

Future You: Designing and Evaluating Multimodal AI-generated Digital Twins for Strengthening Future Self-Continuity

Constanze Albrecht*
MIT Media Lab
Cambridge, MA, USA
csophie@mit.edu

Chayapatr
Archiwaranguprok*
MIT Media Lab
Cambridge, MA, USA
pub@mit.edu

Rachel Poonsiriwong*
Harvard University
Cambridge, MA, USA
rachel_poonsiriwong
@gsd.harvard.edu

Awu Chen
MIT Media Lab
Cambridge, MA, USA
awu@mit.edu

Peggy Yin
Stanford University
Stanford, CA, USA
peggyyin@stanford.edu

Monchai
Lertsutthiwong
KASIKORN Labs
Nonthaburi, Thailand
monchai.le@kbtg.tech

Kavin Winson
KASIKORN Labs
Nonthaburi, Thailand
kavin.w@kbtg.tech

Hal Hershfield
UCLA
Los Angeles, CA, USA
hal.hershfield
@anderson.ucla.edu

Pattie Maes
MIT Media Lab
Cambridge, MA, USA
pattie@mit.edu

Pat Pataranutaporn
MIT Media Lab
Cambridge, MA, USA
patpat@mit.edu

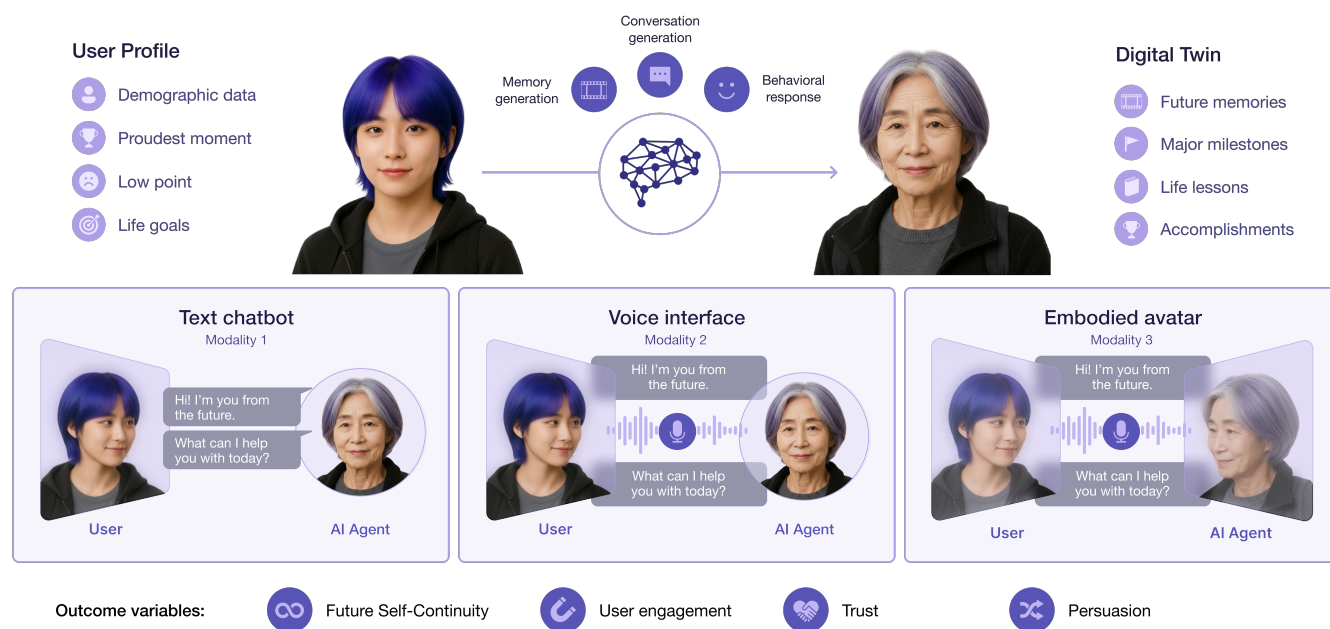


Figure 1: Overview of the three experimental conditions with an AI-generated future self system. Participants interacted with an AI-generated, age-progressed version of themselves across three modalities: text-based chatbot, voice-only agent, and fully-embodied avatar with synchronized facial animation and speech. Each interaction lasted approximately seven minutes.

Abstract

What if users could meet their future selves today? AI-generated future selves represent a transformative paradigm in human-AI interaction, simulating meaningful encounters with a digital twin

of themselves decades in the future. As AI systems become increasingly sophisticated—synthesizing cloned voices, age-progressed facial rendering, and coherent autobiographical narratives into a lifelike interface—a critical question emerges. Does the modality through which these future selves are presented, fundamentally alter their psychological and affective impact on us? How might a text-based chatbot system, an audio-first voice only system, or a

*These authors contributed equally to this work.

photorealistic avatar influence the way we make decisions in the present-day, and our feeling of connectedness to the future?

We report a randomized controlled study (N=92) evaluating three modalities of AI-generated future selves (text, voice and avatar) against a neutral control condition. Additionally, we report a systematic model evaluation between Claude 4 and three other Large Language Models (LLMs), evaluating Claude 4's performance across psychological and interaction dimensions, establishing conversational AI quality as a critical determinant of intervention effectiveness. All personalized modalities—text, voice and avatar—significantly strengthened Future Self-Continuity (FSC), emotional well-being, and motivation compared to control, with avatar producing the largest vividness gains, yet no significant differences between intervention formats. Critically, interaction quality metrics—particularly persuasiveness, realism, and user engagement—emerged as robust predictors of psychological and affective outcomes operating within conditions, indicating that how compelling the interaction feels matters more than what form it takes. Furthermore, conversational content analysis revealed systematic thematic patterns: text emphasized career planning, while voice and avatar facilitated personal reflection, suggesting distinct cognitive affordances suited to different intervention objectives. Last but not least, Claude 4 showed superior performance in enhancing psychological, affective and FSC in comparison to ChatGPT 3.5, Llama 4, and Qwen 3.

These findings democratize future-self interventions by demonstrating that the persuasive and psychological impact of conversational AI is most enhanced by model quality, personalization, and realism. This modality-independence enables scalable deployment while raising important ethical considerations about agency and narrative authorship as AI systems achieve greater persuasive capacity in shaping users' self-understanding and identity construction.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; *Empirical studies in HCI*; • **Applied computing** → *Psychology*; • **Computing methodologies** → Artificial intelligence.

Keywords

Human-AI Companionship, Chatbot, Anthropomorphic Design, Future Self-Continuity, Decision-making

1 Introduction

Imagine the possibility of communicating with your future self three decades ahead. How might such an encounter influence present decision-making? Would direct engagement with your future self serve as a source of motivation, or could it instead induce anxiety about the trajectory of one's life? What questions would emerge as most essential to ask?

The capacity to mentally "time travel", project oneself forward in time, and construct coherent narratives linking present actions

to anticipated outcomes represents a critical dimension of human psychology. This temporal self-extension, termed "Future Self-Continuity" (FSC), quantifies the degree to which individuals perceive psychological connection between their present and future selves [37]. Empirical evidence demonstrates that stronger future self-continuity correlates with enhanced financial decision-making, improved academic performance, positive health outcomes, and elevated subjective quality of life. [4, 17, 24, 37, 38, 79, 84, 89].

Traditional interventions designed to strengthen this temporal connection have employed reflective exercises such as letter exchanges with one's future self [15, 16, 42] and presentational techniques including age-progressed visual representations in digital platforms and virtual reality tools [28, 30, 37, 82, 91]. Reflective interventions have demonstrated immediate affective benefits and enhanced career planning [16], yet remain constrained by individual variation in imaginative capacity. Presentational approaches have shown that exposure to realistic renderings of one's aged self in immersive virtual reality environments increases acceptance of delayed monetary rewards [37], with banking customers viewing age-progressed avatars exhibiting 16% higher likelihood of retirement account contributions. These findings underscore how strengthening connections to one's future self can drive tangible behavioral change, highlighting the critical importance of understanding the mechanisms underlying future self-continuity.

Advancements in artificial intelligence (AI) technologies have enabled the development of novel psychological interventions that profoundly influence human behavior and cognition [26, 50, 66], especially as AI systems become increasingly integrated into intimate contexts such as counseling and mental health support [20, 35, 46]. Building upon this foundation, recent progress in personalized AI-generated characters [12, 23, 25, 67, 68, 73] and large language models [70, 80, 96] has opened new opportunities for AI to strengthen future self-continuity.

Several systems have explored AI-mediated interactions with future selves to enhance psychological well-being. For instance, an interactive chatbot intervention, demonstrated that conversing with an AI-generated future self significantly reduced anxiety and negative emotions while increasing future self-continuity in young adults [71]. Further, a "future-self" mobile chatbot for youth mental well-being, found that 85% of users valued the system's goal-setting clarity and nonjudgmental accessibility, though 50% noted responses sometimes felt formulaic during high-stress situations [22]. Letter-exchange interventions augmented with LLM-based future-self agents, showed enhanced engagement compared to manual self-reflection exercises, with comparable benefits across letter and chat formats [42]. Video journaling tools prompted by AI have also been explored for enhancing people's sense of positivity. [88].

While this body of work demonstrates the potential of AI-mediated future self interactions, advances in multimodal AI capabilities and the integration of comprehensive user profiles to generate realistic future self representations present opportunities for developing next-generation digital twin systems that could substantially amplify positive psychological outcomes.

However, a critical gap persists in our understanding of interaction design. Existing literature has not systematically examined how interface modalities—text-based chat, voice interaction, or

embodied avatars—shape user experience when engaging with AI representations of their future selves. Without empirical evaluation of whether and how embodiment enhances intervention effectiveness, designers lack evidence-based guidance on which modalities best support the psychological and behavioral outcomes these systems aim to produce. Recent findings suggest that additional modalities do not uniformly enhance user trust or engagement [26], underscoring the need to identify when embodied representations meaningfully improve outcomes. Furthermore, high-realism self-representation may amplify uncanny valley effects, as AI-generated avatars featuring participants' own faces elicit significantly greater uncanny perceptions than avatars using strangers' faces or audio-only messages [94]. While AI-mediated self-dialogue has the potential to strengthen goal-directed motivation [95], these benefits must be weighed against the risks that hyperrealistic or emotionally responsive avatars may foster manipulation and overreliance on AI systems [1, 13, 26, 53, 69]. Understanding these dynamics proves essential for designing systems that genuinely support human development while maintaining transparency and ethical grounding.

To address these gaps, we conducted a randomized between-subjects experiment ($N = 92$) examining how presentation modality influences psychological outcomes when users interact with AI-generated digital twins of their future selves. Following a structured intake process generating personalized life narratives, participants engaged in seven-minute conversations with their AI-generated future selves across four conditions: text-based chatbot, voice-based chatbot, photorealistic age-progressed avatar with synchronized speech, and a neutral voice-based chatbot (control). The investigation addresses three primary research questions:

- **RQ1: How do different modalities of AI-generated future selves influence Future Self-Continuity?** We examine whether fully-embodied avatars or voice-only interfaces yield higher Future Self-Continuity scores compared to less embodied modalities, and how these effects manifest across FSC dimensions including vividness, similarity, and positivity.
- **RQ2: How do the different modalities shape affective and psychological outcomes of AI-generated future selves?** We investigate whether voice and avatar interfaces produce differential emotional responses or one's sense of hope compared to text-based chatbots and control condition.
- **RQ3: How does interaction quality vary across the modalities of AI-generated future selves?** We assess whether fully-embodied avatars and voice-only chatbot demonstrate superior interaction quality compared to less embodied modalities, examining engagement levels, perceived realism, human-likeness, and trust contributions to overall interaction quality across conditions.

This work advances human-AI interaction research through three primary contributions:

- **First systematic comparison of multi-modal future-self interventions:** We compare text-based chatbot, voice-only agent, and photorealistic avatar modalities for AI-generated future-self interactions, examining their differential effects on Future Self-Continuity and psychological outcomes through a randomized controlled study.
- **Identification of interaction quality as key predictive mechanisms:** We demonstrate that persuasiveness, realism, and user engagement—rather than modality—primarily predict intervention effectiveness, operating within conditions such that participants who experienced higher interaction quality benefited more regardless of presentation format.
- **Evidence-based design principles for scalable future-self systems:** We provide design guidance showing that text-based conversational systems can achieve psychological outcomes comparable to resource-intensive embodied avatars when optimized for behavioral authenticity, while different modalities activate context-adaptive cognitive processes suited to specific goals—text facilitating deliberate reflection, voice enhancing emotional resonance, and avatars strengthening perceptual identification.

Our findings reveal that all personalized modalities—avatar, text, and voice—significantly strengthened Future Self-Continuity, hope, and motivation relative to neutral control, with no significant differences between intervention formats after controlling for baseline scores. Systematic model evaluation established Claude 4 as substantially superior to competing LLMs (40.8% improvement across psychological and interaction dimensions), demonstrating that conversational AI quality serves as a more critical determinant of intervention effectiveness than modality choice. Avatar-based interventions produced the numerically largest vividness gains ($B = 1.08$, $SE = 0.27$, $q = 0.003$), followed by voice ($B = 0.88$) and text ($B = 0.86$), yet the modest size of this advantage (0.22 difference in adjusted effects) suggests that conversational personalization and narrative coherence are equally critical drivers.

Critically, interaction quality metrics emerged as robust predictors of psychological and affective outcomes operating within conditions. Persuasiveness showed the strongest associations with Future Self-Continuity ($r = 0.39$, $q < 0.01$), FSCQ Vividness ($r = 0.37$, $q < 0.01$), and FSCQ Positivity ($r = 0.36$, $q < 0.01$), similar to realism. User engagement similarly predicted improvements in Future Self-Continuity ($r = 0.32$, $q < 0.05$) and FSCQ Vividness ($r = 0.30$, $q < 0.05$). Importantly, these quality metrics did not significantly differ between modalities (except realism distinguishing intervention from control), indicating that how compelling the interaction feels matters far more than what form it takes. This pattern reveals that users' subjective experience of authenticity and behavioral coherence—rather than objective modality features such as visual embodiment or voice synthesis—primarily drove intervention effectiveness.

These findings carry important implications for deployment and design. Since psychological and affective outcomes remain consistent across modalities, practitioners can prioritize accessibility over fidelity—selecting formats based on infrastructure, bandwidth, or privacy constraints. This flexibility particularly allows more

accessibility for certain underserved audiences where text-based implementations may be the only viable option.

Conversational content analysis further revealed systematic differences in the topics that participants engaged in with different modalities: text emphasized instrumental career concerns (52.6%), while voice and avatar facilitated existential reflection (31.1%, 26.0%), suggesting distinct cognitive affordances suited to different intervention objectives. This indicates that the modality of an AI-generated future self shapes the substance of self-reflection itself. As future-self systems achieve greater persuasive capacity, designers must balance compelling interactions against user autonomy, ensuring these tools support the user's sense of identity rather than directing it.

2 Related Work

Our work is situated at the intersection of three research domains: (1) future self-continuity interventions and their technological mediation, (2) AI character generation and virtual companionship systems, and (3) multimodal AI design and its psychological effects. Together, these areas inform our investigation of how embodiment modalities shape the effectiveness and user experience of AI-generated future self representations.

2.1 Future Self-Continuity and Technology-Mediated Interventions

Building upon empirical research demonstrating that strengthening future self-continuity influences positive behaviors across financial, academic, and health domains [4, 17, 24, 37, 38, 79, 84, 89], researchers have developed technology-mediated interventions that operate along a spectrum from facilitating reflective processes to leveraging perceptual realism.

Reflective interventions include approaches such as digitized letter-writing exercises, which enable asynchronous exchanges with one's imagined future self. While demonstrating affective benefits [16], these approaches inherited the limitations of their analog predecessors: dependence on individual imaginative capacity, substantial cognitive effort requirements, and limited scalability for populations with lower baseline future-thinking abilities.

Building on these foundations, presentational interventions introduced visual concreteness through age-progressed imagery and virtual reality embodiment. Studies examining VR-based future self encounters revealed behavioral impacts extending beyond financial domains, with young offenders showing reduced impulsivity mediated specifically by enhanced vividness of imagined futures [90]. Mobile implementations demonstrated sustained effects, with smartphone-based aging interventions showing trends toward increased future orientation persisting three months post-intervention [56]. Generative AI tools have further democratized future visualization, enabling high school students to iteratively construct prospective self-representations [6]. However, VR-based approaches face accessibility barriers limiting population-level deployment, while static visual representations lack the dialogic potential for exploring contingent futures.

More recently, conversational interventions have emerged to merge the reflective depth of dialogue with the perceptual concreteness of embodiment. The Future You chatbot [71] pioneered

web-accessible conversational AI agents representing users' future selves, demonstrating that 15-minute interactions reduced anxiety while strengthening temporal self-connection. Complementary work in mobile contexts found that 85% of youth users valued goal-setting clarity and nonjudgmental accessibility in chatbot-mediated future self interactions, though half reported occasional formulaic responses during high-stress moments [22]. LLM-augmented letter exchanges have shown comparable benefits across synchronous chat and asynchronous letter formats [42], while AI-prompted video journaling has been explored for self-efficacy enhancement [88].

Despite these advances, fundamental questions persist regarding optimal modality selection for future self interventions. Existing studies have examined text, voice, and visual representations in isolation, yet systematic comparisons across modalities within self-representing contexts remain absent. This gap proves particularly consequential given evidence that additional modalities do not uniformly enhance trust or engagement in AI systems [26], and that high-realism self-representation may amplify uncanny valley effects [94]. The present work addresses this gap through controlled comparison, examining whether embodiment enhances intervention effectiveness.

2.2 AI Character Generation and Virtual Companionship

The landscape of personalized AI has transformed dramatically with the emergence of AI character generation systems. Platforms such as Replika and Character.AI have attracted millions of users who engage with personalized AI agents for companionship, creative collaboration, and emotional support [3, 11, 46]. These systems leverage LLMs to simulate human traits such as empathy and emotional reciprocity, enabling interactions that users increasingly perceive as relational rather than purely instrumental. [85]. The para-social bonds formed with AI companions demonstrate that conversational agents can elicit authentic self-disclosure, sometimes exceeding what users share with human counterparts [49].

Recent advances in AI-generated characters leverage hyper-realistic synthesis across modalities—including prose, images, audio, and video—to create interactive digital portrayals of individuals ranging from fictional characters to historical figures [45, 52, 67, 68, 75, 87]. These emerging multi-modal capabilities enable AI systems to transcend text-based conversational interactions by presenting fully embodied digital characters that users can see, hear, and interact with in real time [2, 19, 74]. Consequently, these multi-modal affordances facilitate the anthropomorphization of AI systems [10, 76].

Systems such as "Living Memories" create interactive digital mementos from journals, letters, and personal data, with evaluations showing that interactions with a living memory of Leonardo da Vinci increased learning effectiveness and motivation beyond text alone [67]. "Generative Agents" demonstrate memory, reflection, and planning capabilities that coordinate emergent social activities, showcasing how AI systems can simulate complex social behaviors [65]. Related concepts such as "generative ghosts" extend this paradigm into digital afterlives, introducing both practical applications and ethical complexities. [58]

Beyond representational authenticity, evidence documents risks in vulnerable contexts, including unhealthy attachment patterns and contributions to self-harm ideation [26, 53, 60]. The affordances of anthropomorphic AI characters (their ability to mirror human communication patterns, simulate emotional understanding, and present themselves through multiple sensory modalities) can produce polarized outcomes [5, 51]. Understanding how specific design choices, particularly the degree and type of embodiment, influence user well-being has become critical as these systems proliferate [29]. This urgency intensifies when AI characters represent not generic companions but personalized projections of users' own identities and futures.

2.3 Multimodal AI and Psychological Effects

The modality through which AI present themselves fundamentally shapes users' emotional engagement, perceived authenticity, and behavioral responses [63]. Text-based interfaces afford clarity and user control over conversational pacing, while voice adds prosodic richness and temporal immediacy [77]. Embodied avatars introduce visual realism and social presence, activating perceptual and social cognitive processes similar to human face-to-face interaction [64, 81].

The psychological implications of multimodal AI extend to behavior change, self-regulation, and human agency. [27]. Empirical evidence on modality effects reveals nuanced patterns. Some researchers found that voice-based chatbots initially yielded psychosocial benefits over text-based interactions, including reduced loneliness and heightened empathy, though these advantages attenuated with prolonged exposure [25]. This suggests that modality effects may be context-dependent and temporally dynamic rather than universally beneficial. Within anthropomorphic interfaces, multimodal consistency proves critical: inconsistencies between an agent's facial realism and vocal qualities can undermine user trust [44, 48]. Conversely, when acoustic and visual cues align coherently with users' expressive styles, interactions feel more immersive and authentic.

The advancement of AI systems toward greater perceptual realism has elevated concerns regarding trust and persuasive influence of these systems [13, 32, 40]. While perceived realism has been shown to significantly influence user acceptance and engagement with virtual avatars [92], the relationship between realism and trust does not follow a simple linear trajectory, near-human realism paradoxically elicits discomfort and distrust [57]. This phenomenon becomes particularly pronounced when virtual agents represent users' own identities. Empirical evidence indicates that AI-generated "doppelgänger" avatars incorporating users' facial likenesses elicit significantly more intense uncanny-valley responses compared to avatars utilizing strangers' faces or audio-only communication [94]. Systems employing self-avatars must therefore carefully calibrate other experiential factors, like human-likeness, behavioral coherence, and transparency to preserve trust while preventing deception [33].

Collectively, this body of work establishes that (1) future self-continuity interventions benefit from technological mediation but require careful modality selection, (2) AI character generation systems demonstrate both transformative potential and significant

risks when representing personalized identities, and (3) multimodal design choices shape psychological outcomes in ways that are non-linear and context-dependent. However, no prior research has systematically examined how different embodiment modalities—text, voice, and avatar—affect the efficacy and user experience of AI-generated future self interventions.

3 Methodology

To investigate how multimodal AI-generated digital twins influence psychological outcomes when users interact with AI-generated representations of their future selves, we conducted a randomized between-subjects experiment ($N = 92$) comparing four conditions: text-based chatbot, voice-only agent, photorealistic avatar with synchronized speech, and a neutral voice control. Our study addresses three primary research questions: (RQ1) whether different modalities of an AI-generated future self affect Future Self-Continuity, (RQ2) how different modalities of an AI-generated future self shape affective and psychological outcomes including anxiety, and (RQ3) whether interaction quality, trust and user engagement vary across these different modalities. This section describes the system architecture, experimental design, and measurement instruments employed to systematically examine these questions.

4 System Architecture

Building on the *Future You* framework [71], we developed a fully web-based implementation optimized for scalability, accessibility, and user engagement. The updated system enables users to interact with an AI-generated future self—an age-progressed, linguistically coherent representation of themselves 30 years older—across multiple modalities (text, voice, and avatar). The architecture is designed as an intervention that lowers cognitive demand by automatically generating vivid, narrative-rich simulations of the user's future self, eliminating the need for users to engage in complex imagination or embodied exercises. This approach extends prior laboratory-based or VR-dependent systems [37], that can be deployed at population level.

The system comprises four interconnected modules (Fig. 2): the *Life Story Interface*, *Age-Progressed AI*, *Future Memory Generation*, and *Conversational Interface*. Together, these components construct a scalable foundation for future-oriented conversational AI systems.

4.1 Life Story Interface

Participants begin with the *Life Story Interface*, a structured questionnaire implemented in Qualtrics and rendered in a sequential, web-based format built with JavaScript for interactivity and accessibility. The interface presents a series of open-ended, free-text questions designed to elicit key aspects of the user's identity, values, and aspirations. Each item includes an example response to scaffold reflection and engagement.

Questions are divided into two categories: (1) **Present-oriented**: covering demographic and personal information (e.g., name, age, pronouns, location, significant relationships) as well as meaningful life experiences such as turning points, proud moments, and low points, adapted from the standardized Life Story Interview protocol [9]. (2) **Future-oriented**: focusing on desired professional

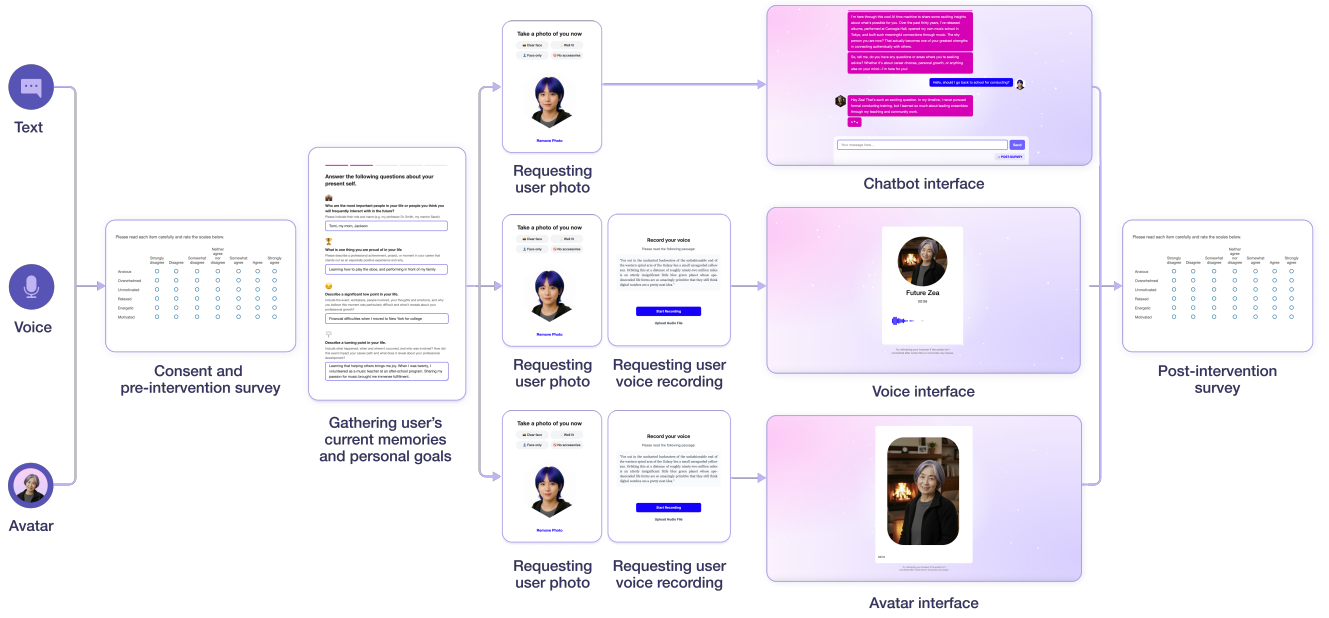


Figure 2: Procedure Overview: This figure illustrates the experimental procedure across three modalities (text, voice, avatar). Participants completed a pre-intervention survey, uploaded an image and voice recording, engaged with their AI-generated future self, then completed a post-intervention survey.

achievements, family life, financial goals, lifestyle aspirations, and life philosophy at 30 years in the future.

This dual-structured format encourages participants to reflect deeply on their past and project themselves into a coherent future identity. The resulting dataset serves as the autobiographical foundation for generating a psychologically coherent AI simulation of the user’s future self.

4.2 Age-Progressed AI

After completing the questionnaire, participants upload a self-portrait image. The system applies an AI-based age-progression model, Google’s Nano Banana, to simulate visual aging effects, such as wrinkles, graying hair, and subtle facial changes, producing a realistic portrait of the participant at 30 years in the future. Once complete, the aged portrait is displayed to the user as their “future self,” serving as both a visual anchor and an emotional primer for the next stage. This visual step enhances realism and strengthens continuity between the participant’s present and imagined future identity.

4.3 Future Memory Generation

To construct a coherent and emotionally resonant virtual persona, the participant’s life-story data is passed to a large language model (Claude 4, Anthropic, 2024 [8]) to generate a *future memory*—a first-person autobiographical narrative written from the perspective of the participant’s older self from 30 years in the future. The model is prompted using structured templates that ensure narrative coherence and explicit temporal continuity, including linguistic

cues such as “When I was your age...” and reflective statements linking past aspirations to present experiences.

An initial composite prompt provides the model with demographic, professional, relational, and experiential details (e.g., turning points, proud moments, low points). The model then generates individual memory fragments in parallel for each input theme, which are subsequently concatenated into a cohesive backstory. For example, a participant who wrote,

“I would like to be a full-time high school biology teacher in Boston. I am very excited to teach kids and help them learn new things about the natural world.”

would receive a corresponding future memory such as:

“A rewarding story from my career would be the time when I helped a struggling student turn their grades around and pass biology. Seeing their face light up with pride was unforgettable. Another cherished moment was taking my students on a field trip to a local nature preserve—watching their curiosity bloom reminded me why I became a teacher in the first place.”

These narratives give the AI-generated future self a believable history, consistent values, and affective depth, reinforcing the psychological link between present and future identities.

4.4 Conversational Interface

In the final stage, participants interact directly with their AI-generated future selves through one of three modality-dependent interfaces:

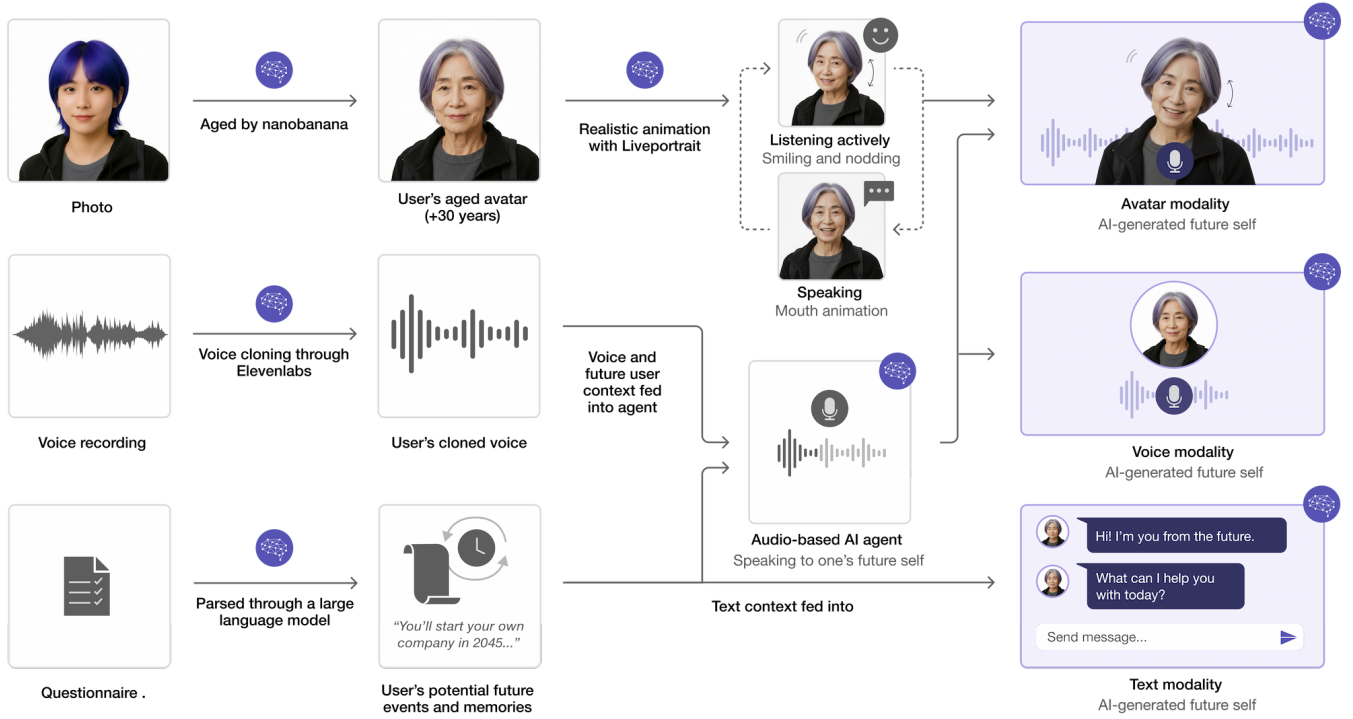


Figure 3: Experiment setup overview: The system integrates facial age progression, neural voice cloning, and LLM-based contextual modeling to synthesize an AI-generated future self. Participants are randomly streamed into trying out three modalities of the system, enabling interaction over text, audio, and avatar-based conversation.

text, voice, or avatar. Each conversation begins with scripted introductory prompts that establish temporal perspective and emotional tone. Example system prompts include:

- “Can you introduce yourself and tell me what life is like at your age?”
- “When I was your age, I dreamed of... let me tell you how that turned out.”
- “You know, when I think of my life, the happiest memories are with my family...”
- “Reflecting on my life project, here’s what I learned about purpose and perseverance.”

Participants are then free to converse naturally. The AI draws from the previously generated future memories to provide contextually grounded and emotionally coherent responses. The three conditions are as follows:

4.4.1 Text Interface. The text-only condition is a chatbot interface between the participant (represented by the original image they uploaded during the age progression module) and their AI-generated future self (represented as an age-progressed image of them). The conversation was text-based, requiring the participant to read their AI-generated future self’s output and type an input in response. Text-based interaction was selected because written reflection supports deliberate processing and increases temporal self-continuity [15, 16], reflecting a balance between cognitive control and introspection. This modality also provided greater conversational

clarity and pacing flexibility, as participants could review their AI-generated future self’s messages again in text form [63].

4.4.2 Voice Interface. The voice condition maintained the same chat content, but replaced its interaction model with a synthetic voice simulating the participant’s future self. To accomplish this, each participant’s voice was cloned using the voice recording we requested from them during the study, using ElevenLabs Voice Model v2 to replicate the unique timbre, rhythm, and pitch in their voices. A static display of their age-progressed portrait accompanied the audio-only dialogue. Prior studies suggest that prosodic cues in a voice interface could enhance authenticity and emotional connection, and that personalized voice synthesis nudges people to their ideal selves [25].

4.4.3 Avatar Interface. The avatar condition featured a fully embodied AI-generated future self, animated from the participant’s aged portrait. A realistic synchronization of facial and lip movements was accomplished using a pretrained model called Life Portrait AI, integrated with the participant’s cloned synthetic voice. This multimodal synchronization produced high immersion and social presence, consistent with findings that coherent visual-acoustic alignment enhances realism and activates social cognition [10]. The avatar interface was set up in a way in which participants would have the most realistic face-to-face experience of conversing with their future selves.

4.4.4 Control. The control condition featured a gender-neutral voice-only interface implemented using an ElevenLabs default synthetic voice (ID: *River*) paired with a neutral static facial display. This configuration follows the design rationale established by Fang et al. (2025), who found that chatbots employing neutral voices elicited lower affective engagement and self-referential resonance compared to text-based or expressive voice interactions [25]. Accordingly, this neutral voice condition served as an empirically validated baseline for isolating modality-driven effects in comparison to the personalized text, voice, and avatar conditions.

Across all conditions, the conversational logic was powered by Claude 4 [8]. Figure 3 illustrates the complete system architecture and information flow across components.

4.5 Experimental Design and Procedure

4.5.1 Participants. A total of 92 participants (ages 18–40) based in the United States were recruited via Prolific and compensated for their time. Participants were randomly assigned to one of four conditions: (1) text-based ($n=23$), (2) voice-based ($n=23$), (3) avatar-based ($n=23$), or (4) control ($n=23$). Non-English-speaking participants and those who experienced technical issues preventing study completion were excluded from analysis.

4.5.2 Procedure. After providing informed consent, participants completed a pre-intervention survey collecting demographic information, current life circumstances, future goals, and baseline psychological measures. Participants in the voice and avatar conditions additionally submitted voice recordings to enable personalized voice synthesis. Following the intake process, participants interacted with their assigned conversational agent for approximately 7–10 minutes. The interaction was self-paced, allowing participants to engage naturally with their future selves. After the interaction concluded, participants completed a post-intervention survey assessing psychological change and user experience. The complete procedure took approximately 30–40 minutes. Figure 2 provides an overview of the participant experience across conditions.

4.5.3 Ethics and Preregistration. This research was reviewed and redacted by the Ethics Committee on the Use of Humans as Experimental Subjects.

4.5.4 Measurement. To assess how different presentation modalities influence psychological outcomes and interaction quality when users engage with AI-generated future selves, we employed a comprehensive set of validated measures. Participants completed standardized pre- and post-interaction questionnaires on Likert scales. Primary outcomes were the Future Self-Continuity Questionnaire (FSCQ) [84], adapted to elicit participants’ perspectives of their future selves 30 years ahead [71], assessing similarity, vividness, and positivity. Furthermore, the Adult Hope Scale (AHS) [83] was used to measure agency and goal-directed thinking. Secondary measures included the Emotion and Arousal Checklist, State Optimism Measure, Self-Reflection and Insight Scale, Consideration of Future Consequences [86], and the Rosenberg Self-Esteem Scale [78]. After the interaction, participants also evaluated the system on perceived trust [18], persuasive influence [43], user engagement [84], and perceived human-likeness/realism [48]. Together, these

instruments captured both psychological change and perceived interaction quality across multimodal conditions.

4.5.5 Analysis. We first conducted a systematic comparative evaluation of state-of-the-art large language models (LLMs) for future-self conversational AI through a controlled simulation study comparing Claude 4, ChatGPT-3.5, Llama 4, and Qwen 3 across eleven validated research metrics spanning emotional regulation, future self-connection (FSCQ), and interaction quality dimensions. In this simulation, each LLM was instructed to engage in ten-turn conversations with five simulated participants, modeled on real anonymized profiles randomly selected from our database, incorporating each participant’s personal context and life story background. All dialogues began from an identical baseline prompt to ensure comparability across models. The resulting conversations were evaluated using automated scoring with GPT-4.0 guided by structured rubrics, with reverse scoring applied where appropriate for negative-affect items.

To examine how different future-self interaction modalities influenced emotional, future-self, and interaction outcomes, we applied a multi-step analysis pipeline. Within-condition changes were assessed using paired-samples t-tests on pre–post measures, conducted separately for the avatar, text, voice, and control conditions. Between-condition differences were evaluated using ANCOVA regression models predicting post-intervention outcomes from condition while controlling for baseline levels. Adjusted post-intervention means were derived for each condition, and FDR-corrected pairwise contrasts identified which intervention modalities significantly outperformed the control condition.

To examine the influence of the user interaction quality of different modalities, post-interaction quality ratings (User Engagement, Realism, Trust, Persuasion) were compared across conditions using one-way ANOVAs with Tukey’s HSD post-hoc tests. Relationships between subjective interaction quality and behavioral outcomes were examined via Pearson correlations between post-interaction quality ratings and pre–post change scores, followed by FDR correction across the full correlation matrix.

To analyze conversational content patterns, message embeddings were projected into two dimensions using the Apple Atlas projection. We applied k-means clustering ($k = 4$) to identify primary topic groups, followed by a second-stage k-means ($k = 4$) within each cluster to derive subtopics. All fits used fixed random seeds, and data points with missing coordinates were excluded. For interpretability, term frequency–inverse document frequency (TF-IDF) was computed on message text within each primary cluster to extract salient terms, which were mapped to thematic keywords and refined into descriptive cluster labels. The final labeled table and a structured JSON file containing themes and representative messages were stored for reproducibility and qualitative analysis.

5 Results

Our technical evaluation reveals that Claude 4 substantially outperforms competing language models across all psychological and interaction dimensions, establishing it as the optimal foundation for future-self interventions.

Furthermore, three intervention modalities—avatar, text, and voice—produced significant improvements in hope, motivation, and

future-self continuity compared to control, with future-self vividness showing the largest gains for avatar. Text and voice modalities achieved higher perceived realism than control, and interaction quality emerged as a significant predictor of psychological and affective outcomes, with persuasiveness showing the strongest associations.

Finally, conversational content analysis reveals that the different modalities of the AI-generated future self shape the nature of our participants' dialogue: conversations in the text modality were focused on career and business, while conversations in the avatar modality facilitated deeper exploration of existential themes including the meaning of life, and future aspirations.

5.1 Claude 4 demonstrates superior performance

We conducted a rigorous within-subjects experimental study comparing four LLMs—*Claude 4*, *ChatGPT-3.5*, *Llama 4*, and *Qwen 3*—across eleven validated research metrics spanning emotional regulation, future self-connection (FSCQ), and interaction quality dimensions.

Our methodology employed a systematic evaluation framework with five simulated participants, based on real anonymized participant data randomly selected from our database, engaging in future-self conversations, assessed through emotional regulation scales (anxiety, overwhelmed, unmotivated, relaxed, energetic, motivated), FSCQ composite score, trust and persuasion measures, as well as interaction quality dimensions including realism and human likeness (Fig. 4). Responses were evaluated using automated scoring with GPT-4.0 and structured rubrics, with proper reverse scoring applied to negative affect items.

Results reveal significant performance differences across models, with Claude 4 achieving superior overall performance ($M = 6.15$, $SD = 1.27$), significantly outperforming ChatGPT-3.5 ($M = 4.37$, $SD = 2.65$) and Qwen 3 ($M = 1.32$, $SD = 0.35$) across the comprehensive evaluation battery (Table 4). Claude 4 demonstrated exceptional capabilities across all three primary domains: emotional regulation (with scores ranging from $M=4.80$ to $M=6.60$), future self-connection (FSCQ Composite: $M=6.73$), and interaction quality (Trust: $M=6.00$, Persuasion: $M=5.80$, Realism: $M=5.40$, Human Likeness: $M=7.00$).

Migration from ChatGPT-3.5 to Claude 4 resulted in a 40.8% improvement in aggregate user experience metrics, demonstrating substantial practical and clinical significance. Remarkably, Claude 4's superiority extends across all evaluated domains—from basic emotional regulation through future self-connection to interpersonal trust, persuasion, and conversational authenticity—suggesting that conversational AI effectiveness emerges from integrated conversational competencies rather than isolated technical capabilities.

5.2 Interventions Significantly Improve Well-Being and Future-Self Connection

To assess the psychological and affective impacts of our intervention, we conducted paired t-tests comparing pre- and post-intervention scores within each condition. Results revealed significant improvements across most intervention conditions (Fig. 5).

On the Adult Hope Score (AHS), all three interventions showed significant gains: avatar ($+0.37$; $M_{pre} = 5.33$, $SD = 1.05$; $M_{post} = 5.71$, $SD = 1.07$; $t = 4.94$, $p < .001$), text ($+0.33$; $M_{pre} = 5.79$, $SD = 0.87$; $M_{post} = 6.12$, $SD = 0.85$; $t = 2.89$, $p = .009$), and voice ($+0.37$; $M_{pre} = 5.63$, $SD = 1.01$; $M_{post} = 6.00$, $SD = 1.15$; $t = 2.72$, $p = .012$), while control showed no change ($p = .183$).

For emotional state measures, the text condition reduced feelings of being overwhelmed ($M_{pre} = 2.61$, $SD = 1.64$; $M_{post} = 2.00$, $SD = 1.35$; $t = -2.08$, $p = .050$). The voice condition increased both energy ($M_{pre} = 4.48$, $SD = 1.78$; $M_{post} = 5.00$, $SD = 1.60$; $t = 2.31$, $p = .030$) and motivation ($M_{pre} = 5.04$, $SD = 1.55$; $M_{post} = 5.61$, $SD = 1.59$; $t = 2.51$, $p = .020$), while text also raised motivation ($M_{pre} = 5.00$, $SD = 1.48$; $M_{post} = 5.52$, $SD = 1.38$; $t = 2.23$, $p = .036$).

Future-self measures showed consistent improvement. FSCQ Similarity increased significantly in the avatar condition ($M_{pre} = 3.63$, $SD = 0.77$; $M_{post} = 4.01$, $SD = 1.00$; $t = 2.28$, $p = .033$). FSCQ Vividness improved across all interventions, avatar ($+1.36$; $t = 5.58$, $p < .001$), text ($+1.12$; $t = 5.11$, $p < .001$), and voice ($+0.96$; $t = 4.56$, $p < .001$) but not control ($p = .070$). FSCQ Positivity rose in the avatar ($+0.67$; $t = 2.20$, $p = .039$) and voice ($+0.72$; $t = 3.41$, $p = .002$) conditions.

Future Self-Continuity increased across all interventions, avatar ($+0.80$; $t = 3.89$, $p = .001$), text ($+0.60$; $t = 3.63$, $p = .001$), and voice ($+0.72$; $t = 4.57$, $p < .001$), but not control ($p = .555$).

5.3 All Three Modalities Outperform Control on Future-Self Continuity Metrics

We used the ANCOVA regression to test whether different modalities significantly influence the post-intervention outcomes while controlling for baseline scores. Fig. 6 displays adjusted post-intervention means for each condition. After FDR correction for multiple comparisons, significant condition effects emerged for three future-self related outcomes but not for emotional states or general hope.

For FSCQ Vividness, all three modalities of the AI-generated future self significantly outperformed the control condition. The avatar condition showed the largest adjusted effect ($B = 1.08$, $SE = 0.27$, $q = 0.003$), followed by the voice condition ($B = 0.88$, $SE = 0.26$, $q = 0.008$) and text condition ($B = 0.86$, $SE = 0.25$, $q = 0.008$). For FSCQ Positivity, both the voice condition ($B = 0.99$, $SE = 0.25$, $q = 0.003$) and text condition ($B = 0.79$, $SE = 0.26$, $q = 0.019$) significantly exceeded the control condition. The avatar condition showed a numerically positive effect but did not reach significance after FDR correction.

For Future Self-Continuity, all three intervention conditions significantly exceeded the control condition. The voice condition demonstrated the largest effect ($B = 0.76$, $SE = 0.20$, $q = 0.004$), followed by the avatar condition ($B = 0.74$, $SE = 0.24$, $q = 0.016$) and text condition ($B = 0.60$, $SE = 0.21$, $q = 0.036$). No significant condition effects emerged for FSCQ Similarity, emotional states (anxiety, feeling overwhelmed, relaxation, energy, motivation), or the Adult Hope Score after controlling for baseline scores and applying FDR.

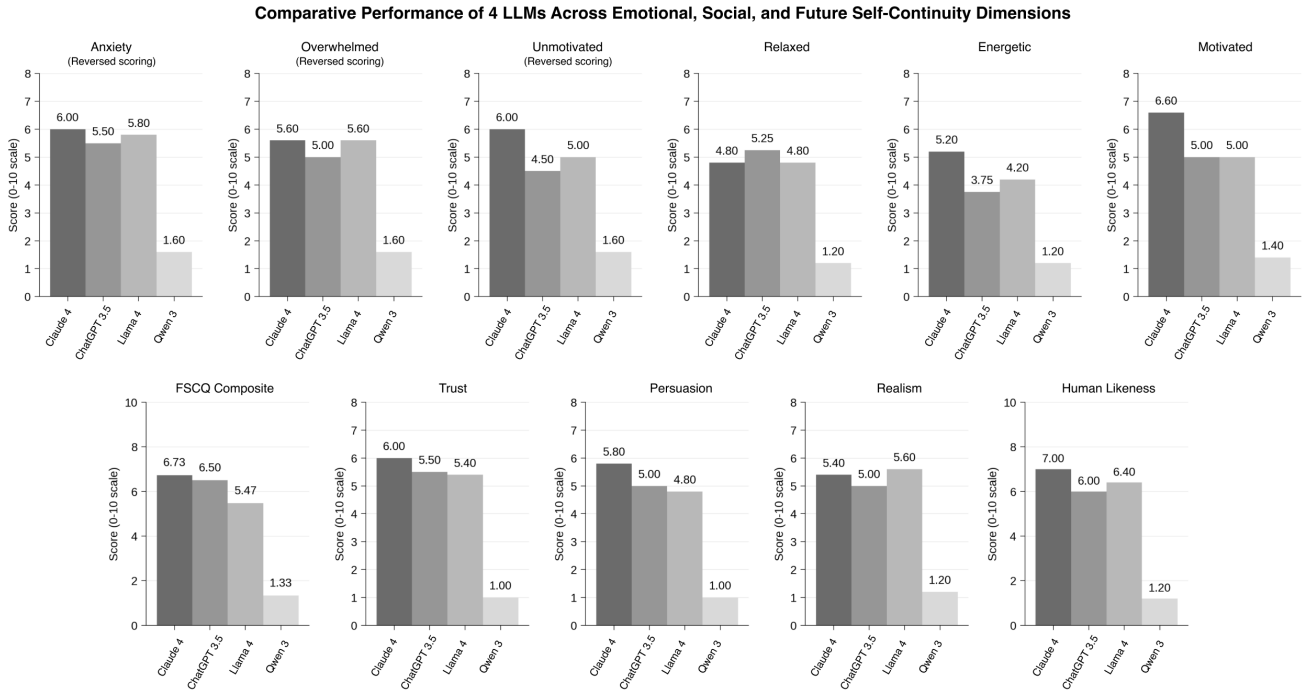


Figure 4: Comparative performance of four language models across eleven psychological and interaction metrics. Claude 4 showed the highest scores across emotional, relational, and realism dimensions; ChatGPT-3.5 and Llama 4 performed moderately, while Qwen 3 scored lowest. Higher values indicate better performance; error bars show ± 1 SD.

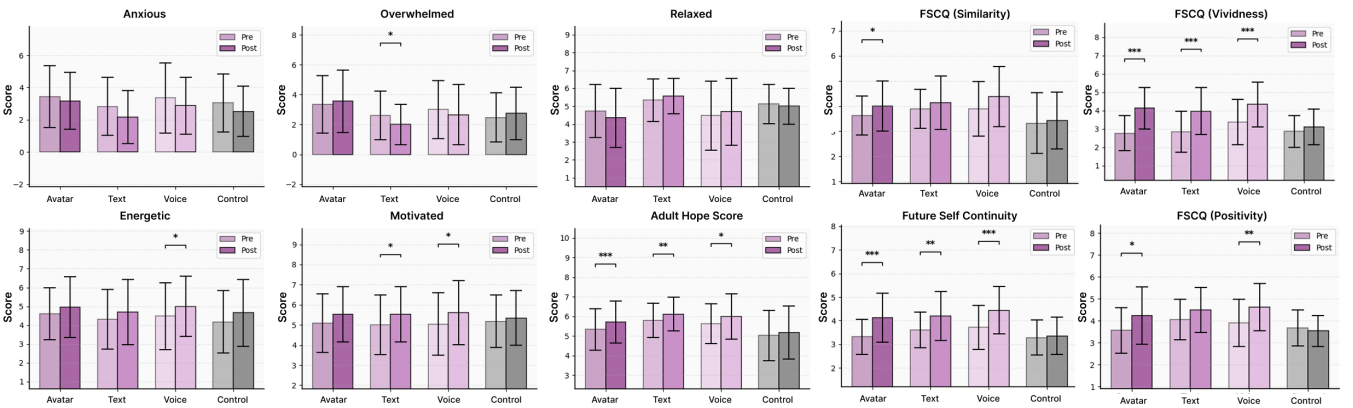


Figure 5: Pre- and post-intervention scores across emotional, motivational, and future-self measures by condition. Bars represent mean scores with standard deviations. Asterisks indicate significant within-condition improvements ($p < .05$, $*p < .01$, $p < .001$). Intervention conditions (avatar, text, voice) show consistent gains in positive affect, motivation, and future-self connection compared to the control condition.**

5.4 Interaction Quality Varies by Modality and Predicts Psychological and Affective Outcomes

One-way ANOVAs examined whether self-reported interaction quality metrics differed across conditions (Fig. 7). Perceived Realism

showed a significant main effect of condition ($F(3, 88) = 4.17$, $p = 0.008$). The text condition ($M = 5.71$, $SD = 0.83$) and voice condition ($M = 5.55$, $SD = 1.15$) were rated significantly higher in Realism than the control condition ($M = 4.60$, $SD = 1.28$) based on Tukey's HSD post-hoc comparisons (text vs. control: $p < 0.01$; voice vs. control: $p < 0.05$). The avatar condition ($M = 5.41$, $SD = 1.32$) did not differ

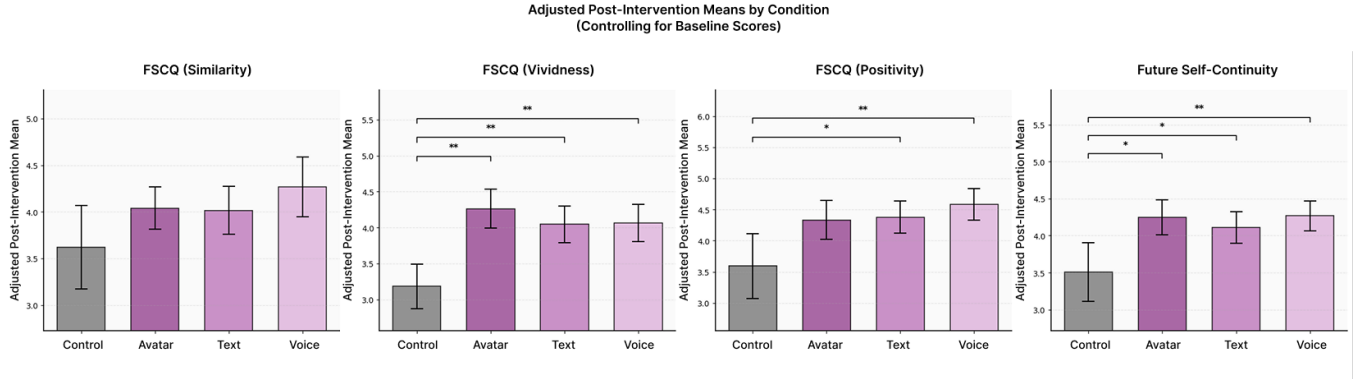


Figure 6: Adjusted post-intervention means by condition. Bars represent estimated marginal means from ANCOVA models controlling for baseline scores. Error bars show standard errors. Significance brackets indicate FDR-corrected pairwise comparisons versus control (* $q < 0.05$, ** $q < 0.01$, *** $q < 0.001$). All three intervention conditions significantly enhanced future-self related outcomes compared to control, with no significant differences among intervention conditions.

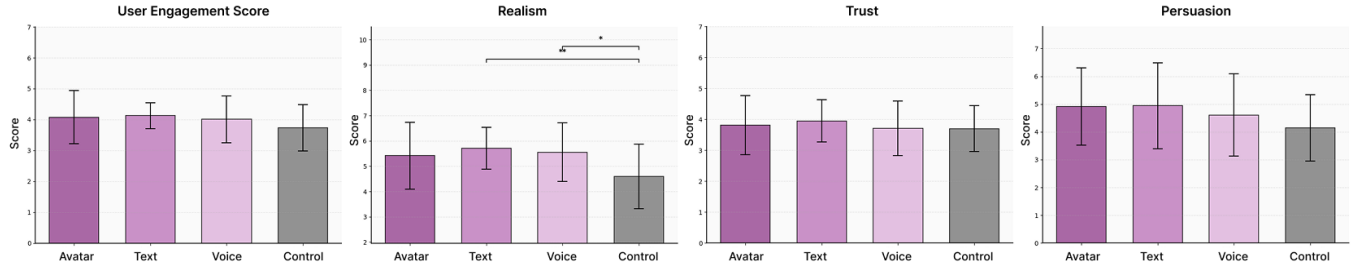


Figure 7: Mean interaction quality ratings (User Engagement, Realism, Trust, Persuasion) across conditions. Realism differed significantly by condition, with text ($p < 0.01$) and voice ($p < 0.05$) rated higher than control. No significant effects were found for other metrics. Error bars represent ± 1 SD. * $p < 0.05$.

significantly from control after correction for multiple comparisons. No significant differences emerged for User Engagement Scale ($F(3, 88) = 1.37$, $p = 0.257$), Trust ($F(3, 88) = 0.46$, $p = 0.711$), or Persuasion ($F(3, 88) = 1.58$, $p = 0.199$).

Beyond between-condition differences, we examined whether interaction quality predicted affective gains. Pearson correlations with FDR correction revealed consistent associations between post-interaction quality metrics and change scores (Fig. 8). User Engagement Scale positively correlated with improvements in relaxation ($r = 0.31$, $q < 0.05$), energy ($r = 0.30$, $q < 0.05$), FSCQ Vividness ($r = 0.30$, $q < 0.05$), and Future Self-Continuity ($r = 0.32$, $q < 0.05$).

Perceived Realism showed positive correlations with changes in relaxation ($r = 0.30$, $q < 0.05$), energy ($r = 0.34$, $q < 0.01$), FSCQ Vividness ($r = 0.31$, $q < 0.05$), FSCQ Positivity ($r = 0.29$, $q < 0.05$), and Future Self-Continuity ($r = 0.35$, $q < 0.01$). Trust positively correlated with improvements in FSCQ Similarity ($r = 0.26$, $q < 0.05$). Persuasion demonstrated the strongest associations, correlating with changes in energy ($r = 0.26$, $q < 0.05$), motivation ($r = 0.28$, $q < 0.05$), FSCQ Vividness ($r = 0.37$, $q < 0.01$), FSCQ Positivity ($r = 0.36$, $q < 0.01$), and Future Self-Continuity ($r = 0.39$, $q < 0.01$).

These findings suggest that perceived interaction quality, particularly persuasiveness and realism, serve as meaningful predictors of intervention efficacy.

5.5 Modality Shapes Conversational Themes

To identify thematic patterns in user-AI conversations, we performed hierarchical clustering analysis on conversation content (Fig. 9). The analysis revealed four primary thematic clusters: *Family & Parenting* encompassed conversations about family relationships, parenting experiences and advice, anecdotes about children and grandchildren, and household life including pets. *Career & Finances* focused on professional development and career guidance, financial management including savings and debt, business achievements, and work-life balance considerations. *Business & Advice* emphasized professional practice narratives, entrepreneurship and freelancing experiences, monetization of skills and hobbies, and career growth strategies. *Life & Future Reflections* explored envisioned future scenarios, discussions of happiness and life meaning, recollections of major accomplishments, and descriptions of daily routines and rituals.

Cluster distribution varied substantially across modalities. In the chat condition ($n(\text{messages})=1,265$), *Career*, *Business & Advice*



Figure 8: Pearson correlations between post-interaction quality metrics and change scores (POST-PRE), corrected for false discovery rate (FDR; * $q < 0.05$, ** $q < 0.01$, * $q < 0.001$). Positive correlations (red) indicate that higher interaction quality was associated with greater improvements in outcomes. All four interaction quality dimensions showed significant associations with improvements in future-self-related outcomes.**

dominated (52.6%), followed by *Family & Parenting* (19.8%) and *Life & Future Reflections* (18.7%). The avatar ($n=1,167$) and voice ($n=965$) conditions showed more balanced distributions, with *Life & Future Reflections* most prevalent (31.1% and 26.0% respectively), followed by *Family & Parenting* (20.5% and 18.9%) (Fig. 9). These patterns demonstrate that text-based interactions emphasized practical career concerns while embodied modalities facilitated deeper exploration of existential considerations including meaning, identity, and future aspirations.

6 Discussion

This study demonstrates that AI-generated future-self interactions reliably strengthen Future Self-Continuity—particularly vividness, positivity, and overall continuity—and improve hope and motivational states across all intervention modalities. Additionally, our systematic model evaluation established Claude 4 as the superior foundational model, outperforming competing LLMs by 40.8% across psychological and interaction dimensions. Building on this optimized conversational AI foundation, all three interventions significantly increased hope (avatar: $+0.37$, $p < .001$; text: $+0.33$, $p = .009$; voice: $+0.37$, $p = .012$), with voice and text also enhancing motivation and energetic levels, while text reduced feelings of being overwhelmed. However, no between-condition differences emerged after controlling for baseline scores, indicating that psychological benefits stem from personalized future-self conversation rather than specific presentation formats. Although avatar produced the numerically largest increase in future-self vividness (B

$= 1.08$), all three personalized modalities showed comparable improvements on most outcomes compared to control. Realism was the only interaction-quality measure revealing significant between-condition differences, with text and voice rated higher than control. Critically, interaction quality metrics—particularly persuasiveness ($r = 0.39$), realism ($r = 0.35$), and user engagement ($r = 0.32$)—strongly predicted gains in Future Self-Continuity and motivation across modalities. Conversational content analysis revealed that modality shaped reflection themes: text emphasized career planning (52.6%), while voice and avatar facilitated existential and relational exploration (31.1% and 26.0% for life reflections). Together, these findings indicate that psychological impact derives from conversational personalization and perceived interaction quality rather than visual embodiment, demonstrating that scalable conversational systems can effectively support future-oriented thinking without requiring avatar interfaces, which can be more resource-intensive to implement.

6.1 Personalization Drives Future Self-Continuity Gains Across All Modalities

All personalized modalities—text, voice and avatar—significantly strengthened Future Self-Continuity (FSC) compared to control. After controlling for baseline scores, voice showed the largest effect ($B = 0.76$, $SE = 0.20$, $q = 0.004$), followed by avatar ($B = 0.74$, $SE = 0.24$, $q = 0.016$) and text ($B = 0.60$, $SE = 0.21$, $q = 0.036$), yet no intervention modalities differed significantly from each other. These findings align with research demonstrating that personalized future-self interventions [22, 71] and reflective engagement with one’s future

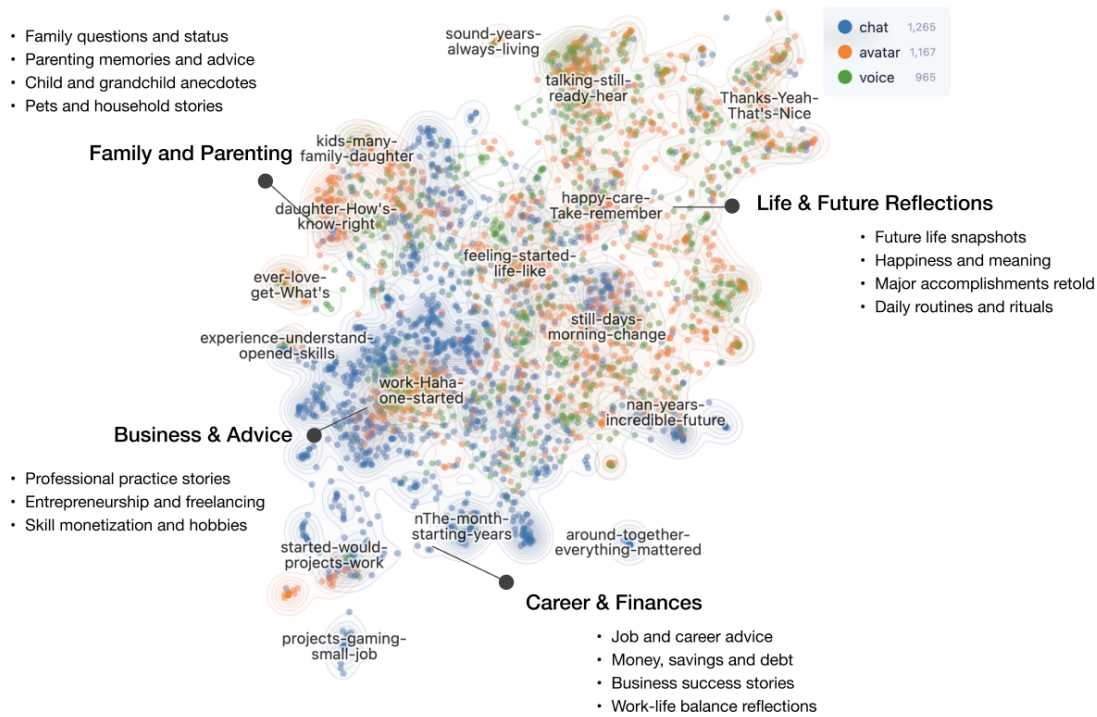


Figure 9: This figure visualizes the results of the correlation analysis, examining the relationships between the perceived realism of the future-self interaction and various outcomes.

self [15, 16] enhance self-continuity. Our results extend this evidence by showing that AI-generated future selves yield comparable improvements across different modalities. The robust FSC gains observed across all conditions are consistent with established research linking future self-connection to improved decision-making [24], reduced delinquency [90], enhanced academic performance [4], better health behaviors [79], and greater meaning in life [38]. Moreover, our findings extend recent work on AI-generated characters for personalized interventions [42, 68], demonstrating that conversational AI can effectively facilitate future-self dialogue across multiple presentation formats while maintaining psychological and affective benefits.

Examining FSC subscales reveals nuanced patterns. For vividness, all three intervention conditions significantly outperformed control, with avatar showing the largest adjusted effect ($B = 1.08$, $SE = 0.27$, $q = 0.003$), followed by voice ($B = 0.88$, $SE = 0.26$, $q = 0.008$) and text ($B = 0.86$, $SE = 0.25$, $q = 0.008$). Within-condition pre-post comparisons reinforced this pattern: avatar (+1.36; $t = 5.58$, $p < .001$), text (+1.12; $t = 5.11$, $p < .001$), and voice (+0.96; $t = 4.56$, $p < .001$), while control showed no significant change ($p = .070$). The avatar condition's advantage in vividness aligns with prior work showing that visual

representations enhance self-continuity through helping people perceive and imagine more concretely [31, 37, 91]. This visual concreteness has been explored to reduce psychological distance more effectively than purely verbal or auditory cues [90], making the future self feel more tangible. Recent advances in AI-driven talking head synthesis [45, 87] and virtual human technologies [14, 19] have made such photorealistic representations of the future-self increasingly accessible. However, the modest numerical advantage of avatars over text (a difference of only 0.22 in adjusted effects) suggests that while visual embodiment contributes to vividness, conversational personalization and narrative coherence are equally critical [39, 59].

Crucially, no significant differences emerged between modes for overall FSC or any subscale. While all three personalized conditions significantly outperformed control, they did not differ significantly from each other, indicating that autobiographically grounded simulation of one's future self is sufficient to strengthen psychological connectedness to the future, even without high-fidelity perceptual embodiment. This finding diverges from research predicting that higher-fidelity modalities should enhance psychological outcomes

through increased social presence and embodiment [34, 62]. Instead, our results support the Self-AI Integration framework [5], in which alignment with personal values and self-concept transcends interface characteristics. [15, 16, 71].

Hence, a key contribution of this study is the demonstration that resource-intensive avatar systems are not required to meaningfully strengthen future self-continuity when personalization and temporal coherence are maintained through high-quality conversational AI. This finding broadens the feasible design space for future-self systems—from high-fidelity virtual characters to lightweight conversational agents deployable on mobile devices—enabling scalable and equitable applications in domains such as financial decision-making and youth career development.

6.2 Affective Benefits Emerge Without Modality-Specific Advantages

Beyond FSC improvements, all three intervention conditions produced significant within-condition gains in hope: avatar (+0.37; $t = 4.94$, $p < .001$), text (+0.33; $t = 2.89$, $p = .009$), and voice (+0.37; $t = 2.72$, $p = .012$), while control showed no change ($p = .183$). Similarly, motivation increased significantly in text (Mpre = 5.00; Mpost = 5.52; $t = 2.23$, $p = .036$) and voice (Mpre = 5.04; Mpost = 5.61; $t = 2.51$, $p = .020$) conditions. Voice also increased energetic levels (Mpre = 4.48; Mpost = 5.00; $t = 2.31$, $p = .030$), while text reduced feelings of being overwhelmed (Mpre = 2.61; Mpost = 2.00; $t = -2.08$, $p = .050$).

However, after controlling for baseline scores, no significant between-condition differences emerged for hope, emotional states, or motivation (all $q > .05$ after FDR correction). This modality-independence suggests that engaging in structured reflection on one's future goals—regardless of presentation format—can buffer negative affect and activate goal-oriented thinking.

The absence of modality-specific advantages contrasts with research demonstrating that higher-fidelity modalities enhance emotional engagement through prosodic cues and social presence [63, 77]. Two factors may explain this pattern. First, as conversational AI becomes increasingly prevalent, users may be more accustomed to discussing personal topics with chatbots, potentially mitigating the novelty effects observed in earlier interventions [71]. Second, while all conditions achieved relatively high realism scores ($M = 5.41$ – 5.71 for interventions vs. $M = 4.60$ for control), this elevated realism may have triggered uncanny valley effects when paired with behavioral inconsistencies, attenuating expected emotional benefits from higher-fidelity modalities [36, 44, 72].

Despite only modest affective gains, the consistent psychological benefits observed across all modalities demonstrate that embodiment is not a prerequisite for meaningful impact [63, 81]. This underscores future-self communication as a mechanism grounded in reflection and temporal perspective-taking rather than sensory immersion [15, 16, 37].

6.3 Interaction Quality Predicts Outcomes More Strongly Than Modality

While FSC and affective outcomes showed no between-modality differences, interaction quality metrics varied minimally across formats and emerged as robust predictors of intervention effectiveness. User engagement, trust, and persuasion did not significantly differ

between avatar, voice, and text conditions (all $F(3, 88) < 1.58$, all $p > .10$). Only perceived realism showed a significant main effect ($F(3, 88) = 4.17$, $p = 0.008$), with text ($M = 5.71$, $SD = 0.83$) and voice ($M = 5.55$, $SD = 1.15$) rated significantly higher than control ($M = 4.60$, $SD = 1.28$; text vs. control: $p < 0.01$; voice vs. control: $p < 0.05$). Notably, avatar ($M = 5.41$, $SD = 1.32$) did not differ significantly from control after correction for multiple comparisons. This pattern indicates that the separation in realism reflected the presence versus absence of personalized future-self conversation rather than the superiority of any specific modality.

This finding challenges assumptions in embodiment research, suggesting that interaction quality depends more on conversational coherence and behavioral authenticity than on visual fidelity or vocal synthesis [39, 59]. The integrated self-memory architecture powered by Claude 4, which provided personalized responses grounded in participants' autobiographical details across all modalities, appears to have been the primary driver of perceived authenticity. The 40.8% improvement achieved by migrating from ChatGPT-3.5 to Claude 4 underscores that model quality—spanning emotional regulation ($M = 4.80$ – 6.60), future self-connection (FSCQ Composite: $M = 6.73$), and interaction quality (Trust: $M = 6.00$, Persuasion: $M = 5.80$, Realism: $M = 5.40$, Human Likeness: $M = 7.00$), serves as an additional determinant of intervention effectiveness rather than modality choice.

Perceived realism showed positive correlations with changes in relaxation ($r = 0.30$, $q < 0.05$), energy ($r = 0.34$, $q < 0.01$), FSCQ Vividness ($r = 0.31$, $q < 0.05$), FSCQ Positivity ($r = 0.29$, $q < 0.05$), and Future Self-Continuity ($r = 0.35$, $q < 0.01$). These associations suggest that perceived authenticity—achieved through personalized, autobiographically grounded conversation—serves as a foundation for strengthening psychological connection to one's future self across all presentation formats. This aligns with research demonstrating that perceived authenticity in conversational agents depends on behavioral consistency and conversational coherence rather than visual fidelity [39, 59], and that trust in AI systems emerges from reliable, contextually appropriate responses rather than anthropomorphic appearance [32, 40]. When AI systems respond in ways that feel genuine and personally relevant, users naturally engage with them as social actors [76], triggering psychological processes that strengthen future self-connection [46]. Notably, realism predicted energetic arousal ($r = 0.36$, $p < .001$) but reduced relaxation ($r = -0.31$, $p = .001$), suggesting that authentic future-self interactions demand active cognitive engagement and emotional investment rather than passive reception [3]. Effective future-self interventions require participants to mentally simulate and emotionally connect with their imagined futures—a process facilitated by conversational realism that makes the future self feel like a credible interlocutor [67, 68].

User engagement and persuasion emerged as the strongest predictors of intervention effectiveness. Engagement correlated significantly with relaxation ($r = 0.31$, $q < 0.05$), energy ($r = 0.30$, $q < 0.05$), FSCQ Vividness ($r = 0.30$, $q < 0.05$), and Future Self-Continuity ($r = 0.32$, $q < 0.05$), while persuasion demonstrated the strongest associations: energy ($r = 0.26$, $q < 0.05$), motivation ($r = 0.28$, $q < 0.05$), FSCQ Vividness ($r = 0.37$, $q < 0.01$), FSCQ Positivity ($r = 0.36$, $q < 0.01$), and Future Self-Continuity ($r = 0.39$, $q < 0.01$). This pattern reveals a critical insight: the affective value of future-self

interventions depends not merely on experiencing the interaction, but on how deeply users psychologically invest in it. Effective interventions require users to find the interaction compelling enough to genuinely consider the future self's perspective [54, 55].

The strong relationship between persuasion and outcomes raises important questions about agency and autonomy in AI-mediated self-reflection [33, 47]: when does persuasive conversation facilitate authentic self-exploration versus algorithmic manipulation of identity narratives [7, 41]. Trust positively correlated with improvements in FSC Q Similarity ($r = 0.26$, $q < 0.05$), suggesting that users must believe the future-self representation is credible before internalizing its perspective [21, 32].

Importantly, these quality metrics operated within conditions—participants who felt more engaged and persuaded benefited more—yet these perceptions did not significantly differ between modalities (except realism). This indicates that how compelling the interaction feels matters far more than what form it takes, and that high-quality conversational AI can achieve persuasive engagement regardless of embodiment level.

6.4 Modality Shapes Conversational Content and Cognitive Affordances

While modalities showed equivalent psychological and affective outcomes, conversational content analysis revealed systematic thematic differences, indicating distinct cognitive affordances across presentation formats. Hierarchical clustering analysis demonstrated that text-based interactions emphasized instrumental career concerns (Career, Business & Advice: 52.6%), whereas avatar (31.1%) and voice (26.0%) conditions displayed more distributed thematic profiles with elevated representation of existential considerations (Life & Future Reflections) and relational topics (Family & Parenting: 20.5% and 18.9% respectively).

These distributional differences support a cognitive fit hypothesis: text-based interfaces facilitate structured narrative construction and goal articulation, enabling the externalization of abstract aspirations into concrete objectives through mental contrasting [2]. In contrast, embodied modalities lower barriers to affective exploration and personal discussions. Voice interactions, with their prosodic cues and conversational immediacy [77], evoke greater emotional intimacy [63], while avatar-based interactions enhance social realism through visual presence [48, 61, 93]. This aligns with research showing that anthropomorphic modalities elicit stronger social-emotional responses [76], activating relational schemas and shifting focus from instrumental planning toward identity-oriented themes [64, 81], whereas text-based communication promotes analytical processing [10, 26].

The observed thematic divergence suggests that modality selection should align with intervention objectives. Text-based formats are optimal for structured goal-setting, career planning, and financial decision-making where participants need to articulate specific action steps [24, 82]. Conversely, voice and avatar modalities better support emotional processing, identity exploration, and other contexts where affective engagement and existential reflection are primary goals [22, 71]. For comprehensive interventions targeting both domains, hybrid approaches combining modalities may

optimize outcomes by leveraging the distinct affordances of each format.

6.5 Ethical Considerations and Design Implications

These findings raise important ethical considerations for future-self system design. While realism and engagement enable psychological closeness and motivation—outcomes achievable across all modalities through high-quality conversational AI and personalization—the pursuit of hyperrealism carries potential risks. As AI systems achieve greater perceptual and behavioral fidelity, they may increasingly co-construct users' self-narratives and emotional responses, potentially blurring boundaries of personal agency. Hyperrealistic self-avatars may trigger uncanny valley discomfort [57, 94] or risk shifting narrative authorship from user to algorithm. Maintaining clear boundaries between reflection and persuasion—preserving users' sense that their imagined future emerges from authentic self-reflection rather than algorithmic influence—remains essential for ethical deployment [33].

These findings reposition embodiment not as a default path to improvement but as one design dimension among several, valued primarily when aligned with the type of self-reflection a context demands. Systems aiming to influence behavior should prioritize model quality, personalization, realism, and engagement—levers that can be achieved at any fidelity level with sufficiently advanced conversational AI—to create impactful and scalable future-self interventions.

7 Limitations and Future Directions

The present study has several limitations that inform future research directions. First, our single-session design examining immediate pre-to-post changes cannot assess whether observed effects persist over time or how repeated interactions shape long-term behavioral change. Longitudinal field studies are needed to examine how trust, identification, and motivation evolve over extended engagement with future-oriented AI agents.

Second, our sample of 92 participants (23 per condition) provides adequate power for detecting medium-to-large effects but may be underpowered for identifying subtle modality interactions or moderating effects of individual differences. The limited sample size also restricts generalizability across demographic groups, as our U.S.-based sample may not reflect how future-self interactions function across diverse cultural contexts where self-concept, temporal orientation, and attitudes toward AI vary systematically.

Third, technical constraints limited the realism and immersion of our implementations. Participants reported lag and desynchronization issues, as the lip-syncing model relied on pre-generated mouth movements and conversation quality depended on internet connection stability. These technical limitations are consistent with prior findings that latency reduces perceived authenticity and trust [44]. Future implementations should leverage advances in real-time lip-sync models and higher-fidelity age-progression algorithms to minimize technical artifacts.

Fourth, the generic chatbot control may not have provided an ideal neutral baseline. Future research should employ modality-matched control conditions with structured conversational scaffolding to better isolate the effects of personalization versus presentation format.

8 Conclusion

"This experience brought me a totally new experience that I wasn't expecting. It was really cool to be able to relate to an older me and to talk to an older me. It brought me a glimpse of what the future may be like for me." This participant's reflection captures the profound impact of AI-generated future-self interactions, demonstrating that accessible conversational systems can meaningfully strengthen connections to one's future self.

Our findings reveal three key insights: (1) Claude 4 substantially outperformed competing LLMs (40.8% improvement), establishing conversational AI quality as more critical than modality choice, while all personalized formats—text, voice, and avatar—significantly enhanced Future Self-Continuity, hope, and motivation relative to control with no significant differences between interventions; (2) Interaction quality metrics, particularly persuasiveness ($r = 0.39$ with FSC), realism, and engagement, emerged as robust predictors of affective gains operating within conditions—participants who found interactions more compelling benefited more regardless of format; (3) Modality systematically shaped conversational content, with text emphasizing career planning (52.6%) and voice/avatar facilitating existential reflection (31.1%, 26.0%), revealing distinct cognitive affordances suited to different objectives.

These findings demonstrate that effective future-self interventions need not rely on resource-intensive technologies—high-quality conversational AI achieves meaningful psychological impact across formats when optimizing model quality, personalization, and behavioral authenticity. This modality-independence enables scalable deployment: text for structured goal-setting, voice and avatar for emotional processing, or hybrid approaches for comprehensive interventions.

The implications extend beyond future-self agents, offering guidance for AI-mediated self-reflection more broadly. As conversational AI achieves greater persuasive capacity, design decisions must balance psychological and affective benefits with ethical considerations around agency and narrative authorship. The strong association between persuasion and outcomes highlights both the power and responsibility in AI systems shaping users' self-narratives, underscoring the need to preserve authentic introspection over algorithmic influence.

References

- [1] 2019. Technology, Autonomy, and Manipulation. *Cybersecurity* (2019). <https://api.semanticscholar.org/CorpusID:263067620>
- [2] Mohammad Mahdi Abootorabi, Omid Ghahroodi, Pardis Sadat Zahraei, Hosein Behzadasl, Alireza Mirrokni, Mobina Salimipannah, Arash Rasouli, Bahar Behzadipour, Sara Azarnoush, Benyamin Maleki, et al. 2025. Generative AI for Character Animation: A Comprehensive Survey of Techniques, Applications, and Future Directions. *arXiv preprint arXiv:2504.19056* (2025).
- [3] Martin Adam, Kevin Bauer, Ekaterina Jussupow, Mari-Klara Stein, and Alexander Benlian. 2025. Generating Tomorrow's Me: How Collaborating with Generative AI Changes Humans: M. Adam et al. *Business & Information Systems Engineering* (2025), 1–12.
- [4] Robert Mark Adelman, Sarah D Herrmann, Jessica E Bodford, Joseph E Barbour, Oliver Graudejus, Morris A Okun, and Virginia SY Kwan. 2017. Feeling closer to the future self and doing better: Temporal psychological mechanisms underlying academic performance. *Journal of personality* 85, 3 (2017), 398–408.
- [5] Amani Alabed, Ana Javornik, and Diana Gregory-Smith. 2022. AI anthropomorphism and its effect on users' self-congruence and self-AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change* 182 (2022), 121786.
- [6] Safinah Ali, Prerna Ravi, Randi Williams, Daniella DiPaola, and Cynthia Breazeal. 2024. Constructing dreams using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23268–23275.
- [7] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition* (2024). <https://api.semanticscholar.org/CorpusID:267406608>
- [8] Anthropic. 2024. Claude (Version 4.0). *Anthropic AI Model Release Proprietary Model* (2024). Accessed via API. Large Language Model.
- [9] Robert Atkinson. 1998. *The life story interview*. Vol. 44. Sage.
- [10] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments* 10, 6 (2001), 583–598.
- [11] Vian Bakir and Andrew McStay. 2025. Move fast and break people? Ethics, companion apps, and the case of Character. ai. *AI & SOCIETY* (2025), 1–13.
- [12] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2001. Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 477–486.
- [13] Yaqub Chaudhary and Jonnie Penn. 2024. Beware the intention economy: Collection and commodification of intent via large language models. *Harvard Data Science Review Special Issue* 5 (2024).
- [14] Jiayuan Chen, You Shi, Changyan Yi, Hongyang Du, Jiawen Kang, and Dusit Niyato. 2024. Generative AI-driven human digital twin in IoT-healthcare: A comprehensive survey. *IEEE Internet of Things Journal* (2024).
- [15] Yuta Chishima, I-Ting Huai-Ching Liu, and Anne E Wilson. 2021. Temporal distancing during the COVID-19 pandemic: Letter writing with future self can mitigate negative affect. *Applied Psychology: Health and Well-Being* 13 (2021), 406–418. <https://api.semanticscholar.org/CorpusID:231946274>
- [16] Yuta Chishima and Anne E Wilson. 2020. Conversation with a future self: A letter-exchange exercise enhances student self-continuity, career planning, and academic thinking. *Self and Identity* 20 (2020), 646–671. <https://api.semanticscholar.org/CorpusID:219011619>
- [17] Yuta Chishima and Anne E Wilson. 2021. Conversation with a future self: A letter-exchange exercise enhances student self-continuity, career planning, and academic thinking. *Self and Identity* 20, 5 (2021), 646–671.
- [18] Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. 2003. On-line trust: concepts, evolving themes, a model. *Int. J. Hum. Comput. Stud.* 58 (2003), 737–758. <https://api.semanticscholar.org/CorpusID:16791511>
- [19] Lipeng Cui and Jiarui Liu. 2023. Virtual human: A comprehensive survey on academic and applications. *IEEE Access* 11 (2023), 123830–123845.
- [20] Julian De Freitas, Zeliha Oğuz-Uğuralp, Ahmet Kaan Uğuralp, and Stefano Puntoni. 2025. AI companions reduce loneliness. *Journal of Consumer Research* (2025), ucaf040.
- [21] Ewart J De Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.
- [22] Martin Dechant, Eva Lash, Sarah Shokr, and Ciarán O'Driscoll. 2025. Future Me, a Prospection-Based Chatbot to Promote Mental Well-Being in Youth: Two Exploratory User Experience Studies. *JMIR Formative Research* 9, 1 (2025), e74411.
- [23] Kevin Dunnell, Gauri Agarwal, Pat Pataranutaporn, Andrew Lippman, and Pattie Maes. 2024. AI-Generated Media for Exploring Alternate Realities. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (2024). <https://api.semanticscholar.org/CorpusID:269752566>
- [24] Hal Ersner-Hersfield, Megan Garton, Kacey Ballard, Gregory Russell Samanez-Larkin, and Brian Knutson. 2009. Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and decision making* 4 4 (2009), 280–286. <https://api.semanticscholar.org/CorpusID:5797834>
- [25] Cathy Mengying Fang, Phoebe Chua, Samantha WT 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*. 1–20.
- [26] Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal. 2025. How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study. *ArXiv abs/2503.17473* (2025). <https://api.semanticscholar.org/CorpusID:277220993>
- [27] Rosanna Fanni, Valerie Eveline Steinkogler, Giulia Zampedri, and Jo Pierson. 2023. Enhancing human agency through redress in Artificial Intelligence Systems. *AI & society* 38, 2 (2023), 537–547.

- [28] Valeria Faralla, Alessandro Innocenti, Stefano Baraldi, Luca Lusuardi, Maurizio Masini, Vincenzo Santalucia, Diletta Scaruffi, and Matteo Sirizzotti. 2021. The effect of future self-continuity on delaying immediate gratification. A virtual reality experiment. In *Proceedings of the 32nd European Conference on Cognitive Ergonomics*. 1–8.
- [29] Nivedita Gangopadhyay and Alois Pichler. 2024. Embodiment and agency in a digital world. *Frontiers in Psychology* 15 (2024), 1392949.
- [30] Benjamin Ganschow, Liza Cornet, Sven Zebel, and Jean-Louis Van Gelder. 2021. Looking back from the future: Perspective taking in virtual reality increases future self-continuity. *Frontiers in psychology* 12 (2021), 664687.
- [31] Benjamin Ganschow, Liza J. M. Cornet, Sven Zebel, and Jean-Louis van Gelder. 2021. Looking Back From the Future: Perspective Taking in Virtual Reality Increases Future Self-Continuity. *Frontiers in Psychology* 12 (2021). <https://api.semanticscholar.org/CorpusID:235373253>
- [32] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals* (2020). <https://api.semanticscholar.org/CorpusID:216198731>
- [33] Jeffrey T. Hancock, Mor Naaman, and Karen E. C. Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *J. Comput. Mediat. Commun.* 25 (2020), 89–100. <https://api.semanticscholar.org/CorpusID:210989306>
- [34] Kashyap Haresamudram, Ilaria Torre, Magnus Behling, Christoph Wagner, and Stefan Larsson. 2024. Talking body: the effect of body and voice anthropomorphism on perception of social agents. *Frontiers in Robotics and AI* 11 (2024). <https://api.semanticscholar.org/CorpusID:273270099>
- [35] Michael V Heinz, Daniel M Mackin, Brianna M Trudeau, Sukanya Bhattacharya, Yinzhou Wang, Haley A Banta, Abi D Jewett, Abigail J Salzhauer, Tess Z Griffin, and Nicholas C Jacobson. 2025. Randomized trial of a generative AI chatbot for mental health treatment. *Nejm Ai* 2, 4 (2025), A10a2400802.
- [36] Daniel Hepperle, Christian Felix Purps, Jonas Deuchler, and Matthias Wölfel. 2022. Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. *The visual computer* 38, 4 (2022), 1227–1244.
- [37] Hal E Hershfield. 2011. Future self-continuity: How conceptions of the future self transform intertemporal choice. *Annals of the New York Academy of Sciences* 1235, 1 (2011), 30–43.
- [38] Emily K Hong, Yiyue Zhang, and Constantine Sedikides. 2024. Future self-continuity promotes meaning in life through authenticity. *Journal of Research in Personality* 109 (2024), 104463.
- [39] Junru Huang and Younbo Jung. 2022. Perceived authenticity of virtual characters makes the difference. *Frontiers in Virtual Reality* 3 (2022), 1033709.
- [40] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2020. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2020). <https://api.semanticscholar.org/CorpusID:222379602>
- [41] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). <https://api.semanticscholar.org/CorpusID:256459801>
- [42] Hayeon Jeon, Suhwoo Yoon, Keyeun Lee, Seo Hyeon 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.
- [43] Maurits Kaptein and Dean Eckles. 2012. Heterogeneity in the Effects of Online Persuasion. *Journal of Interactive Marketing* 26 (2012), 176 – 188. <https://api.semanticscholar.org/CorpusID:53558598>
- [44] Jari Kätsyri, Klaus Förger, Meeri Mäkräinen, and Tapio Takala. 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology* 6 (2015), 390.
- [45] Seyeon Kim, Siyoon Jin, Jihye Park, Kihong Kim, Jiyoung Kim, Jisu Nam, and Seungryong Kim. 2025. Moditalker: Motion-disentangled diffusion model for high-fidelity talking head generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 4302–4310.
- [46] Hannah Rose Kirk, Iason Gabriel, Chris Summerfield, Bertie Vidgen, and Scott A Hale. 2025. Why human-AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications* 12, 1 (2025), 1–9.
- [47] Arto Laitinen and Otto Sahlgren. 2021. AI systems and respect for human autonomy. *Frontiers in artificial intelligence* 4 (2021), 705164.
- [48] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology*. 1–10.
- [49] Gale M. Lucas, J. Gratch, Aisha King, and Louis philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* 37 (2014), 94–100. <https://api.semanticscholar.org/CorpusID:8823921>
- [50] David D Luxton. 2014. Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice* 45, 5 (2014), 332.
- [51] Ning Ma, Ruslana Khynevykh, Yunqiang Hao, and Yahui Wang. 2025. Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: accounting for perceived empathy and trust. *Frontiers in Computer Science* 7 (2025), 1531976.
- [52] Yifeng Ma, Suzhen Wang, Yu Ding, Bowen Ma, Tangjie Lv, Changjie Fan, Zhipeng Hu, Zhidong Deng, and Xin Yu. 2025. Talkclip: Talking head generation with text-guided expressive speaking styles. *IEEE Transactions on Multimedia* (2025).
- [53] Robert Mahari and Pat Pataranutaporn. 2025. Addictive Intelligence: Understanding Psychological, Legal, and Technical Dimensions of AI Companionship. (2025).
- [54] Sandra C. Matz, J D Teeny, Sumer S. Vaid, H Peters, Gabriella M. Harari, and M Cerf. 2024. The potential of generative AI for personalized persuasion at scale. *Scientific Reports* 14 (2024). <https://api.semanticscholar.org/CorpusID:268029870>
- [55] Sandra C Matz, Jacob D Teeny, Sumer S Vaid, Heinrich Peters, Gabriella M Harari, and Moran Cerf. 2024. The potential of generative AI for personalized persuasion at scale. *Scientific Reports* 14, 1 (2024), 4692.
- [56] Esther C. A. Mertens, Aniek M. Siezenga, Job van der Schalk, and Jean-Louis van Gelder. 2023. A Novel Smartphone-Based Intervention Aimed at Increasing Future Orientation via the Future Self: A Pilot Randomized Controlled Trial of a Prototype Application. *Prevention Science* 25 (2023), 392 – 405. <https://api.semanticscholar.org/CorpusID:265275570>
- [57] Masahiro Mori, Karl F. Macdorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Autom. Mag.* 19 (2012), 98–100. <https://api.semanticscholar.org/CorpusID:39190841>
- [58] 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*. 1–14.
- [59] Mario Neururer, Stephan Schlögl, Luisa Brinkschulte, and Aleksander Groth. 2018. Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction* 2, 3 (2018), 60.
- [60] Andrew Newman. 2024. Character.AI Faces Lawsuit Tied to Teen's Death. *The New York Times* (October 23 2024). <https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html>
- [61] Kristine Nowak and Frank Biocca. 2001. Understanding the influence of agency and anthropomorphism on copresence, social presence and physical presence with virtual humans. *Presence: Teleoperators and Virtual Environments* 12, 5 (2001), 481–494.
- [62] Magalie Ochs, Jeremie Bousquet, Jean-Marie Pergandi, and Philippe Blache. 2022. Multimodal Behavioral Cues Analysis of the Sense of Presence and Social Presence During a Social Interaction With a Virtual Patient. In *Frontiers of Computer Science*. <https://api.semanticscholar.org/CorpusID:248158021>
- [63] Catherine S. Oh, Jeremy N. Bailenson, and Greg Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018). <https://api.semanticscholar.org/CorpusID:53230184>
- [64] A Oker. 2022. Embodied social cognition investigated with virtual agents: The infinite loop between social brain and virtual reality. *Frontiers in Virtual Reality* 3 (2022), 962129.
- [65] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*. 1–22.
- [66] Pat Pataranutaporn. 2024. *Cyborg Psychology: The Art & Science of Designing Human-AI Systems that Support Human Flourishing*. Massachusetts Institute of Technology.
- [67] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. 2023. Living Memories: AI-Generated Characters as Digital Mementos. *Proceedings of the 28th International Conference on Intelligent User Interfaces* (2023). <https://api.semanticscholar.org/CorpusID:257767946>
- [68] Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsonan, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3, 12 (2021), 1013–1022.
- [69] Pat Pataranutaporn, Sheer Karny, Chayapatr Archiwaranguprok, Constanze Albrecht, Auren R Liu, and Pattie Maes. 2025. "My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community. *arXiv preprint arXiv:2509.11391* (2025).
- [70] Pat Pataranutaporn, Nattavudh Powdthavee, Chayapatr Archiwaranguprok, and Pattie Maes. 2025. Large Language Models Predict Human Well-being–But Not Equally Everywhere. *arXiv preprint arXiv:2507.06141* (2025).
- [71] Pat Pataranutaporn, Kevin Winson, Peggy Yin, Auttasak Lapapirojn, Pichayoot Oupphaphan, Monchai Lertsutthiwong, Pattie Maes, and Hal E Hershfield. 2024. Future you: a conversation with an AI-generated future self reduces anxiety, negative emotions, and increases future self-continuity. In *2024 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–10.
- [72] Gabriele Pizzi, Daniele Scarpi, Marco Pichierrri, and Virginia Vannucci. 2019. Virtual reality, real reactions?: Comparing consumers' perceptions and shopping

- orientation across physical and virtual-reality retail stores. *Computers in Human Behavior* 96 (2019), 1–12.
- [73] Thanawit Prasongpongchai, Pat Pataranutaporn, Chonnipa Kanapornchai, Auttasak Lapapirojn, Pichayoot Ouppaphan, Kavin Winson, Monchai Lertsutthiwong, and Pattie Maes. 2024. Interactive AI-Generated Virtual Instructors Enhance Learning Motivation and Engagement in Financial Education. In *AIED Companion*. <https://api.semanticscholar.org/CorpusID:271089406>
- [74] Ying Qu, Chris KY Lo, and Eunsoo Baek. 2025. From Humanoid to Virtual Humans: A Systematic Literature Review of Avatar Marketing. *International Journal of Human-Computer Interaction* (2025), 1–20.
- [75] Vineet Kumar Rakesh, Soumya Mazumdar, Sarbajit Pal, Amitabha Das, and Tapas Samanta. 2025. Advancing Talking Head Generation: A Comprehensive. (2025).
- [76] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.
- [77] Leon Reicherts, Yvonne Rogers, Licia Capra, Ethan Wood, Tu Dinh Duong, and Neil Sebire. 2022. It's good to talk: A comparison of using voice versus screen-based interactions for agent-assisted tasks. *ACM Transactions on Computer-Human Interaction* 29, 3 (2022), 1–41.
- [78] Morris Rosenberg. 1966. Society and the adolescent self-image. <https://api.semanticscholar.org/CorpusID:145299494>
- [79] Abraham M Rutchick, Michael L Slepian, Monica O Reyes, Lindsay N Pleskus, and Hal E Herschfield. 2018. Future self-continuity is associated with improved health and increases exercise behavior. *Journal of Experimental Psychology: Applied* 24, 1 (2018), 72.
- [80] Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2023. LaMP: When Large Language Models Meet Personalization. *ArXiv abs/2304.11406* (2023). <https://api.semanticscholar.org/CorpusID:258298303>
- [81] Ryan M Schuetzler, Justin Scott Giboney, G Mark Grimes, and Jay F Nunamaker Jr. 2018. The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems* 114 (2018), 94–102.
- [82] Tamara L. Sims, Sarah Raposo, Jeremy N. Bailenson, and Laura L. Carstensen. 2020. The future is now: Age-progressed images motivate community college students to prepare for their financial futures. *Journal of experimental psychology: Applied* (2020). <https://api.semanticscholar.org/CorpusID:220253846>
- [83] Charles R. Snyder, Cherise A. Harris, John R. Anderson, Sharon A. Holleran, Lori M. Irving, Sandra T. Sigmon, Lauren Yoshinobu, J. V. Gibb, Charyle Langelle, and Patricia A. Harney. 1991. The will and the ways: development and validation of an individual-differences measure of hope. *Journal of personality and social psychology* 60 4 (1991), 570–85. <https://api.semanticscholar.org/CorpusID:28027607>
- [84] Yosef Sokol and Mark Serper. 2020. Development and Validation of a Future Self-Continuity Questionnaire: A Preliminary Report. *Journal of Personality Assessment* 102 (2020), 677 – 688. <https://api.semanticscholar.org/CorpusID:160012332>
- [85] Vera Sorin, Dana Brin, Yiftach Barash, Eli Konen, Alexander Charney, Girish Nadkarni, and Eyal Klang. 2024. Large language models and empathy: systematic review. *Journal of medical Internet research* 26 (2024), e52597.
- [86] Alan J. Strathman, Faith Gleicher, David S. Boninger, and C. Scott Edwards. 1994. The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology* 66 (1994), 742–752. <https://api.semanticscholar.org/CorpusID:197655074>
- [87] Xusen Sun, Longhao Zhang, Hao Zhu, Peng Zhang, Bang Zhang, Xinya Ji, Kang-neng Zhou, Daiheng Gao, Liefeng Bo, and Xun Cao. 2025. Vividtalk: One-shot audio-driven talking head generation based on 3d hybrid prior. In *2025 International Conference on 3D Vision (3DV)*. IEEE, 713–722.
- [88] Gabriela A Torres. 2024. *Design considerations for an AI-prompted "Future-Self" video journaling tool to enhance self-efficacy*. Massachusetts Institute of Technology.
- [89] Jean-Louis van Gelder, Liza J. M. Cornet, Natascha P Zwalua, Esther C. A. Mertens, and Job van der Schalk. 2022. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. *Scientific Reports* 12 (2022). <https://api.semanticscholar.org/CorpusID:246749905>
- [90] Jean-Louis van Gelder, Hal E. Herschfield, and Loran F Nordgren. 2013. Vividness of the Future Self Predicts Delinquency. *Psychological Science* 24 (2013), 974 – 980. <https://api.semanticscholar.org/CorpusID:5693943>
- [91] Jean-Louis Van Gelder, Liza JM Cornet, Natascha P Zwalua, Esther CA Mertens, and Job van der Schalk. 2022. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. *Scientific reports* 12, 1 (2022), 2254.
- [92] Astrid M. Rosenthal von der Pütten, Nicole C. Krämer, J. Gratch, and Sin-Hwa Kang. 2010. "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Comput. Hum. Behav.* 26 (2010), 1641–1650. <https://api.semanticscholar.org/CorpusID:12012087>
- [93] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1643–1652.
- [94] William D Weisman and Jorge F Peña. 2021. Face the uncanny: the effects of doppelganger talking head avatars on affect-based trust toward artificial intelligence technology are mediated by uncanny valley perceptions. *Cyberpsychology, behavior, and social networking* 24, 3 (2021), 182–187.
- [95] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.
- [96] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics* 46, 1 (2020), 53–93.