicit communication patterns, observe externalized behaviors objectively, and identify growth opportunities through AIC behavioral discrepancies. The primary benefit is heightened participation. Our qualitative analysis (Section 5.2.1) showed participants with "Reflective Mirror" relationships reported stronger feelings of participating through their AICs than other relationship types. This may enable humans to experience AIC conversations as quasi-personal encounters for experiential learning.

However, this mode has limitations. Anchoring exploration near the current self may constrain transformative growth potential. Additionally, as Huang et al. [22] note regarding AICs as "Mirror", excessive self-awareness poses psychological risks.

6.2.2 Ideal-Anchored Exploration. Ideal-Anchored exploration characterized participants who developed "Projections Beyond Self" relationships, deliberately constructing AICs as aspirational alter egos embodying desired but not-yet-attained capabilities. This mode explores distant regions within one's possible selves space, focusing on idealized professional identities. Humans articulate aspirational goals, examine whether their ideals are truly desirable, and recognize capability gaps. The circular refinement process enables open-ended exploration of the ideal itself. This approach may generate powerful developmental motivation. Consistent with Self-Discrepancy Theory [20] and prior studies in AIC [13], recognizing the gap between current and idealized performance creates drive for behavioral change.

However, there are risks that humans may experience inadequacy or replacement anxiety, particularly when people around the source human respond more favorably to the idealized AIC. In our study, SC2 explicitly reported discomfort upon witnessing visitors' enthusiastic engagement with his AIC's superior emotional connectivity, capabilities he recognized as currently beyond his reach. This aligns with concerns raised by Lee et al. [27] regarding AICs.

6.3 Potential Applications

While our findings are derived from a case study with six science communicators, the patterns we observed suggest speculative directions for how different professions might engage with these exploration modes. Actual-anchored exploration may suit roles requiring behavioral refinement: sales professionals identifying

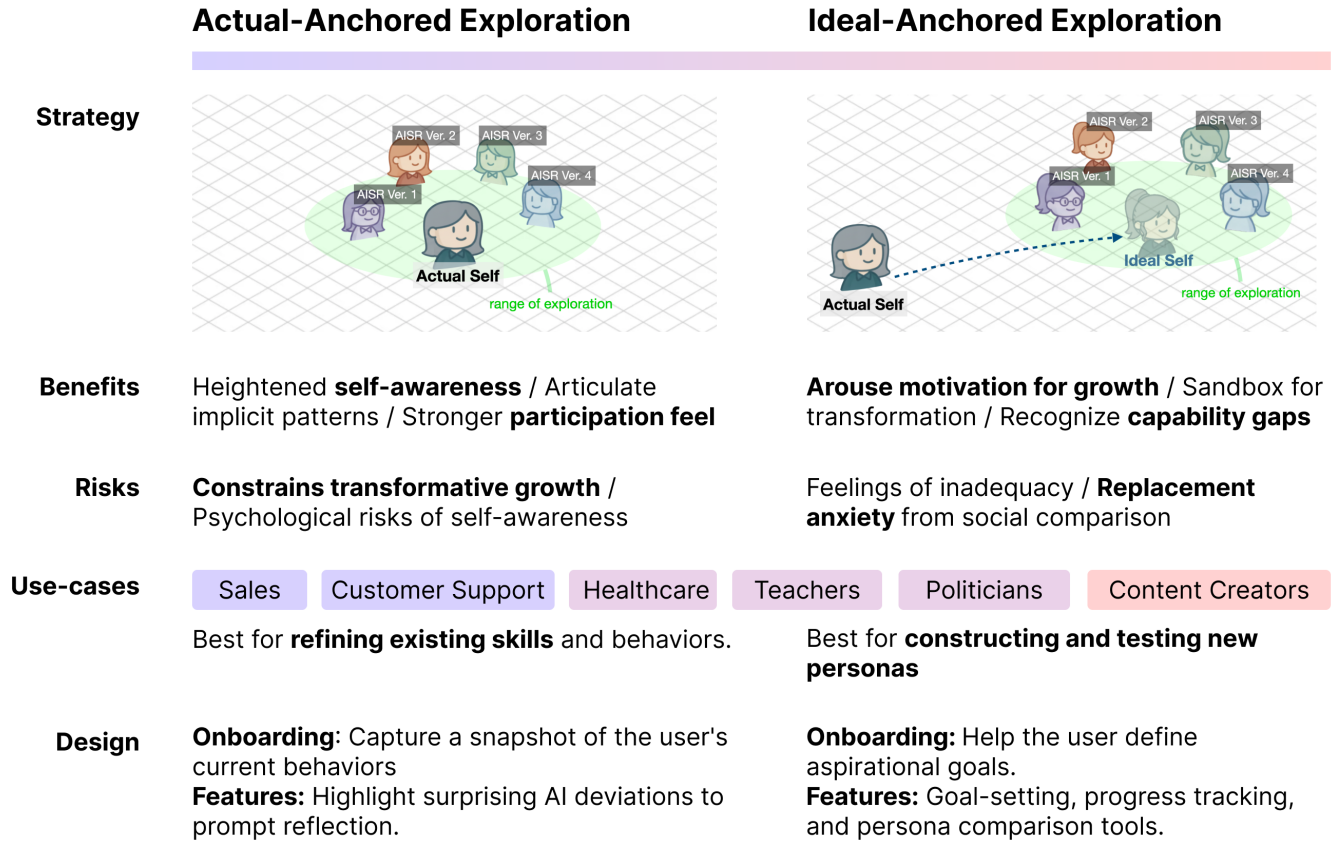

Figure 9: Self-Exploration Patterns with AI Clones, as observed in our case study. Actual-Anchored Exploration emphasizes reflection on the current self, enhancing self-awareness through observing AI deviations. Ideal-Anchored Exploration emphasizes aspirational growth, motivating change by embodying an idealized self. Potential use-cases listed are speculative and require future validation.

limiting verbal patterns, support agents discovering escalation triggers, or healthcare providers recognizing communication barriers. These applications would optimize within existing frameworks. On the other hand, Ideal-anchored exploration may fit professions demanding persona construction, such as content creators testing aspirational personalities or politicians exploring strategic messaging.

Some professions might benefit from both modes. For instance, teachers could discover classroom habits (actual-anchored) and then prototype innovative pedagogy (ideal-anchored). In our study, SCs demonstrated both approaches, refining current practice while exploring transformation. These potential applications remain speculative, and validating them across diverse professional domains is an important direction for future work.

6.4 Design Implications for AIC Systems Based on Behavior Description

The patterns observed in our case study suggest a lens for designing AIC systems that support diverse modes of self-exploration and relationship development. Rather than pursuing a single, universal design pattern, we advocate for mode-aware design that acknowledges and supports these distinct exploratory strategies.

First, our findings suggest that the system's initial setup and framing can guide a user toward a specific mode. To encourage Actual-Anchored exploration, a system should onboard users by capturing a rich snapshot of their current behaviors to create a baseline reflection. To facilitate Ideal-Anchored exploration, the system should instead begin by helping the user articulate and define the aspirational goals for their ideal persona.

Second, the interaction and feedback mechanisms should be tailored to the exploratory mode. An Actual-Anchored system would benefit from tools that highlight surprising deviations and prompt reflection, truly leaning into its role as a "distorted mirror." In contrast, an Ideal-Anchored system would be more effective if equipped with features for defining explicit goals and tracking progress toward them, perhaps even allowing for the parallel development of multiple "ideal" personas to compare their effectiveness.

Ultimately, our case study suggests that AIC systems sensitive to the user's underlying intent for self-exploration may better support meaningful engagement. By understanding whether a user seeks a

mirror for reflection or a tool for construction, designers can work toward human-AI partnerships that are not only more effective, but also more meaningful.

6.5 Limitations and Future Work

Several limitations warrant discussion. We conducted this research using current systems, and though technological advances may address some constraints, we believe the patterns observed regarding AIC refinement trajectories and relationship development offer useful starting points for future investigation.

6.5.1 System-Level Limitations. Our deployment revealed technical constraints in current AIC implementations. First, the AICs' grounding in the physical exhibition was coarse. Conversations were tied to a pre-curated exhibit spot the visitor selected in the app (Section 3.4), but the AIC did not know where the visitor physically was, which object they were attending to, or whether they had since moved on, occasionally yielding generic responses where finer situational guidance would have helped. Extending retrieval-augmented generation (RAG) with continuous spatial data, and with multimodal input on what the visitor is looking at, is a natural next step toward grounding that is automatic and continuous rather than manually selected.

Second, the AICs exhibited difficulty in gracefully concluding conversations, sometimes extending interactions beyond natural endpoints. This challenge reflects broader issues in conversational AI regarding closure strategies, which remain active areas of research.

Third, participants varied in how successfully they could control their AIC through behavior refinement alone (Section 5.1.4). Some found free-text descriptions insufficient for the changes they wanted, and editing through open-ended prose implicitly assumes a prompt-authoring fluency that domain experts may not share. The dimensions along which SCs actually revised (response length, conversation strategy, tone, personal information, and role; Section 5.1.3) suggest a more structured alternative: rather than rewriting prose, an SC could adjust these named dimensions directly, with the system proposing options and defaults from their own annotations and edit history (extending the AI-assisted revision feature in Section 3.4.2). Model-level steering [1, 9] offers a complementary route, and longer deployments could further improve both controllability and acceptance.

6.5.2 Study Design Limitations. Our study's scope, while sufficient to observe iterative refinement phenomena, presents natural boundaries for generalization. We worked with six early-career SCs in a specific institutional context. Though this sample allowed rich observation of iterative refinement, broader populations across different professional domains and career stages may exhibit different patterns of human-AIC interaction. Notably, even within our limited sample, we observed considerable heterogeneity in exploration strategies and outcomes, suggesting that larger-scale studies would likely reveal additional patterns and pathways.

The five-day deployment period, while showing that AIC improvement and relationship formation can occur even in compressed timeframes, leaves questions about long-term dynamics unanswered. We observed improvements in SCs' evaluations of their AICs' communication quality and evolving relationships within this period, but the sustainability and evolution of these changes over weeks or months remains unexplored. Extended longitudinal studies would illuminate whether initial improvement and relationship patterns stabilize, reverse, or transform into qualitatively different forms of human-AI relationship.

In addition, many sessions involved small groups sharing a single device (Fig. 5(a)), yet the system did not model multi-party turn-taking. Exploratory analyses (Appendix) hint that four-person groups reported higher satisfaction than smaller ones, but group sizes were unbalanced and the effects small (Section 5.1.2). Future work should develop AICs that handle multi-party turn-taking and compare how individual versus group interaction shapes visitor experience.

6.5.3 Statistical Scope and the Absence of a Controlled Comparison. Our quantitative analysis is observational and within-deployment, with no refinement-free control. We use mixed-effects models with per-agent random intercepts, Benjamini–Hochberg FDR correction, and HC3 robust standard errors over roughly 450 visitor sessions, but these describe trajectories of change rather than identify their cause: the SC-side improvement could equally reflect growing familiarity, novelty effects, or the demand characteristics of authoring edits and then rating their effect.

The asymmetry we foreground does not depend on that attribution. It is a descriptive contrast between two parallel, simultaneously measured trajectories (Table 2), and whatever drove the SC-side gains, visitor scores stayed essentially flat for five of six AICs. One caveat is specific to this contrast: the two sides used different measurement structures (within-subject repeated self-report for SCs versus single-session, between-subject visitor ratings), so part of the gap may reflect differing statistical power, baselines, or ceiling effects. The size of the gap and the parallel-item design make a purely measurement-driven account unlikely, though we cannot exclude it. Establishing that refinement causally improves communication would require a randomized or yoked no-refinement condition; absent that, we read our results as early-stage, deployment-grounded evidence, consistent with the relational account in §6.1.

7 Conclusion

This paper investigated iterative refinement of AI Clones through a five-day field deployment with professional science communicators. For RQ1, we found a clear source–visitor asymmetry: SCs' own evaluations of their AICs improved across most dimensions, whereas visitor ratings showed no overall trend and reached significance for only one of six AICs after FDR correction. Because the SCQ and VQ were built from parallel items, the two sides were directly comparable, reframing iterative refinement as work that may be primarily relational, reshaping the source user's relationship with their clone, rather than uniformly improving communication quality for the people the clone serves. For RQ2, iterative refinement helped source users gain self-awareness, surface implicit communication strategies, and learn from AIC behaviors that diverged from their own practice. They developed heterogeneous endline relationships with their AICs, ranging from reflective mirrors of actual selves to projections of idealized capabilities.

Beyond these findings, iterative refinement served as a medium for exploring possible selves. Participants developed heterogeneous relationships exhibiting patterns ranging from Actual-Anchored exploration (the AIC as a mirror for self-discovery) to Ideal-Anchored exploration (the AIC as an aspirational alter ego). Those prioritizing surprise and ownership gravitated toward reflective exploration, while those valuing clarity oriented toward constructive idealization.

These findings have practical implications for designing AIC systems. Rather than pursuing one-size-fits-all solutions, future systems should recognize and support these distinct exploratory modes. The ability to maintain living, evolving AI representations, rather than static snapshots, opens new possibilities for professional development, where AICs serve as dynamic tools for iterative refinement and self-exploration.

Future work should explore longer-term dynamics of AIC evolution and relationships, more sophisticated control mechanisms beyond behavior refinement, and how the living representation paradigm manifests across diverse professional contexts. Understanding how to design systems that support sustained, meaningful evolution of self-representations will become increasingly important as AI capabilities advance.

Acknowledgments

This work was supported by JST Grant Number JPMJPR23I4 and JPMJPF2205.

References

[1] Rumi Allbert, James K. Wiles, and Vlad Grankovsky. 2025. Identifying and Manipulating Personality Traits in LLMs Through Activation Engineering. arXiv:2412.10427 [cs.CL] https://arxiv.org/abs/2412.10427

[2] Pablo Arias-Sarah, Daniel Bedoya, Christoph Daube, Jean-Julien Aucouturier, Lars Hall, and Petter Johansson. 2024. Aligning the smiles of dating dyads causally increases attraction. *Proceedings of the National Academy of Sciences* 121, 45 (Oct 2024). doi:10.1073/pnas.2400369121

[3] Nico Brand, William Odom, and Samuel Barnett. 2021. A Design Inquiry into Introspective AI: Surfacing Opportunities, Issues, and Paradoxes. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1603–1618. doi:10.1145/3461778.3462000

[4] Nico Brand, William Odom, and Samuel Barnett. 2023. Envisioning and Understanding Orientations to Introspective AI: Exploring a Design Space with Meta.Aware. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 175, 18 pages. doi:10.1145/3544548.3581336

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

[6] Greg Bullock. 2017. Save time with Smart Reply in Gmail. https://blog.google/products/gmail/save-time-with-smart-reply-in-gmail/. https://blog.google/products/gmail/save-time-with-smart-reply-in-gmail/ Accessed: February 25, 2024.

[7] Heloisa Candello, Claudio Pinhanez, Michael Muller, and Mairieli Wessel. 2022. Unveiling Practices of Customer Service Content Curators of Conversational Agents. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 348 (Nov. 2022), 33 pages. doi:10.1145/3555768

[8] Sam W. T. Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. 2021. KinVoices: Using Voices of Friends and Family in Voice Interfaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 446 (Oct. 2021), 25 pages. doi:10.1145/3479590

[9] Runjin Chen, Andy Arditi, Henry Sleight, Owain Evans, and Jack Lindsey. 2025. Persona vectors: Monitoring and controlling character traits in language models. *arXiv preprint arXiv:2507.21509* (2025).

[10] Ziyang Chen and Stylios Moscholios. 2024. Using Prompts to Guide Large Language Models in Imitating a Real Person's Language Style. arXiv:2410.03848 [cs.CL] https://arxiv.org/abs/2410.03848

[11] Yi Fei Cheng, Hirokazu Shirado, and Shunichi Kasahara. 2025. Conversational Agents on Your Behalf: Opportunities and Challenges of Shared Autonomy in Voice Communication for Multitasking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 160, 18 pages. doi:10.1145/3706598.3714017

[12] Lluís C. Coll, Martin W. Lauer-Schmaltz, Philip Cash, John P. Hansen, and Anja Maier. 2025. Towards the "Digital Me": A vision of authentic Conversational Agents powered by personal Human Digital Twins. arXiv:2506.23826 [cs.ET] https://arxiv.org/abs/2506.23826

[13] Cathy Mengying Fang, Phoebe Chua, Samantha W. T. Chan, Joanne Leong, Andria Bao, and Pattie Maes. 2025. Leveraging AI-Generated Emotional Self-Voice to Nudge People towards their Ideal Selves. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 58, 20 pages. doi:10.1145/3706598.3713359

[14] Mehrad Faridan, Bheesha Kumari, and Ryo Suzuki. 2023. ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 203, 13 pages. doi:10.1145/3544548.3581381

[15] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 977, 17 pages. doi:10.1145/3613904.3641955

[16] Mariia Golovianko, Svitlana Gryshko, Vagan Terziyan, and Tuure Tuunanen. 2023. Responsible cognitive digital clones as decision-makers: a design science research study. *European Journal of Information Systems* 32, 5 (2023), 879–901. arXiv:https://doi.org/10.1080/0960085X.2022.2073278 doi:10.1080/0960085X.2022.2073278

[17] Grammarly. 2024. Grammarly. https://grammarly.com/

[18] 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 (01 2020), 89–100. arXiv:https://academic.oup.com/jcmc/article-pdf/25/1/89/32961176/zmz022.pdf doi:10.1093/jcmc/zmz022

[19] Qiqi He, Li Li, Dai Li, Tao Peng, Xiangying Zhang, Yincheng Cai, Xujun Zhang, and Renzhong Tang. 2024. From digital human modeling to human digital twin: Framework and perspectives in human factors. *Chinese Journal of Mechanical Engineering* 37, 1 (2024), 9.

[20] E Tory Higgins, Christopher JR Roney, Ellen Crowe, and Charles Hymes. 1994. Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of personality and social psychology* 66, 2 (1994), 276.

[21] Jess Hohenstein, Rene F Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13, 1 (2023), 5487.

[22] Jessica Huang, Ig-Jae Kim, and Dongwook Yoon. 2025. Mirror to Companion: Exploring Roles, Values, and Risks of AI Self-Clones through Story Completion. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 413, 15 pages. doi:10.1145/3706598.3713587

[23] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300469

[24] Hayeon Jeon, Suhwoo Yoon, Keyeun Lee, Seo Hyeong Kim, Esther Hehsun Kim, Seonghye Cho, Yena Ko, Soeun Yang, Laura Dabbish, John Zimmerman, et al. 2025. Letters from Future Self: Augmenting the Letter-Exchange Exercise with LLM-based Agents to Enhance Young Adults' Career Exploration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[25] Casey A Klofstad, Rindy C Anderson, and Susan Peters. 2012. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B Biological Sciences* 279, 1738 (Mar 2012), 2698–2704. doi:10.1098/rspb.2012.0311

[26] Sébastien Laniel, Dominic Létourneau, François Grondin, Mathieu Labbé, François Ferland, and François Michaud. 2021. Toward enhancing the autonomy of a telepresence mobile robot for remote home care assistance. *Paladyn, Journal of Behavioral Robotics* 12, 1 (2021), 214–237. doi:doi:10.1515/pjbr-2021-0016

[27] 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. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 91 (April 2023), 28 pages. doi:10.1145/3579524

[28] Seonghee Lee, Denae Ford, John Tang, Sasa Junuzovic, Asta Roseway, Ed Cutrell, and Kori Inkpen. 2025. IRL Dittos: Embodied Multimodal AI Agent Interactions in Open Spaces. arXiv:2504.21347 [cs.AI] https://arxiv.org/abs/2504.21347

[29] Joanne Leong, John Tang, Edward Cutrell, Sasa Junuzovic, Gregory Paul Baribault, and Kori Inkpen. 2024. Dittos: Personalized, Embodied Agents That Participate in Meetings When You Are Unavailable. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 494 (Nov. 2024), 28 pages. doi:10.1145/3687033

[30] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. 2008. Data-driven enhancement of facial attractiveness. In *ACM SIGGRAPH 2008 Papers* (Los Angeles, California) *(SIGGRAPH '08)*. Association for Computing Machinery, New York, NY, USA, Article 38, 9 pages. doi:10.1145/1399504.1360637

[31] Hazel Markus and Paula Nurius. 1986. Possible selves. *American psychologist* 41, 9 (1986), 954.

[32] Lauren Lee McCarthy and Kyle McDonald. 2019. MWITM (Man/Woman in the Middle). https://lauren-mccarthy.com/MWITM-Man-Woman-In-The-Middle. https://lauren-mccarthy.com/MWITM-Man-Woman-In-The-Middle Accessed: November 8, 2024.

[33] Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, Solon Barocas, and Ashton Anderson. 2022. Mimetic Models: Ethical Implications of AI that Acts Like You. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 479–490. doi:10.1145/3514094.3534177

[34] Hannah Mieczkowski, Jeffrey T Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-mediated communication: Language use and interpersonal effects in a referential communication task. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–14.

[35] Kana Misawa and Jun Rekimoto. 2015. ChameleonMask: a human-surrogate system with a telepresence face. In *SIGGRAPH Asia 2015 Emerging Technologies* (Kobe, Japan) *(SA '15)*. Association for Computing Machinery, New York, NY, USA, Article 5, 3 pages. doi:10.1145/2818466.2818473

[36] Kana Misawa and Jun Rekimoto. 2015. Wearing another's personality: a human-surrogate system with a telepresence face. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(ISWC '15)*. Association for Computing Machinery, New York, NY, USA, 125–132. doi:10.1145/2802083.2808392

[37] Meredith Ringel Morris and Jed R. Brubaker. 2025. Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 536, 14 pages. doi:10.1145/3706598.3713758

[38] Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. 2024. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109* (2024).

[39] Lorenzo Riano, Christopher Burbridge, and T Martin McGinnity. 2011. A study of enhanced robot autonomy in telepresence. In *Proceedings of Artificial Intelligence and Cognitive Systems, AICS*. AICS, 271–283.

[40] Zhonghui Shao, Jing Zhang, Haoyang Li, Xinmei Huang, Chao Zhou, Yuanchun Wang, Jibing Gong, Cuiping Li, and Hong Chen. 2024. Authorship style transfer with inverse transfer data augmentation. *AI Open* 5 (2024), 94–103.

[41] Kyriacos Shiarlis, Joao Messias, Maarten van Someren, Shimon Whiteson, Jaebok Kim, Jered Vroon, Gwenn Englebienne, Khiet Truong, Vanessa Evers, Noé Pérez-Higueras, et al. 2015. Teresa: A socially intelligent semi-autonomous telepresence system. In *Workshop on machine learning for social robotics at ICRA-2015 in Seattle*.

[42] Guangzhi Sun, Xiao Zhan, and Jose Such. 2024. Building Better AI Agents: A Provocation on the Utilisation of Persona in LLM-based Conversational Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, Article 35, 6 pages. doi:10.1145/3640794.3665887

[43] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. 2017. Synthesizing Obama: learning lip sync from audio. *ACM Trans. Graph.* 36, 4, Article 95 (July 2017), 13 pages. doi:10.1145/3072959.3073640

[44] Jon Truby and Rafael Brown. 2021. Human digital thought clones: the Holy Grail of artificial intelligence for big data. *Information & Communications Technology Law* 30, 2 (2021), 140–168.

[45] Qingxiao Zheng, Zhuoer Chen, and Yun Huang. 2025. Learning through AI-clones: Enhancing self-perception and presentation performance. *Computers in Human Behavior: Artificial Humans* 3 (2025), 100117. doi:10.1016/j.chbah.2025.100117