Introspectus AI: Long-term AI-Driven Dialogue Training To Promote Self-Reflection

SHENGYIN LI*, Keio University Graduate School of Media Design, Japan
GUANGYAO ZHU, Waseda University, Japan
DANYANG PENG, Keio University Graduate School of Media Design, Japan
XIMING SHEN, Keio University Graduate School of Media Design, Japan
CHENYU TU, Waseda University, Japan
XIARU MENG, Keio University Graduate School of Media Design, Japan
YUN SUEN PAI, University of Auckland, Inclusive Reality Lab, School of Computer Science, New Zealand
GIULIA BARBARESCHI, University of Duisburg-Essen, Research Center Trustworthy Data Science and Security, Germany
KOUTA MINAMIZAWA, Keio University Graduate School of Media Design, Japan

Fig. 1. (a) People commonly think about how to improve their performance on a variety of tasks. (b) A user is recording an episode of their daily life. (c) Introspectus AI probes its users to reflect on their daily activities.

*Corresponding Author

Authors' Contact Information: Shengyin Li, l.lishengyin@gmail.com, Keio University Graduate School of Media Design, Yokohama, Japan; Guangyao Zhu, zhuzgy@akane.waseda.jp, Waseda University, Kitakyushu, Japan; Danyang Peng, pengdanyang111@gmail.com, Keio University Graduate School of Media Design, Yokohama, Japan; Ximing Shen, ximing.shen@kmd.keio.ac.jp, Keio University Graduate School of Media Design, Yokohama, Japan; Chenyu Tu, tuchenyu@asagi.waseda.jp, Waseda University, Kitakyushu, Japan; Xiaru Meng, xiarumeng2019@outlook.com, Keio University Graduate School of Media Design, Yokohama, Japan; Yun Suen Pai, yun.suen.pai@auckland.ac.nz, University of Auckland, Inclusive Reality Lab, School of Computer Science, Auckland, New Zealand; Giulia Barbareschi, giulia.barbareschi@uni-due.de, University of Duisburg-Essen, Research Center Trustworthy Data Science and Security, Duisburg, Germany; Kouta Minamizawa, kouta@kmd.keio.ac.jp, Keio University Graduate School of Media Design, Yokohama, Japan.

Introspectus AI is a generative AI-based system designed to enhance self-reflection and support positive behavior change. By leveraging multimodal information from users' daily life recordings, it provides personalized and detailed feedback, aiming to deepen self-awareness and facilitate positive behavioral adjustments. This study explores the short-term and long-term impacts of interacting with Introspectus AI, focusing on its potential to enhance reflective practices and improve the acceptance of generative AI tools. Following the user experience was defined through an initial round of workshops with four experts. The resulting system was evaluated through a long-term study involving 64 participants. The results demonstrate that AI-supported interventions significantly improved engagement in self-reflection, the need for reflection, and insight, while also increasing user acceptance of generative AI over time. These findings underscore the potential of generative AI as a practical tool for self-improvement, offering insights into its broader applicability in promoting well-being and personal growth.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **Empirical studies in HCI**; • **Computing methodologies** → *Activity recognition and understanding*; Natural language generation.

Additional Key Words and Phrases: Generative AI, Self-reflection, Long-term intervention, Video understanding, Chat-bot

ACM Reference Format:
Shengyin Li, Guangyao Zhu, Danyang Peng, Ximing Shen, Chenyu Tu, Xiaru Meng, Yun Suen Pai, Giulia Barbareschi, and Kouta Minamizawa. 2025. Introspectus AI: Long-term AI-Driven Dialogue Training To Promote Self-Reflection. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW435 (November 2025), 43 pages.
https://doi.org/10.1145/3757616

1 Introduction

The ability to look back on our personal experiences, and analyze our thoughts, feelings, and actions to build a new understanding that can support decision-making around key aspects of our life is important for both personal well-being and social dynamics [95]. It is thus unsurprising that unpacking how individuals engage with self-reflection and investigating how technologies can support existing practices and offer opportunities for new forms of engagement, has been a topic of significant interest in HCI and CSCW communities [7, 12, 24, 85, 114]. Technologies for supporting self-reflection can encompass the use of personal informatics and data visualization in a health context and beyond [25, 26], serious games for behaviour change [45, 87], immersive technologies for perspective taking [61, 113], and mobile applications for in-situ journalism [53, 119].

Chatbots and voice-based conversational agents have been shown to be both effective and acceptable tools to support self-reflection as they facilitate natural dialectic processes which are commonly leveraged in both everyday life with other people in our social circles, and as part of professional consultations with mental health practitioners [7, 52, 58, 59, 68, 112].

While traditional self-reflection tools such as journaling or audio recording offer convenience and support emotional articulation from the point of view of the individual, they often fail to fully capture the situational context and behavioral nuances essential for meaningful insight. Ultimately, these methods depend heavily on memory recall, which may distort or selectively omit critical actions, reactions, or emotional cues which can only be captured in situ [40, 91, 103].

In contrast, first-person video recordings can capture behaviors, emotions, and social dynamics as they unfold—providing a rich, multi-layered account of lived experience. Collecting videos of ourselves performing a particular action which can be scrutinized after the fact, is an effective way to be able to re-examine our past and reflect on how we want to improve and has been extensively

ACM 2573-0142/2025/11-ARTCSCW435
https://doi.org/10.1145/3757616

used in various fields including sports, work and personal life [6, 60, 81, 104, 105]. The use of online tools or digital portfolios for categorization and analysis, not only facilitates self-reflection but can also enhance professional development[23, 31, 35].

The interrogation of such recordings offer an opportunity for grounded, multimodal feedback that connects reflection with actual embodied experience. This makes video-based reflection particularly relevant in action-oriented or interpersonal contexts, such as skill training, conversation dynamics, and daily routine awareness. By integrating AI with video-based input, it becomes possible to not only support deeper behavioral insight but also to enhance the temporal and emotional fidelity of reflective processes[30, 74].

Recent research has begun to explore how artificial intelligence—particularly large language models (LLMs)—can be leveraged to support the analysis and self-reflection of personal video data. However, these investigations have largely been application-specific, addressing narrow domains or tasks[75, 102]. As a result, it remains unclear how adaptable such AI-driven systems are to the diverse and evolving needs of individuals, who may wish to engage in self-reflection for different purposes depending on personal preferences and situational contexts[6, 7, 15].

Moreover, despite growing interest in generative AI tools, studies have shown that users remain cautious about adopting these systems in personal contexts, raising concerns about trust, transparency, and data privacy [46, 83]. Little is known about how people engage with AI-supported reflection over time, or how such technologies are perceived when used in flexible, self-directed ways in everyday life.

Our goal is to explore how a generative AI system can support self-reflection across a wide range of real-world situations—allowing individuals to engage with their own recorded experiences in personally meaningful and contextually relevant ways.

To guide this investigation, we pose the following research questions:

- **RQ1:** How does interacting with a generative AI agent grounded in users' own video recordings affect self-reflection outcomes—particularly users' perceived reflective ability?
- **RQ2:** How do users perceive the relevance, depth, and emotional resonance of feedback provided by such AI systems?
- **RQ3:** How do users' acceptance and engagement evolve when using generative AI for reflection in a real-world, multi-day context?

In this study, as shown in Fig 1, we present the design and deployment of Introspectus AI, a flexible web-based platform that allows users to upload previously recorded videos of specific activities and engage in dialogue with a purposefully designed chatbot agent that provides support for self-reflective practice through a conversation with their past selves. Introspectus AI operates by analyzing multimodal information from users' daily life recordings to provide personalized and detailed feedback, thereby enhancing self-reflection and supporting behavior change. As shown in Fig 2, we conducted a initial set of design workshops with four experts to determine the requirement of the system and define the user experience. The system was deployed as part of a week-long study to evaluate acceptability and perceived impact on self-reflection, compared to traditional journaling methods, with 64 participants utilizing a mixed methods approach. Our results show Long-term interaction with Introspectus AI not only enhances various self-reflection abilities, but also gradually increases user acceptance of generative AI applications.

Our study makes the following contributions to the CSCW community:

(1) The design and implementation of Introspectus AI, the first web-based platform purposefully designed for supporting dialectical self-reflection with an AI agent based on the re-examination of personal videos

(2) An evaluation of the usability, acceptability and effectiveness of AI-powered chatbot for self-reflection based on videos of past actions
(3) Discussion around implications around the use of generative AI platform to support self-reflection in the form of dialogue with past selves, highlighting both the potential and risks of such an approach.

2 Related Works

2.1 Self-reflection Training

Self-reflection is a key process in professional development and learning, serving as an essential method for cultivating self-awareness. It involves self-analysis, self-evaluation, and self-observation, representing a process of critically examining one's current and past thoughts, emotions, and behaviors [37, 66, 117, 118]. Neuroimaging research substantiates that self-reflection is a fundamental human behavior, with specific brain regions, particularly the ventral and dorsal medial prefrontal cortex, being activated during self-reflection [108, 108, 117]. The engagement of these brain regions underscores self-reflection role in developing self-awareness and emotional intelligence and allows individuals to critically analyze their thoughts, actions, and interactions with others[48, 79].

Self-reflection training has been found to enhance self-esteem and improve performance. Research by Johnson demonstrated that for individuals with high levels of self-criticism, engaging in self-reflection can improve mood and reduce negative emotions[50]. In the workplace, it enhances problem-solving skills and the ability to learn from experience [62]. In Cognitive Behavioral Therapy (CBT) training, self-practice and self-reflection (SP/SR) have been reported to benefit trainees [10]. In educational training, self-reflection exercises can help students redefine professional knowledge, improve academic performance, and strengthen self-efficacy beliefs [78, 122].

Various forms of reflection training, such as CBT, are widely used in mental health, professional development, leadership training, and education [10, 11, 34, 122]. One of the most common forms of reflection training is reflective dialogue, in which self-awareness, intuition, reflection, and listening are key elements[107]. Boyd et al. proposed that dialogic learning more effectively promotes the practice of critical reflection compared to discussion-based learning[13]. In reflective dialogue, participants engage in structured conversations with a coach, therapist, or even a peer to encourage reflection on their experiences and decisions, this process fosters deeper self-awareness, helps individuals identify patterns in their behavior, and supports emotional growth by providing a safe space for exploring thoughts and feelings, ultimately enhancing critical thinking and decision-making skills[47, 89, 100, 117].

Video-based self-reflection has proven valuable in enabling individuals to critically observe and analyze their behaviors. Research indicates that video recordings enhance teachers' ability to critically reflect on their practices, offering more analytical insights than memory-based methods[84, 97]. Similarly, web-based video tools have been shown to deepen reflective thinking, especially in classroom management and professional knowledge[60]. Moreover, previous research has documented how integrating digital video into ePortfolios supports self-reflection and peer feedback—video-based reflection's helps to foster professional growth and improve teaching across diverse educational settings[23, 81]. There is limited understanding of how to sustain and deepen self-reflection over time. Although traditional self-reflection methods can provide insights, they often rely heavily on personal motivation. In contrast, objective feedback and guidance can more effectively promote personal growth[56].

2.2 Design and Application of Chatbot

Dialogue enables individuals or groups to exchange information, share viewpoints, express emotions, and solve problems through two-way interaction, playing a vital role in understanding, relationship-building, and collaboration [5]. In 1960, Alan Turing proposed the Turing Test with the question, "Can machines think?" [106]. Building on this question, the concept of chatbot—online human-computer dialogue systems using natural language—began to attract increasing attention and find broader applications [1, 18].

With the rapid development of computer science, particularly advances in natural language processing and artificial intelligence [36, 64], chatbots, under the guidance of contemporary technological paradigms, have evolved from simple text-based interactions to multi-modal interfaces incorporating speech and images. This has made human-machine dialogue more natural and diverse, positioning chatbots as a significant research direction in human-computer interaction [2, 33, 93]. Natural language processing (NLP) enables chatbots to interpret user input by recognizing entities, intent, and context [73]. Techniques such as tokenization, named entity recognition (NER), and part-of-speech tagging are commonly employed in this process. Pre-trained language models like BERT [28] and GPT [16] have substantially enhanced the ability of chatbots to process natural language. Python libraries like NLTK and JSON are commonly used for efficient chatbot development, and some models achieve up to 93% accuracy in responding to queries [88]. Recent advances in large language models (LLMs) have significantly improved chatbot technology, conversational agents have rapidly evolved from rule-based chatbot to AI-driven chatbot, making it a valuable tool across various aspects of daily life [8, 44, 86].

Chatbots have become especially prominent in customer service, healthcare, education, and personal productivity. Kwon et al. found that chatbot-based writing practice improves second language learners' skills while offering a comfortable learning environment[65]. Evebot is used to diagnose negative emotions through positive, suggestive responses, alleviating stress and anxiety in adolescents to help prevent depression [116]. Simultaneously, several studies showed that nearly half of U.S. adults use digital voice assistants, primarily on smartphones, to perform daily tasks. Approximately one in five teenagers also utilize ChatGPT [1] to support their academic work [82, 99]. This finding highlights the significant role that chatbots and virtual assistants, such as Siri[2] and Alexa[3], play in simplifying everyday activities by managing tasks and improving efficiency in users' lives.

At the same time, the empathy and customization of chatbots are also being continuously explored. Jonell developed a chatbot capable of sensing user's physiological signals, such as skin temperature, respiratory rate, and pupil dilation. The neural network processes the collected physiological data to generate personalized responses, thereby enhancing the chatbot's adaptability to each user[51]. The tone-aware chatbot designed by Hu et al. for customer service uses deep learning to generate empathetic and enthusiastic responses based on customer tone, enhancing user satisfaction[43]. However, challenges such as data privacy and cultural sensitivity in chatbot responses remain [96]. In the future, empathetic chatbots are expected to understand human emotions better, reinforcing their role as essential tools in modern life.

2.3 AI-Assisted Self-Reflection

From rule-based response systems to Artificial Intelligence(AI) chatbots that increasingly attempt to understand user emotions, AI is being applied more frequently to support users' emotional

[1] https://openai.com/index/gpt-4/
[2] https://www.apple.com/siri/
[3] https://www.alexa.com/

well-being [2, 63, 120]. By leveraging big data analysis, AI can provide insights that assist individuals in exploring their thoughts, behaviours, and emotions in a manner that fosters self-awareness and introspection.

AI can evaluate emotional expressions in users' tone, text, and even voice through sentiment analysis, identifying emotional states and changes to better adapt to their emotional patterns [80, 120]. Research by Chen et al. and Lee et al. showed that AI can facilitate creativity and emotional self-expression, contributing to AI-based interventions in mental health education, therapy, and counselling[21, 69]. For example, based on GPT-2, Rajcic and Mccormack used facial expression recognition and poetry generation to encourage emotional self-reflection in users[92]. Torres utilized AI to guide users in recording video selfies featuring narratives addressed to their future selves, encouraging users to repeatedly watch the videos to facilitate emotional self-reflection[104]. These data-driven Emotion AIs help individuals understand emotional triggers and responses, supporting efforts to regulate emotions, identify sources of stress, or recognize sources of joy [52].

AI chatbots are increasingly being applied to support emotional reflection and promote mental well-being [20, 109]. Examples include Woebot [4], VOS [5], and Wysa [6], which assist users with daily schedule planning while encouraging journaling for self-reflection and tracking users' mental states through big data analysis. These chatbots promote AI-based therapy by engaging users in conversations, encouraging them to describe their experiences, and employing cognitive behavioral therapy (CBT) techniques to help individuals reframe negative thoughts [22, 29, 90, 94]. Likewise, AI is combined with traditional meditation practices to guide users in mindfulness exercises, helping them focus on the present moment, recognize past thoughts to potentially enhance awareness, and increase willingness for self-reflection and engagement while observing internal experiences [55, 80, 98].

The role of chatbots and AI in supporting and assisting users with emotional regulation and enhancing self-awareness has been widely demonstrated[21, 29, 92, 109]. In Mashhadi et al.'s research, the need for long-term, actionable feedback to effectively promote self-reflection is highlighted[77]. Most AI-based chatbots rely on users' ability to recall past activities, and relay events that are then unpacked through written or spoken dialogue. This can make it difficult to engage in in-depth analysis of complex activities for which it might be hard for a person to remember specific details which are important for self-reflection. In our study, we explore the integration of LLMs for video analysis and AI-based chatbot via the "Introspectus AI" web platform to support flexible self-reflective practices in daily life assessing both the effectiveness and acceptability of such an approach.

3 Methodology

In the era of LLMs, there are ongoing discussions about how to design AI systems that can support personal well-being and be seamlessly integrated into human life. Introspectus AI is designed to facilitate in-depth self-reflection by utilizing LLMs to analyze user-recorded videos of their daily behaviors. Unlike traditional self-reflection methods, which rely on memory or subjective recall [57, 101], Introspectus AI allows users to capture specific aspects of their daily lives via video and receive personalized, AI-driven feedback through a dialectical interaction with a chatbot. The system helps users gain deeper insights into their actions, promoting enhanced self-awareness and behavioral improvement through structured and targeted reflection.

[4]https://woebothealth.com/
[5]https://vos.health/en/
[6]https://www.wysa.com/

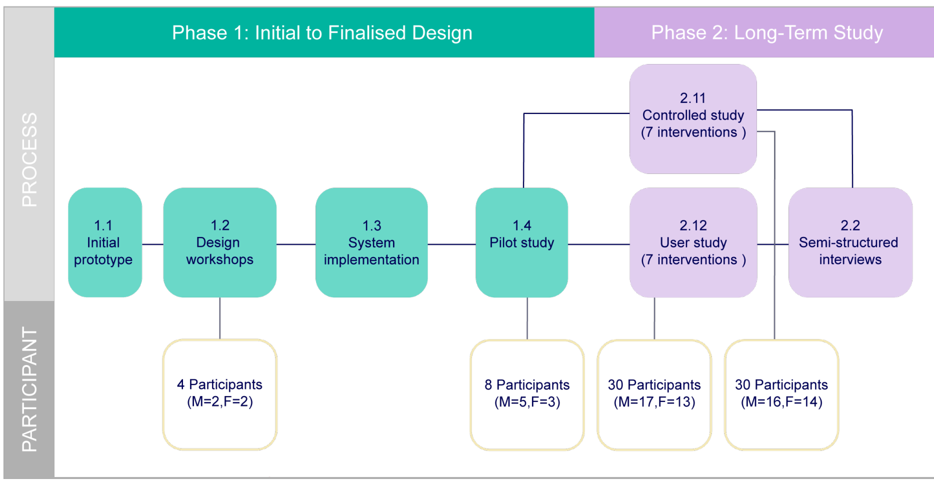

Fig. 2. The overall research workflow including the system development procedures and deployment of the long-term study.

3.1 Preprocessing of Introspectus AI

3.1.1 Processing Visual Information from Video. As show in Figure 3, Current LLMs face challenges in processing videos longer than 1 minute due to computational constraints and difficulties in maintaining contextual continuity [71, 111]. To address this challenge we propose an approach which encompasses two key aspects:

(1)Combine the Frame and Enhance with Key Frame. To ensure that long-duration videos could be effectively processed by LLMs, we implemented a hybrid frame representation strategy that combines uniform interval sampling with key frame selection. Interval sampling provides consistent temporal coverage, while key frames are selected based on scene changes and visual saliency, ensuring that semantically meaningful moments are captured. For each video, selected frames from both strategies were merged and grouped into sets of nine, which were then combined into 3 × 3 grid images. This method preserves the sequence of events while significantly reducing the total number of tokens passed to the model.

Each grid image serves as a condensed visual summary of a specific video segment. The LLMs was instructed to analyze these grids by inferring what occurred between frames, reconstructing timelines, and interpreting user behavior in context. The prompt included clear instructions to respond from a first-person perspective, such as: "Here are some frames from a video. Describe, based on the timeline, what happened from a first-person perspective, including the environment and reactions from people around me." (see Appendix A.1.3).

(2)Key frame. To improve the informativeness and semantic coverage of visual summaries used in dialogues with *"Introspectus AI"*, we enhanced traditional uniform frame sampling with an intelligent keyframe extraction algorithm. This method dynamically identifies meaningful visual transitions by combining three complementary techniques: Structural Similarity Index (SSIM)

analysis, histogram difference evaluation, and cumulative frame variation tracking[31, 105, 110]. Together, these approaches allow the system to preserve key moments in the video, particularly those relevant to user behavior, environmental changes, and social dynamics.

2.1 SSIM-Based Selection. We first evaluate perceptual similarity between adjacent frames using the Structural Similarity Index (SSIM), which considers luminance, contrast, and structural patterns[110]. If a sudden change occurs—such as a scene cut or gesture shift—the SSIM value drops significantly. We convert the frames to grayscale to reduce complexity and compare adjacent frames. A frame is selected as a keyframe if it is visually distinct from the previous one. We let I_t and I_{t-1} denote the grayscale frames at time t and $t-1$. A keyframe is selected if:

$$\mathrm{SSIM}(I_t, I_{t-1}) < \tau_s \tag{1}$$

τ_s is the threshold for keyframe detection, empirically set to 0.6.

2.2 Histogram-Based Selection. To capture changes in color distribution or composition, we compute normalized histograms in RGB or HSV color spaces for each frame. By measuring the divergence between histograms of consecutive frames, we detect visual differences even in scenes without structural variation[105]. We employ Bhattacharyya distance, Chi-square test, and Kullback–Leibler divergence as metrics. A frame is flagged as a keyframe if the histogram difference exceeds a threshold.

We let H_t be the normalized color histogram of frame I_t in RGB or HSV space. The histogram difference is computed as:

$$D_t = \delta(H_t, H_{t-1}) \tag{2}$$

where δ can be one of the following:

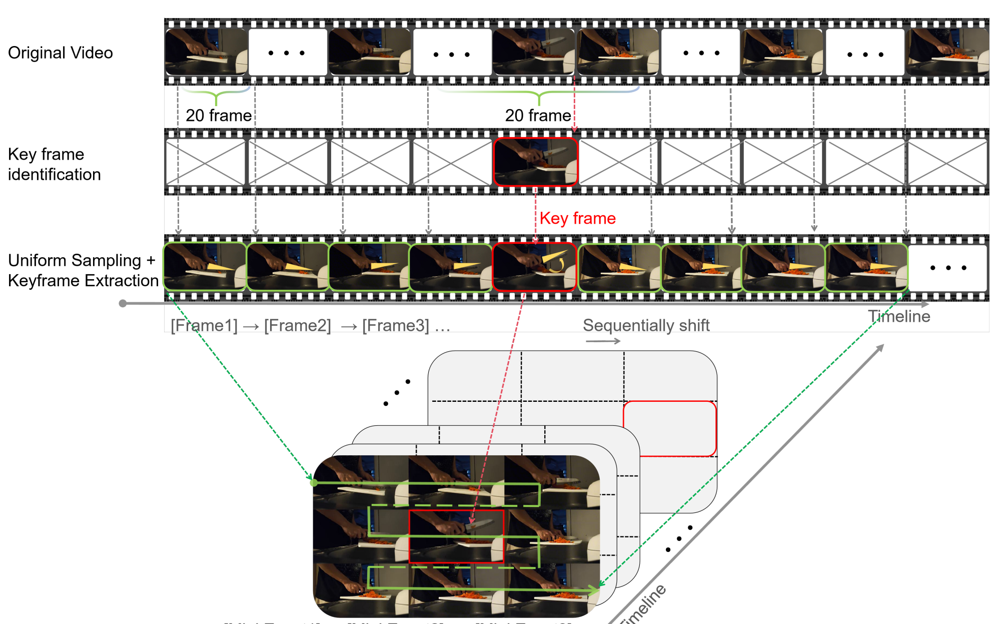

Fig. 3. Hybrid Frame Sampling and Keyframe Selection for Semantically Efficient Long-Video Representation

- **Chi-square distance**:

$$\chi^2(H_t, H_{t-1}) = \sum_i \frac{(H_t(i) - H_{t-1}(i))^2}{H_t(i) + H_{t-1}(i) + \varepsilon} \tag{3}$$

- **Kullback-Leibler divergence**:

$$D_{\mathrm{KL}}(H_t \parallel H_{t-1}) = \sum_i H_t(i) \log \left(\frac{H_t(i)}{H_{t-1}(i) + \varepsilon} \right) \tag{4}$$

A keyframe is selected if $D_t > \tau_h$, where τ_h is typically set to 0.3 and ε is a small constant (e.g., 10^{-8}) for numerical stability.

2.3 Cumulative Frame Change Detection. While abrupt changes are easily captured, slow transitions—such as gradual gestures or posture changes—are harder to detect when utilizing keyframes. To address this, we calculate cumulative variation over a sliding window. If small differences persist across several frames, the accumulated change will exceed a threshold, indicating a key moment, which ensures that important but non-dramatic transitions are not overlooked[105]. To detect slow-changing but meaningful segments, we define a cumulative change score over a sliding window of size W:

$$C_t = \sum_{i=t-W+1}^{t} \mathrm{Change}(I_i, I_{i-1}) \tag{5}$$

where $\mathrm{Change}(\cdot)$ may be either SSIM difference or histogram difference. A keyframe is selected if:

$$C_t > \tau_c \tag{6}$$

2.4. Final Keyframe Decision Rule. A frame I_t is marked as a keyframe if any one of the above conditions is satisfied[31]:

$$\mathrm{Keyframe}(I_t) = \begin{cases} 1, & \text{if } \mathrm{SSIM}(I_t, I_{t-1}) < \tau_s \\ 1, & \text{if } \delta(H_t, H_{t-1}) > \tau_h \\ 1, & \text{if } C_t > \tau_c \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

This multi-criteria decision rule ensures that both abrupt and subtle transitions are preserved, improving the temporal and emotional richness of video representations fed into the LLMs while maintaining manageable computational load. As shown in Figure 3, this approach successfully identified a previously omitted key moment: when a knife was raised and about to descend, which had been missed by uniform frame sampling. By recovering this frame as a keyframe, the system was able to complete the understanding of the "cutting" action sequence, enhancing the semantic continuity of the event.

3.1.2 Processing Audio Information from Video. For better understanding of the video content, we then separately processed the audio information of the uploaded video. We selected iFlytek's API [7] for its stability and comprehensive capabilities in speech-to-text conversion, since it allows speech recognition in noisy environments, speaker role differentiation, and emotional analysis.

[7] https://www.xfyun.cn/doc/asr/voicedictation/API.html

3.1.3 Performance Evaluation of the method. To quantitatively evaluate the effectiveness of our visual preprocessing strategies, we adopted a multiple-choice question-answering (MCQA) framework inspired by prior multimodal understanding benchmarks [35, 72]. We aimed to test whether multimodal large language models (LLMs) could accurately interpret our grid-based video summaries when supported by both visual and audio inputs.

1. QA Design. We selected 10 videos uploaded by users and manually annotated each with 10 multiple-choice questions. Each question offered 4 answer choices, only one of which was correct. This standardized format allowed for consistent scoring and comparison across different input formats.

2. Multimodal Inputs. Each question was answered by prompting the LLM with both the processed video frames (i.e., grid-based composite images) and the associated audio transcript. This setting allowed us to evaluate the model's ability to integrate information across visual and auditory modalities.

3. Question Types. The questions were categorized into three classes:

- *Perception:* Questions focused on object recognition, environmental understanding, or identifying human actions.
- *Reasoning:* Questions requiring logical inference, such as temporal ordering or causal reasoning based on observed sequences.
- *Information Synopsis:* Questions that required summarizing or synthesizing key information across the full video context.

A concrete example of the question set is provided in Appendix A.3.

Video FPS	Processing Method Variants					
	Full Input	Frame Every 20 Frames	+ Keyframes	+ 3×3 Grid	+ Keyframes + 3×3 Grid	+ Keyframes + 4×4 Grid
1080p24	Exceeded API processing limit / None	48s	57s	42s	48s	31s
1080p30	Exceeded API processing limit / None	55s	68s	43s	49s	31s
1080p60	Exceeded API processing limit / None	103s	125s	58s	63s	32s

Table 1. Processing times for different video frame rates and visual abstraction methods. Full video input exceeded API processing limits at all tested frame rates. Under identical resolution and frame rate conditions (1080p24, 1080p30, and 1080p60), the keyframes + 4×4 grid configuration resulted in the shortest processing times (31–32 seconds), and the keyframes + 3×3 grid configuration showed slightly longer times (48–63 seconds).

4. Evaluation Metrics. We computed the accuracy of model responses by comparing them to the annotated ground truth answers.

We evaluated five frame sampling and compression strategies by comparing model responses to annotated ground truth answers across 120 multiple-choice questions derived from 10 videos.

As shown in Table 2, the highest accuracy (95%) was achieved by the combination of keyframes and a 3×3 grid. This configuration proved to be the most effective, striking an optimal balance

Evaluation Metric	**Processing Method Variants**					
	Full Input	**Frame Every 20 Frames**	**+ Keyframes**	**+ 3×3 Grid**	**+ Keyframes + 3×3 Grid**	**+ Keyframes + 4×4 Grid**
10 Videos, 120 Questions (rounded)	Exceeded API limit or None	76%	94%	80%	**95%**	88%

Table 2. Accuracy across different frame sampling and compression methods (120 evaluation questions). The highest accuracy (95%) was achieved using keyframes with a 3×3 grid. While the 4×4 grid with keyframes was faster, its accuracy dropped to 88%, indicating a trade-off between efficiency and comprehension quality.

between semantic retention and token efficiency. It preserved critical temporal and contextual information, enabling the model to perform more accurate reasoning over long-form visual input.

In contrast, while the keyframes with 4×4 grid variant reduced processing time to approximately 31–32 seconds (Table 1), its accuracy dropped to 88%, indicating a significant loss in comprehension quality. Thus, despite its advantage in low-latency scenarios, the 4×4 method sacrifices semantic depth.

Overall, the grid-based visual abstraction—especially when enhanced with keyframes—not only improved inference efficiency but also enabled the model to scale from short clips (1–2 minutes) to extended videos of up to 30 minutes. The integration of key frames further enhanced the model's ability to comprehend transitions, interactions, and critical behaviors, making the system robust against temporal sparsity and improving its inference quality in multi-modal scenarios.

3.2 Design workshops to define user experience

While our initial prototype of Introspectus AI featured an efficient and reliable architecture for video analysis with a simple text base interface which could be used for querying, the interaction experience was not specifically geared towards facilitating self-reflection practices. To this end we conducted a series of design workshops involving four experts who helped us to design how Introspectus AI would present information to the user, the level of details included in personalised responses to various queries, as well as defining an optimal layout for the application.

Method. Experts joined the design workshop individually in a 1-1 session with the first author Workshops were conducted online and lasted approximately one hour. Sessions were recorded with participants consent and the researcher took notes to highlight the key feedback points provided. Participants received a compensation equivalent to $10 for their time.

Participants and Procedure. We recruited 4 participants (two males and two females; mean = 26.5 years old, SD =1.118) to take part in the design workshops. Among the participants, two were researchers in relevant fields, one was a practitioner in the domain of AI, and one was a researcher in the medical field. All participants had prior experience using LLMs in their daily academic, work, and personal activities.

At the start of the session participants were asked a series of introductory questions focused primarily on their habits related to daily reflection, as well as their expectations and perceptions of involving LLMs in the reflective process. Experts were then introduced to the core functionalities of the Introspectus AI. The participants then tested the implemented features, and gave feedback on the system's functionality, effectiveness, and practical requirements for supporting self-reflection. The workshop guide and the questions asked to expert are included for clarity in the Appendix (see B). The video from the design workshops were analysed alongside the researcher notes using affinity diagrams to identify key feedback and user experience requirements highlighted by the experts [76].

Findings. Based on the analysis of data from the design workshops, we finalized the system's configuration across four key areas. These categories were derived through qualitative affinity diagramming of expert feedback, which clustered their suggestions into recurring themes. The process involved iterative coding and grouping of transcripts and researcher notes, revealing shared expectations regarding the chatbot's interaction style, content depth, identity framing, and platform usability (Representative examples are summarized in Appendix (see B.2)):

- Chatbot Tone and Communication Style: The language style of Introspectus AI was defined as warm and friendly.
- Response Content: The AI's feedback will be set to focus on provide a detailed and accurate review of events, including guiding questions and logical solutions.
- AI Identity: The identity of the AI was designed as the user's self at a specific point in the past.
- Platform Design: The platform was established as a lightweight, easily accessible website.

3.3 System Improvement and Final Prototype of Introspectus AI

3.3.1 Integration of multimodal information. Based on experts' feedback, we improved our prompts by asking GPT-4o to summarize existing information into key categories before generating feedback. As shown in Figure 4, the prompt structure ensured GPT-4o to focus on: (1) events and scenes, (2) user actions, (3) positive and negative user behaviors, and (4) user interactions. Here, we present example prompts that were used to guide GPT-4o in analyzing the recorded video (see complete

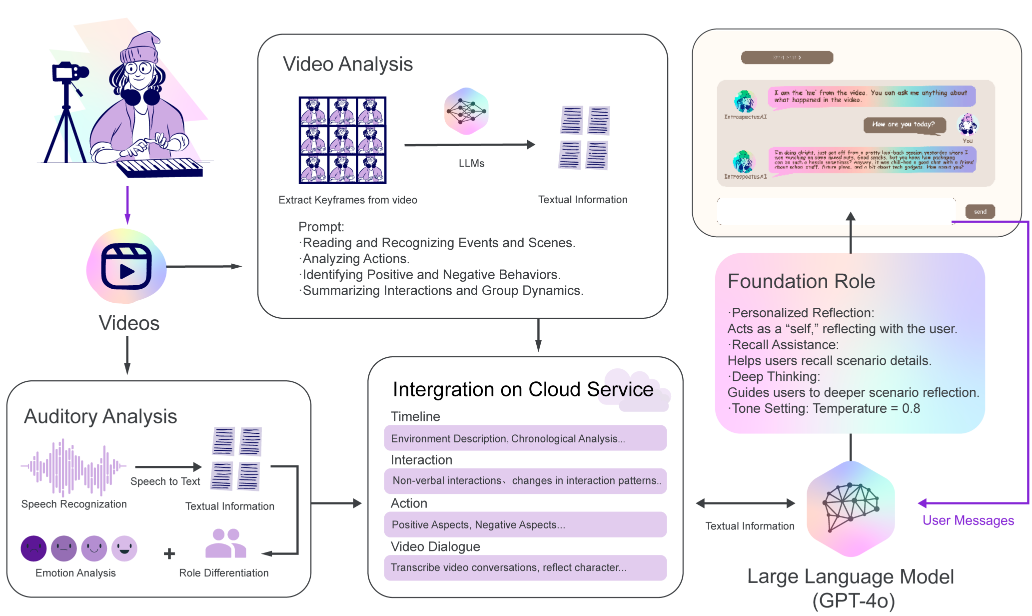

Fig. 4. An illustration of the Introspectus AI system

prompt in the Appendix A). To ensure LLM's memory retention, prevent hallucinations, and maintain coherence during interactions, we categorized and labelled all obtained information.

(1)Scene Description and Event Chronology: Extract the environmental setting and timeline of events from visual summaries. Prompt Sample:*"The environment and setting (if indoor): describe the room, furniture, table arrangements, and any objects present (e.g., notebooks, dishes). Include lighting, colors, and background details. Maintain chronological order to reflect the timeline accurately."*
This ensures the model infers the correct flow of time and spatial context from image grids.

(2)User Action Interpretation: Identify and sequence user behaviors—both physical and social. Prompt Sample:*"Focus on what people are doing (e.g., pouring drinks, passing food, taking notes). If my actions changed during the interaction, provide a sequential breakdown, including subtle actions such as nodding or avoiding eye contact."*
This module helps highlight what the user did in the provided video and how actions evolved.

(3)Behavioral Valence Evaluation: Distinguish helpful or problematic behaviors from a reflective standpoint. Prompt Sample:*"Identify any potential mistakes, misunderstandings, or negative impacts caused by my actions. Highlight actions that could be considered helpful, polite, or constructive."*
This supports value-based reflection and the identification of growth opportunities.

(4)Social Interaction and Group Dynamics: Assess interpersonal dynamics through non-verbal and conversational cues. Prompt Sample:*"Focus on non-verbal cues: facial expressions, eye contact, gestures, and posture. Identify if the interaction was collaborative, tense, or neutral. Mention how my behavior may have influenced others' responses."*
This segment helps users understand the potential social-emotional impacts of their actions.

3.3.2 Generating Personalized Responses. We then prompted the foundation role of the chatbot using ChatGPT-4o (temperature = 0.8). The additional settings for the chatbot include: (1) acting as an independent "self," generating personalized responses to the content of the recorded scenario and engaging in collaborative reflection with the user; (2) comprehensively assisting the user in recalling details related to the scenario in response to their queries; and (3) proactively guiding the user to think more deeply about the scenario (see Appendix A.1.1 for the complete prompts used in this process). We tested several temperature parameters and found that the temperature setting of 0.8 can balance providing sufficient suggestions to users while maintaining a warm tone.

3.4 System Integration

Finally, we integrated the entire Introspectus AI system into a web platform using Wix[8] for the front-end interface as shown in Fig 5, which allows users to upload recorded first-person life video clips directly through the website, as illustrated in Fig 4. Upon uploading, the front-end automatically transferred the video data to the back-end system hosted on Amazon Web Services [9].

The back-end managed all core processing stages including the data pre-processing, summarizing visual information via GPT-4o, and processing audio(codec='libmp3lame', bitrate='64k') using iFlytek's API and performing role separation('duration': '200', 'roleType': '1'). The back-end system imposed constraints on video uploads, limiting the duration to between 5 and 25 minutes and the file size to under 150 MB. Depending on the video size, processing times ranged from 3 to 8 minutes. After processing, the information was compiled and returned to the front-end as the final prompt material. Users could initiate interaction with the material by clicking the "Start chat" button. Once the prompt material is received, it directly connects to GPT-4o (temperature=0.8).

[8] https://ja.wix.com/
[9] https://aws.amazon.com

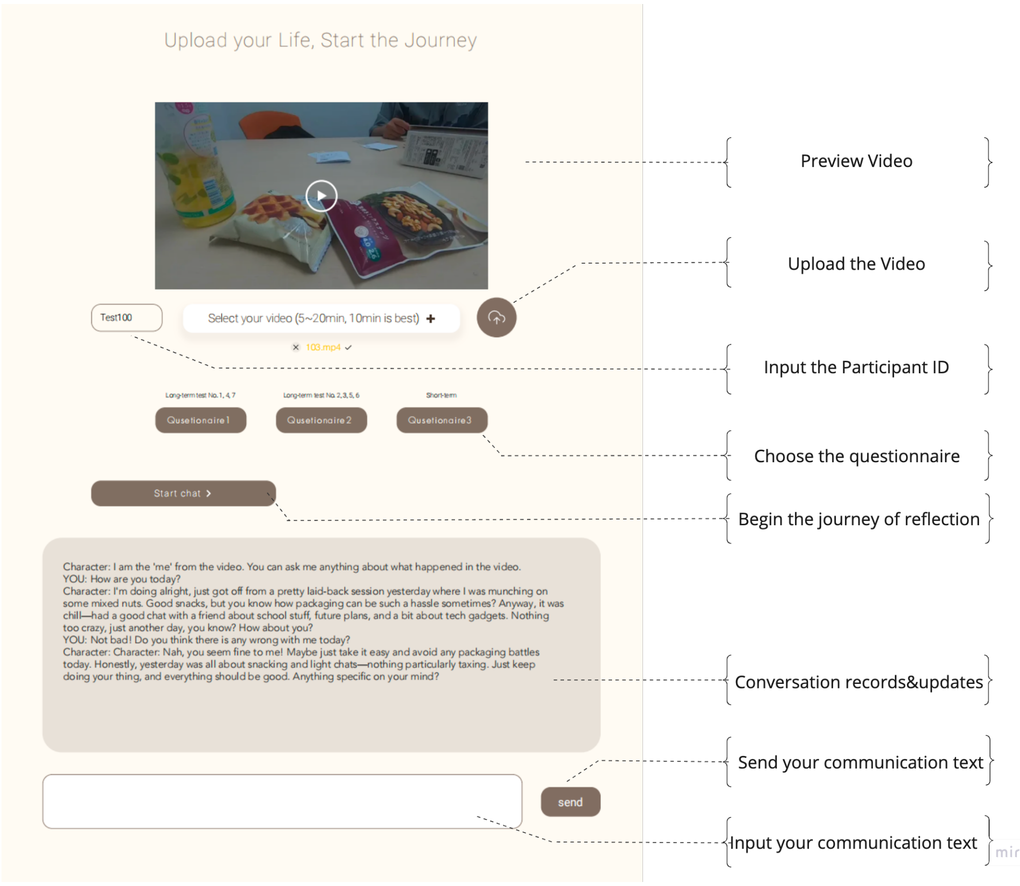

Fig. 5. The website of Introspectus AI

This setup enabled seamless communication with Introspectus AI through an embedded chatbox on the website.

4 User Study on the Long-term Impact of Introspectus AI

The study design was approved by our institutional review board, and we conducted this study between September 2024 and March 2025. To refine the interview script and recruitment strategy, we initially conducted a pilot study involving eight participants recruited from the authors' social networks. These participants represented both technical and non-technical backgrounds. Following each pilot interview, researchers reflected on the entire process, documented observations, and made iterative modifications to the interview script to improve its quality and efficacy.

Our research interest and objective were to investigate whether Introspectus AI can serve as an intervention or support tool to inspire individuals' reflective awareness or enhance their capacity for reflection, thereby fostering improved self-development. Additionally, we sought to explore factors influencing the acceptance and integration of GenAI in daily life. To ensure that any observed effects could be attributed to the AI-driven interaction itself rather than general reflective activity, we included a control group that engaged in a structured daily diary without using Introspectus AI.

4.1 Methods

Recent research demonstrates the short-term effectiveness of an AI-driven tool in enabling users to express emotions and promote self-reflection[32]. We are interested in understanding how interactions with Introspectus AI, as well as the duration of these interactions, impact users' reflective abilities, awareness of reflection, and acceptance of integrating GenAI systems into their daily lives. To evaluate these outcomes, we conducted a comparative study involving both an experimental group (who interacted with Introspectus AI) and a control group (who engaged in a structured daily diary). This design allowed us to assess how sustained interaction with the Introspectus AI impacts participants' reflective habits, behavior modification, and acceptance of generative AI in everyday life, as compared to traditional self-reflection methods.

4.1.1 Hypothesis. Here, we describe our hypotheses. *Hypothesis 1*: Repeated interactions with Introspectus AI over a long-term period (a) increase participants' engagement in reflective practices and (b) enhance the effectiveness of their self-reflective abilities.

Hypothesis 2: The use of Introspectus AI (a) increases participants' acceptance of generative AI applications in their daily lives and (b) long-term use does more significantly compared to initial session.

4.1.2 Questionnaires. We used the Self-Reflection and Insight Scale (SRIS) [39] and the Generative Artificial Intelligence Acceptance Scale (GAIAS) [115] to probe participants' changes in reflective attitude and generative AI acceptance throughout the 7-day period. Participants were asked to complete the SRIS before and after each interaction on a daily basis, and the GAIAS before and after the initial interventions, as well as after their 4th and 7th interventions. Participants in both the experimental and control groups completed the same SRIS and GAIAS questionnaires at equivalent time points to enable direct comparison across conditions.

Additionally, during the post-intervention interview, both groups were asked about the perceived depth and quality of their reflections. For the experimental group, we focused on their experiences interacting with Introspectus AI, including their perceptions of the AI's usefulness, relevance, and whether the feedback prompted behavioral or emotional insights. Example questions included: *"How has your experience been interacting with Introspectus AI?"*, *"What was your goal in using the system, or which behavior did you aim to reflect on?"*.

For the control group, the interview explored their experience writing a structured daily diary, challenges in maintaining the habit, the clarity of their reflections, and their openness to using AI-based tools for similar purposes in the future. Questions included: *"Was it easy to identify what to reflect on each day?"*, *"Do you think an AI assistant might support your reflection in the future?"*.

4.1.3 Recording first-person perspective videos. Considering individual differences in personal goal and daily activities, we did not restrict the themes of the recording. The only requirement was that participants should capture purposeful tasks, such as shopping, eating, or exercising, instead of simple repetitive activities like wandering or daydreaming. Based on our pilot study and subsequent experiments, we found that recording a main task generally took at least 5 minutes. Taking into account the completeness of information transmission, stability, and processing speed, participants were instructed to take videos of 5-25 minutes and with the file size under 150 MB for uploading. This video recording protocol was only applied to the experimental group, as the control group engaged in text-based reflection through a structured daily diary and did not collect or submit video recordings.

4.1.4 Communication with Introspectus AI(Experimental Group). Considering that the individual differences in users' baseline levels of self-reflection, extroversion or introversion, and prior experience with LLMs could lead to difficulties in sustaining conversations or initiating new topics with the Introspectus AI, we provided each participant with a reference guidebook containing potential questions to ask(see Appendix C). They could ask questions freely or refer to the guide for inspiration. Each interaction with Introspectus AI was required to last a minimum of five minutes, with no upper time limit. Here, we describe the five main categories of questions provided in the guidebook:(1)Scenario Recap and Understanding, (2)Areas for Improvement, (3)Positive Behaviors and Strengths, (4)Emotional Awareness and Management, (5)Explore Broader Reflection and Pattern.

4.1.5 Control Group Procedure and Reflection Guidelines. Participants in the control group were instructed to engage in self-reflection using a structured journaling method over the same one-week period. We provided the same five reflective dimensions used in the experimental group as daily prompts to guide their writing process: (1) what happened, (2) what was done well, (3) what could be improved, (4) how to improve it, and (5) how to maintain the habit of self-reflection. These suggestions served as a reference for participants when composing their daily entries.

Although participants in the control group did not interact with the Introspectus AI system, they were allowed to use general-purpose tools such as ChatGPT to help organize their thoughts or articulate their reflections.

4.2 Participants

We recruited 64 participants for the one-week study. Four participants were excluded due to incomplete data or discontinuation of participation midway. The final sample included 33 males and 27 females, with a mean age of 29.23 years (SD = 9.50). Participants were based in Japan, New Zealand, China, and Germany. Eligibility criteria required individuals to be at least 18 years old and proficient in either English, Japanese, or Chinese. The sample comprised a diverse demographic, including students, working professionals, and six retired individuals aged 50–60. Participants were randomly assigned to either the experimental or control group, with stratified sampling to ensure balanced distribution across age ranges, occupations, and gender. Each participant received a compensation equivalent to $40 upon completing the study.

4.3 Procedure

4.3.1 User Onboarding. As shown in Fig 6, participants provided informed consent before taking part in the follow-up investigation, Participants provided informed consent before taking part in the follow-up investigation, agreeing to participate in either the experimental group or control group and to allow all survey and interview data to be used anonymously. For participants in the experimental group, additional consent was obtained for uploading videos to the Introspectus AI system, and participants confirmed that all individuals appearing in their videos were aware of and agreed to the intended use. Participants then filled in the initial questionnaires including their demographic information, the SRIS and the GAIAS scales. The questionnaire also included subjective questions on familiarity with generative AI and engagement in self-reflective activities rated on 7-point Likert Scale.

4.3.2 Daily Intervention with Introspectus AI. Over the subsequent week, participants were required to use Introspectus AI at least seven times, without requiring daily use. On a daily basis, participants selected and recorded segments of their daily lives and then upload these recordings. The participants then interacted with Introspectus AI for at least 5 minutes following the guidebook or

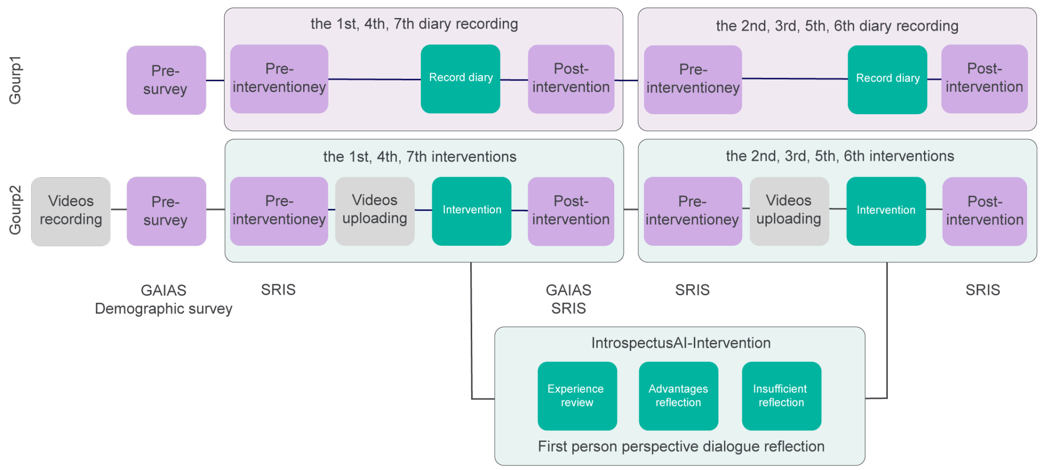

Fig. 6. Procedure of long-term user study: Group 1 served as the control group, completing structured daily diaries for reflection, while Group 2 was the experimental group, using Introspectus AI to engage in AI-assisted self-reflection.

based on their own interest. There was no upper time limit for each intervention. The participants filled in the SRIS before and after the daily intervention.

Additionally, after the fourth and seventh use of Introspectus AI as an intervention, participants were asked to complete the GAIAS scale to observe changes in their acceptance of GenAI applications over time.

Participants in the control group completed a structured daily diary instead, under the same weekly schedule and reflection frequency (see Section 4.1.5).

4.3.3 Exit Interview. At the end of the long-term study, each participant were invited to a 30-minute semi-structured interview regarding their overall experience. The interview protocol was adapted for the experimental and control groups to reflect their different modes of reflection (see Appendix D section for details).

4.4 Data Analysis

We used SPSS to analyze the quantitative data [10]. Regarding the qualitative analysis, all uploaded videos, chat logs, and interview data were anonymized and analyzed using inductive thematic analysis to gain deeper insights [14]. The primary researcher categorized and coded the chat logs and transcribed interview discussions into themes. These identified themes were then reviewed and reread by a secondary researcher, followed by a third review by other members of the research team, ensuring consistency.

5 RESULTS

Here, we describe our findings on the long-term use of Introspectus AI's impact on self-reflection and changes in acceptance of generative AI in daily use.

[10]https://www.ibm.com/jp-ja/products/spss-statistics

5.1 Impact on self-reflection

5.1.1 Quantitative findings. We compared changes in participants' self-reflection abilities using the SRIS score. Measurements were taken at four timepoints: pre-1st session (baseline), and post-1st, post-4th, and post-7th session. The analysis included two groups: an experimental group that engaged with Introspectus AI, and a control group that completed structured daily diaries.

Shapiro–Wilk tests were performed to check normality. When the normality assumption was met, repeated-measures ANOVA was conducted; otherwise, the non-parametric Friedman test was used. For each subscale, we analyzed within-group changes across timepoints and computed change scores from Day 1 to Day 7. To further investigate between-group differences in self-reflection-related outcomes, we conducted Mann–Whitney U tests on the difference scores (post-7th minus pre-1st session) for each questionnaire subscale.

The following results demonstrate divergent trajectories across the two groups, with the experimental group showing significantly pronounced improvements in reflective capacity over time.

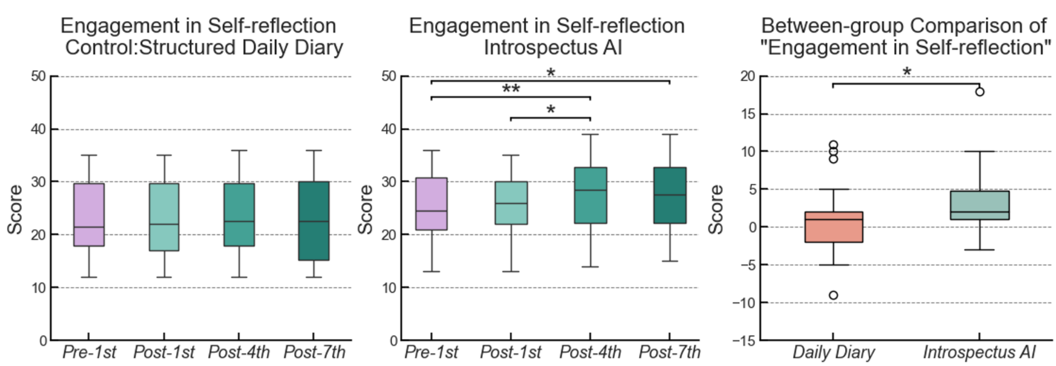

Fig. 7. Significant increases in self-reflection engagement were found in the experimental group (Introspectus AI) from pre-1st to post-4th (p = .004 **) and post-7th (p = .014 *), and from post-1st to post-4th (p = .049 *); no significant change was observed in the control group. The improvement from pre-1st to post-7th was also significantly greater in the experimental group than in the control group (p = .034 *). ($p < .05$, $p < .01$, $p < .001$; *, ***, *** respectively).

Engagement in Self-Reflection. As shown in Fig 7, for the experimental group using Introspectus AI, a repeated-measures ANOVA was conducted to assess changes in *Engagement in Self-Reflection* across four timepoints. Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(5) = 60.38$, $p < .001$; therefore, Greenhouse–Geisser correction was applied. The analysis revealed a significant main effect of time, $F(1.75, 99.78) = 7.93$, $p = .001$, $\eta^2 = .119$.

Bonferroni-corrected post-hoc comparisons indicated significant increases from pre-1st to post-4th (p = .004, **) and to post-7th (p = .014, *), as well as from post-1st to post-4th (p = .049, *).

For the control group (structured daily diary), the data violated normality assumptions. Therefore, a non-parametric Friedman test was used, which showed no significant change across timepoints, $\chi^2(13) = 8.978$, $p = .775$ (ns.).

The Introspectus AI group showed significantly greater improvement ($M_{\text{rank}} = 35.25$) than the control group ($M_{\text{rank}} = 25.75$), $U = 307.5$, $Z = -2.12$, $p = .034$.

These findings demonstrate that AI-supported reflection significantly increased participants' engagement in self-reflection, while journaling alone had no measurable effect over the same period.

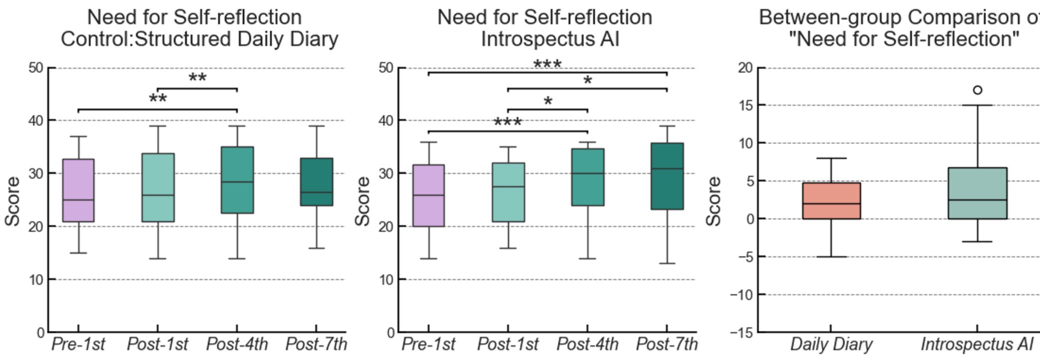

Fig. 8. Significant increases were observed in the experimental group (Introspectus AI) from pre-1st to post-4th ($p < .001$ ***) and post-7th ($p < .001$ ***), and from post-1st to post-4th ($p = .025$ *) and post-7th ($p = .031$ *). In the control group (Daily Diary), pre-1st to post-4th ($p = .003$ **) and post-1st to post-4th ($p = .029$ *) also showed significant increases. No between-group difference was found ($p = .216$). ($p < .05$, $p < .01$, $p < .001$; *, ***, *** respectively).

Need for Self-Reflection. As shown in Fig 8, for the experimental group using Introspectus AI, a non-parametric Friedman test was conducted to assess changes in *Need for Self-Reflection* across four timepoints. The analysis revealed a significant main effect of time, $\chi^2(3) = 36.29$, $p < .001$.

Bonferroni-corrected post-hoc comparisons indicated significant increases from pre-1st to post-4th ($p < .001$, ***) and to post-7th ($p < .001$, ***), as well as from post-1st to post-4th ($p = .025$, *) and to post-7th ($p = .031$, *).

For the control group (structured daily diary), a Friedman test also revealed a significant effect of time, $\chi^2(3) = 16.62$, $p = .001$. Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant increases from pre-1st to post-4th ($p = .0029$ **) and from post-1st to post-4th ($p = .029$ *), while all other comparisons were not significant.

No significant difference was found between groups, $U = 366.5$, $Z = -1.24$, $p = .216$.

These findings suggest that although both groups exhibited increases in their perceived need for self-reflection, the improvement in the AI-supported condition was more sustained and emerged earlier during the intervention period, but was not significantly greater in magnitude compared to the control group.

Insight. As shown in Fig 9, to evaluate how participants' insight evolved throughout the intervention, we compared *Insight* scores at four timepoints. In the experimental group, Shapiro–Wilk tests indicated violations of normality; thus, a non-parametric Friedman test was applied. The result revealed a significant main effect of time, $\chi^2(3) = 24.80$, $p < .001$ ***. Bonferroni-corrected pairwise comparisons showed significant increases from pre-1st to post-4th ($p = .002$ **) and to post-7th ($p < .001$ ***).

In contrast, the control group's Insight scores met the assumption of normality, and a repeated-measures ANOVA was used. The result showed a marginally significant main effect of time, $F(3, 87) = 2.60$, $p = .057$. Further pairwise comparisons indicated a significant increase from pre-1st

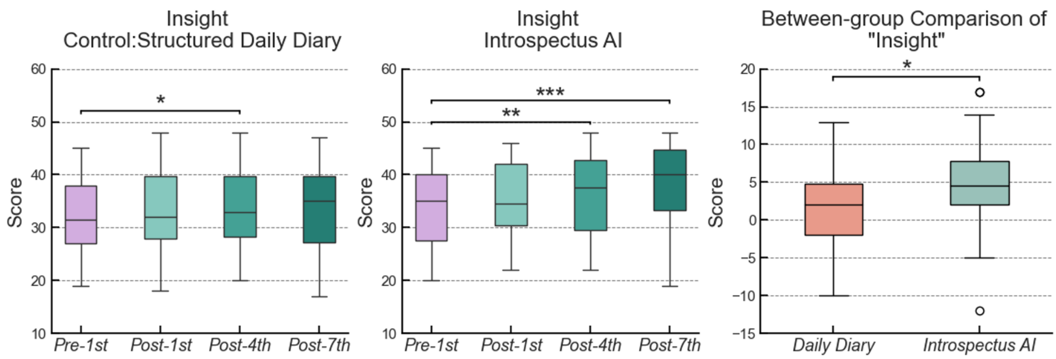

Fig. 9. Significant increases were observed in the Introspectus AI group from Pre-1st to Post-4th ($p = .002$ **) and to Post-7th ($p < .001$ ***). The control group showed only a temporary increase from pre-1st to post-4th ($p = .022$ *), with no sustained improvement. Between-group comparison revealed significantly higher insight gains in the AI group than in the control group ($p = .035$ *). ($p < .05$, $p < .01$, $p < .001$; *, ***, *** respectively).

to post-4th ($p = .022$ *, Bonferroni corrected), but no significant difference between pre-1st and post-7th ($p = .615$, ns.), suggesting that the increase in insight was temporary and not sustained by journaling alone.

The Introspectus AI group had significantly higher gains ($M_{\text{rank}} = 35.25$) than the control group ($M_{\text{rank}} = 25.75$), $U = 307.5$, $Z = -2.11$, $p = .035$.

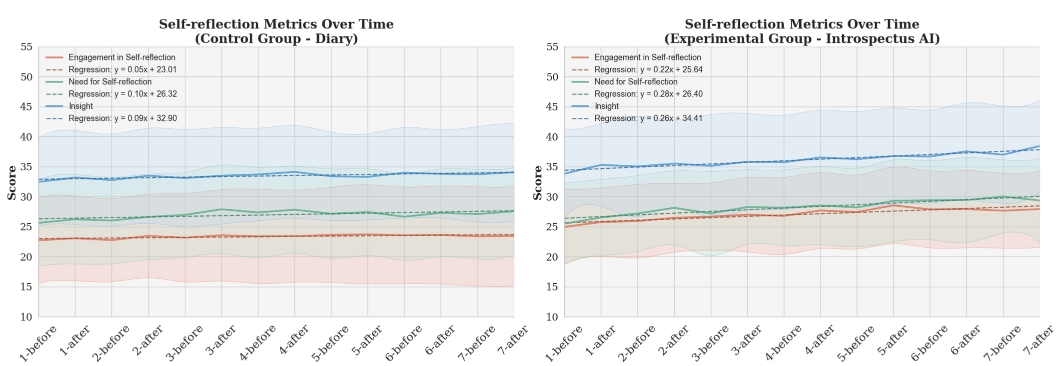

Fig. 10. Regression analyses across 7 sessions revealed linear increases in all SRIS subscales for the experimental group (Introspectus AI). Engagement followed the model $y = 0.22x + 25.64$ ($R^2 = 0.8118$, 95% CI = 0.22 ± 0.06), Need for Self-Reflection showed a steeper rise ($y = 0.28x + 26.40$, $R^2 = 0.8595$, CI = 0.28 ± 0.06), and Insight demonstrated the strongest fit ($y = 0.26x + 34.41$, $R^2 = 0.8849$, CI = 0.26 ± 0.05). By contrast, the control group exhibited weaker trends: Engagement ($y = 0.05x + 23.01$, $R^2 = 0.5133$), Need ($y = 0.10x + 26.32$, $R^2 = 0.4090$), and Insight ($y = 0.09x + 32.90$, $R^2 = 0.5931$) showed flatter slopes and lower model fits. These findings support greater and more consistent reflective gains in the AI-supported condition.

Regression Trends Across Time. : This trend is further illustrated in Fig. 10, which presents regression models of the 7-day progression for each SRIS subscale.

In the experimental group, Engagement in Self-Reflection followed a linear trajectory with the model $y = 0.22x + 25.64$, $R^2 = 0.8118$, and a 95% confidence interval for the slope of 0.22 ± 0.06.

Need for Self-Reflection showed a steeper increase, modeled by $y = 0.28x + 26.40$, $R^2 = 0.8595$, with a 95% CI of 0.28 ± 0.06. Insight exhibited the highest goodness of fit, with $y = 0.26x + 34.41$, $R^2 = 0.8849$, and a slope confidence interval of 0.26 ± 0.05.

In contrast, the control group showed weaker trends.

These regression models visually reinforce the earlier statistical findings, suggesting more consistent and accelerated improvement across all reflective dimensions in the AI-supported condition.

5.1.2 Qualitative findings. Our qualitative data further support the quantitative findings with the following themes being conceptualised.

Reflective positivity. Regarding the system's usage, we observed that most users initially displayed curiosity and an exploratory attitude towards the features of Introspectus AI rather than trying to use to for reflection. "*At first, I was curious about what AI could recognize, but over time, I started focusing more on reflection and summarizing my actions.*"(P5); "*I think my experience can be split into two phases. In the beginning, I'd ask more open-ended questions, and later, I mainly used it for reflection.*"(P12); "*At first, I just used it as a way to summarize situations. Later on, I realized it could help me understand my actions and suggest ways to improve.*"(P11),

Among older participants (aged 50 and above), the system also demonstrated positive influence. "*It made me constantly want to record a video and look back to see what new things I could find.*"(P22, 57 years old) P23 noted, "*I want to talk to it every day,*"(P23, 55 years old) and "*It made me start learning how to speak and behave all over again.*"(P24, 58 years old)

After becoming familiar with the system, users demonstrated a positive attitude towards engaging in reflective behavior. "*Yes, it's made me realize there are way more opportunities for self-reflection in life than I thought.*"(P2)"*After interacting with the AI, I reflect 30% more often, and my interest in self-reflection has increased by 30%-50%.*"(P3); "*It's really made me rethink behaviors I usually ignore.*"(P13); "*Using this AI system has indeed become an opportunity for reflection.*"(P16).

While encouraging a tendency towards self-reflection, several users also mentioned that interacting with Introspectus AI inspired them to consider making self-reflection a long-term habit. "*I've gradually gotten used to reflecting on my actions and summarizing each day. It's becoming a habit.*"(P1); "*Now I'm more willing to take a deeper look back and reflect on my life experiences.*"(P8). This stands in contrast to responses from the control group (P31–P60), who often perceived the act of journaling as a routine task. While a few participants acknowledged potential benefits: "I want to keep up the habit of journaling"(P41), many expressed more passive or even negative sentiments: "*I just wrote what I did every day. It didn't really change the way I think.*" (P35), "*It felt like homework, and during the last few days it became quite boring.*" (P49), and "*Although GPT offered some prompts, the process still required me to actively provide all the information, which made it feel less engaging.*" (P52).

Reflective ability. The significant increase in Insight was also reflected in our analysis of the interview. With the help of Introspectus AI, participants consistently demonstrated an improvement in the level and depth of their self-reflection: "*There's some improvement. The AI focuses more on details than self-reflection itself, helping me reexamine small habits*"(P15); "*Sometimes, I don't think too deeply in the moment, but it makes me want to do a summary afterward.*"(P11);

"*AI offers new perspectives, pointing out details I hadn't noticed.*"(P21). Moreover, during long-term usage, participants became more aware of the improvement in their reflective abilities, which encouraged continued use and created a positive feedback loop: "*...I found AI really helped me notice things I hadn't paid attention to before. Now, I'm happy to keep using it.*"(P8); "*In long-term use, it points out things I didn't notice, which is really nice.*"(P21). This trend was also observed among older participants. "*It made me think more carefully about how I speak and behave, which I never did before.*" (P24, 58 years old).In contrast, older adults in the control group (P31–P60) generally reported limited reflection. For example, P60 (57 years old) commented, "*Even recalling what happened is difficult for me, let alone going further.*"Most diary participants emphasized recounting events rather than analyzing them: "*I just wrote down what happened; it felt more like a memory exercise.*" (P37). Among the few who found it useful, the main benefit was the introduction of reflective thinking itself: "*This was the first time I seriously reviewed what I did each day—journaling gave me a new perspective.*" (P51).

Impact on Behavioral Changes. Beyond enhancing reflective thinking, we found that long-term use of Introspectus AI had a notably positive impact on participants' behavior and habits.

Most participants indicated that the interaction had a positive impact on their behavior."*In presentations, it suggested me to cut down on hand gestures and focus more on eye contact.*"(P17); "*.....It influenced me a bit—I might go for a healthier salad instead of fries next time, maybe add some nuts or protein.*"(P1); "*The AI pointed out how overly focused I was while gaming and reminded me to keep an eye on time management.*"(P12); ""*It suggested me to focus on eating and cut down on multitasking. That advice was helpful.*"(P7); Older participants also reported behavior-related impacts: "*It taught me how to pose when taking photos, which I never thought about before.*" (P24, 57 years old). A few participants even reported that this positive impact translated into their real-life behavior within the week. ""*The AI suggested me to tidy up my lab desk, and I actually did it.*"(P6); "*......(This experience) helped me be more patient, especially in board games where I'm more inclined now to wait for others to make their move, and I noticed myself doing this the next day when we played.*"(P10); "*(When practicing mindfulness meditation) the AI's tips advised me to relax and shift my mindset, which helped me get into the right state more effectively in my next attempt.*"(P3) In contrast, most participants in the control group (P31–P60) reported minimal behavioral change: "*There wasn't much change. Even if I reviewed my behavior in writing, I couldn't tell whether I should change anything.*" (P49); "*Maybe behavioral change would require more specific guidance.*" (P51).

5.2 Acceptance of GenAI applications in daily life

5.2.1 Quantitative findings. We compared changes in participants' acceptance of generative AI applications using the GAIAS (Generative AI Acceptance Scale), which includes four subscales: Performance Expectancy, Effort Expectancy, Facilitating Conditions, and Social Influence. Measurements were taken at pre-1st session (Baseline) and three post-intervention timepoints: post-1st session (Day1-after), post-4th session (Day4-after), and post-7th session (Day7-after).

The analysis included two groups: an experimental group that engaged with Introspectus AI, and a control group that completed structured daily diaries. The control group used a guided diary template for daily reflection, and were allowed to optionally consult general-purpose generative AI tools (e.g., ChatGPT) for brainstorming or phrasing support. Shapiro–Wilk tests revealed violations of the normality assumption for several items across both groups. Therefore, we used the non-parametric Friedman test to analyze within-group changes over time, and Mann–Whitney U tests to compare 7-day change scores between groups. The following results examine each group separately, followed by a between-group comparison.

As shown in Fig 11, we conducted Friedman tests to examine within-group changes in participants' acceptance of generative AI across four timepoints (pre-1st post-1st, post-4th, post-7th). Significant improvements were observed across all four GAIAS subscales.

Performance Expectancy increased significantly, $\chi^2(3) = 28.810$, $p < .001$. Post-hoc comparisons revealed significant improvements from pre-1st to post-4th ($p = .016$ *), and from pre-1st to post-7th ($p < .001$ **).

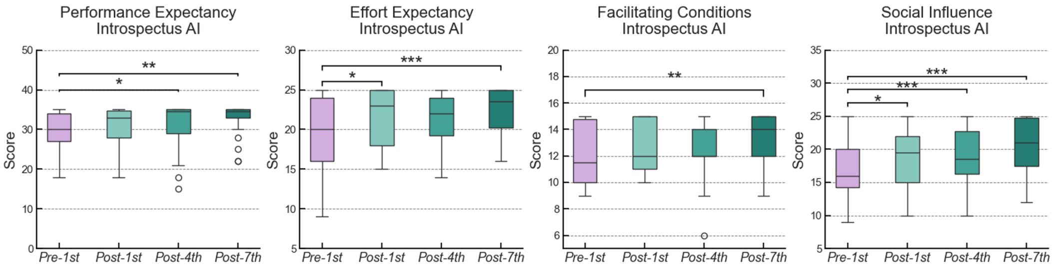

Fig. 11. Significant increases were observed in the Introspectus AI group from Pre-1st to Post-4th (p = .016 *) and to Post-7th (p < .001 **) on Performance Expectancy. Effort Expectancy also improved significantly from Pre-1st to Post-1st (p = .026 *) and to Post-7th (p < .001 ***). Facilitating Conditions showed a moderate but significant increase from Pre-1st to Post-7th (p = .007 **). Social Influence exhibited the most robust increase, with significant gains from Pre-1st to Post-1st (p = .012 *), Post-4th (p < .001 ***), and Post-7th (p < .001 ***). (p < .05, p < .01, p < .001; *, ***, *** respectively).

Effort Expectancy also showed a significant increase, $\chi^2(3) = 25.747$, $p < .001$, with significant pairwise differences between pre-1st and post-1st (p = .026 *) and between pre-1st and post-7th (p < .001 ***).

Facilitating Conditions demonstrated a moderate but significant improvement, $\chi^2(3) = 14.301$, $p = .003$. The comparison between pre-1st and post-7th reached significance after correction (p = .007 **).

Social Influence showed the most robust upward trend, $\chi^2(3) = 44.189$, $p < .001$, with significant pairwise increases between pre-1st and post-1st (p = .012 *), pre-1st and post-4th (p < .001 ***), and pre-1st and post-7th (p < .001 ***).

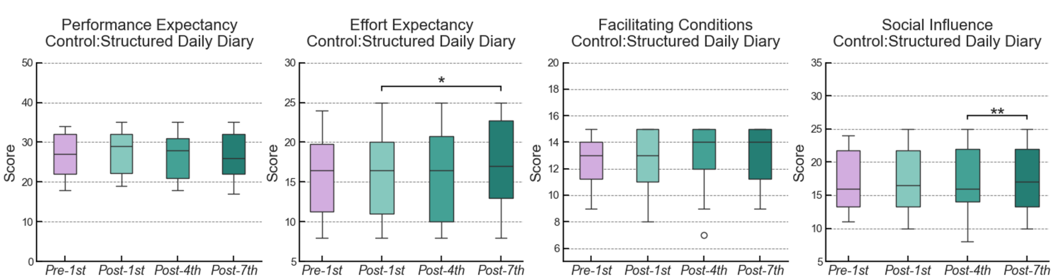

Fig. 12. A significant increase was observed in the control group from Post-1st to Post-7th (p = .014 *) on Effort Expectancy. Social Influence also improved significantly from Post-4th to Post-7th (p = .008 **). No significant improvements were observed for Performance Expectancy (p = .155) or Facilitating Conditions (p = .049; no pairwise differences remained significant after adjustment). (p < .05, p < .01, p < .001; *, ***, *** respectively).

As shown in Fig 12, we also conducted Friedman tests for the control group to assess changes in participants' acceptance of generative AI across three post-intervention timepoints. No significant differences were observed in most GAIAS subscales.

For *Effort Expectancy*, the test yielded $\chi^2(3) = 13.221$, $p = .004$. A significant increase was observed between post-1st and post-7th (p = .014 *), but not between pre-1st and post-7th (p = .121, ns.). For *Social Influence*, the overall test was also significant: $\chi^2(3) = 12.418$, $p = .006$. Post-hoc comparisons revealed a significant difference between post-4th and post-7th (p = .008 *); however, the change from pre-1st to post-7th was not significant (p = .093, ns.), indicating limited long-term impact. *Performance Expectancy* and *Facilitating Conditions* did not show reliable improvement over time.

Taken together, these within-group results suggest that while participants who interacted with Introspectus AI became increasingly confident in the usefulness, usability, and social viability of generative AI in daily contexts, the diary group showed only limited or inconsistent gains across these dimensions.

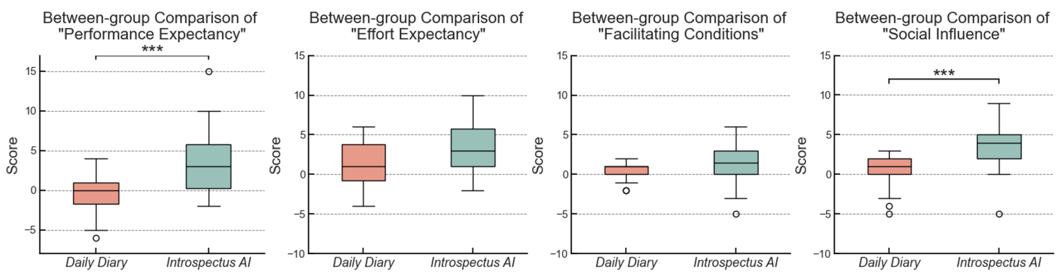

Fig. 13. Between-group comparison of acceptance of generative AI applications after 7-day use. Significant differences were observed in Performance Expectancy (p < .001 ***) and Social Influence (p < .001 ***), with the Introspectus AI group showing greater improvement than the control group. No significant differences were found in Effort Expectancy (p = .064) and Facilitating Conditions (p = .072). (p < .05, p < .01, p < .001; *, ***, *** respectively).

Between-group Comparison. As shown in Fig 13, to further evaluate how Introspectus AI influenced participants' long-term acceptance of generative AI applications compared to the control group, we conducted Mann–Whitney U tests on the change scores (post-7th minus pre-1st session) across four GAIAS subscales.

Performance Expectancy. The Introspectus AI group showed significantly greater improvement (M_{rank} = 39.13) than the control group (M_{rank} = 21.87), U = 191.0, Z = −3.86, p < .001.

Effort Expectancy. The difference between groups was not statistically significant, U = 325.5, Z = −1.85, p = .064.

Facilitating Conditions. No significant difference was found between groups, U = 331.0, Z = −1.80, p = .072.

Social Influence. The Introspectus AI group demonstrated significantly greater increases (M_{rank} = 40.50) than the control group (M_{rank} = 20.50), U = 150.0, Z = −4.47, p < .001.

These results indicate that repeated interaction with the generative AI system significantly enhanced users' perceived usefulness, ease of use, and social receptiveness toward AI-assisted reflection, with most effects becoming stronger over time.

The qualitative analysis of interview results allows us to contextualize the conclusions of the quantitative study and helps us explore how GenAI applications like Introspectus AI can be integrated into daily life.

5.2.2 *Qualitative Findings.* Our qualitative findings allowed us to gain deeper insights around the specific reasons that determined participants acceptance of GenAI application for self-reflection.

Perceived Value and Benefits. An important component of AI acceptance is the user's performance expectations of the application's functionality. This aligns with the analysis of "Performance

Expectancy" discussed in Section 5.2.1. A majority of participants expressed surprise and significant interest in the functionality: "*I think it was a very novel experience*"(P4); *"I didn't know that large language models were capable of this kind of technology now! I wonder how you made it happen!*"(P9); Besides, participants provided highly positive evaluations of the effectiveness of Introspectus AI, especially in the analysis of scenario details, personal behaviors, and emotions: "*I think it(Analysis of the scenario) matches really well, which surprised me a lot.*"(P3); "*Its summary of daily activities is quite accurate*"(P14); "*The AI captures the overall picture well.*"(P21); "*No misunderstandings so far, and it accurately sums up what I'm focused on.*"(P20) Middle-aged participants also showed surprise and excitement toward Introspectus AI: "*I only recently started learning about large language models, and this made me feel once again how fast the world is changing.*" (P24, 57 years old). Under

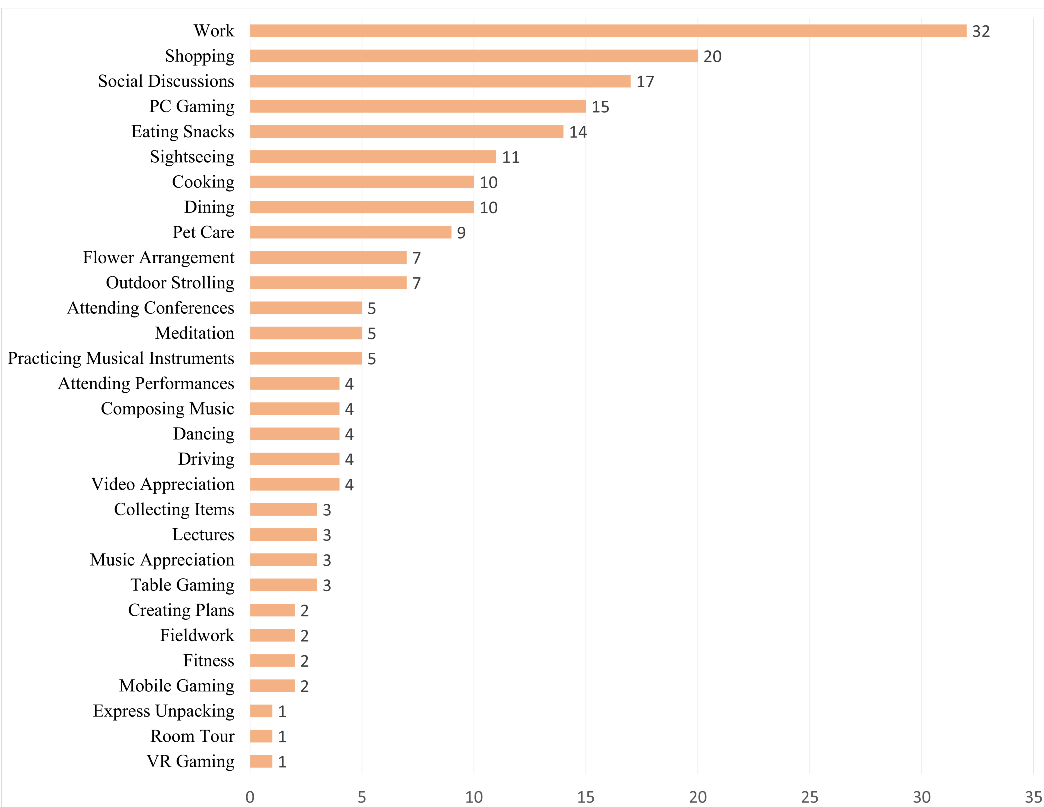

Fig. 14. A summary of the type of activities participants chose to self-reflect.

this premise, users tend to proactively use it in various scenarios in their daily lives. As shown in Fig 14, the uploaded videos included a total of 30 different events, ranging from common activities like work (32 times), shopping (20 times), and eating (Snacks 14 times, Dining 10 times) to more personalized choices such as playing VR game (1 time), practicing musical instruments (5 times),flower arrangement(7 times, same person) and dancing (4 times) This indicates that trust in and expectations of the functionality have contributed to the acceptance of GenAI in everyday scenarios.

Ease of Integration into Daily Life. We also analyzed the clarity and simplicity of Introspectus AI for users, as well as its impact on usage and the potential for long-term integration of GenAI into daily life. Participants provided positive feedback regarding the clarity and ease of use of this system:"*I didn't encounter any difficulties while using it.*"(P8);"*There wasn't anything particularly difficult, it was pretty clear.*"(P13); Interestingly, the system also received favorable responses from middle-aged and older participants. Despite potential concerns about digital literacy, these users found Introspectus AI to be accessible and straightforward. One older participant noted, "*Even though I don't usually use many apps, this was really easy to follow.*" (P21). Another commented, "*I didn't feel overwhelmed. The steps were clear, and I appreciated that it didn't require logging in or installing anything complicated.*" (P25). Their responses further support the idea that the system's simplicity reduces technical barriers, even for those less familiar with newer technologies. "*The system is simple and super easy to understand how to use it.*"(P15); The simplicity of the system and its clear instructions allowed users to quickly become proficient with Introspectus AI, even enabling them to personalize their experience after only a few uses. The limitations users faced during their experience were also evident. To ensure the system was lightweight, participants were required to prepare their own recording equipment, leading nearly all of them to use their smartphones. However, recording in first-person perspective with a phone inevitably caused some inconvenience. When we asked participants about their willingness for long-term use, while the majority of participants demonstrated a positive attitude, they also highlighted how self-video recording could become cumbersome: "*Yes, I'd enjoy using it as a tool to transcribe videos or keep records.*"(P16); "*I'd probably use it if it were reasonably priced or even free.*"(P7); "*I'm interested in keeping use it.*"(P9). "*If the recording device were easier to use, like smart glasses, I'd be more willing.*"(P6), "*If it doesn't rely on video uploads, I'd be open to it.*"(P13) This indirectly supports the findings from the analysis in Section 5.2.1, where "Effort expectancy" of GenAI acceptance showed a significant increase with long-term use, "Facilitating Conditions" showed significant improvement only in the latter half of the study, as participants became more familiar with the system and gradually learned to manage usability challenges.

Social Acceptance and Influence. The current acceptance of generative AI usage in everyday life is essentially a societal issue. Due to the misuse of generative AI for activities like copyright infringement and online fraud, there is ongoing public apprehension about its application. Although the usage was only a user test where we ensured all individuals featured in the uploaded videos were informed and gave their consent, and we guaranteed no misuse of uploaded data, we analyzed users' willingness to engage with Introspectus AI, particularly focusing on the prospect of long-term use and any privacy concerns they may have regarding uploading personal videos for processing by generative A, a small number of participants indicated unconditional acceptance: "*I'm not really worried about privacy. I feel fairly confident about the protection in place.*"(P10); *I'm personally not too worried.*"(P7). However, the vast majority indicated that they would be willing to use it for an extended period in the future if security conditions could be imposed to safeguard information: *"I feel fairly comfortable, but I'd like to see some encryption or legal measures to help users trust it more.*"(P3), *"If the videos being uploaded are filtered first, I think it's safe.*"(P5), *"I do have privacy concerns. I hope future updates add settings like letting users choose what to upload or block certain scenes.*"(P12). Interview responses also highlighted that participants' comfort with the system was strongly influenced by the explicit privacy protections and content restrictions communicated during the study: *"I trust your team so I don't worry*"(P2); *"I'm not too worried because it doesn't require showing my face and I believe the information is well protected.*"(P8). Notably, older participants also expressed high levels of acceptance and trust in the system. As one participant stated, "*As long as I know what data I uploaded, I feel it's acceptable.*" (P24, 57 years old). Another shared, "*I believe that* *if this technology exists, someone must be responsible for keeping it safe—so I feel at ease.*" (P23, 55 years old). These comments suggest that transparency about data usage and belief in responsible oversight can be key factors in fostering generative AI acceptance across age groups. This highlights one of the strengths of Introspectus AI—its ability to foster trust and acceptance even among older users, by maintaining transparency and lowering the cognitive barrier to participation.

This indicates that the use of Introspectus AI, while ensuring functional settings, has effectively provided users with a sense of security regarding their personal information. This confidence in privacy measures has contributed to participants' increased acceptance of Introspectus AI, and generative AI in general, for long-term use.

6 DISCUSSION

6.1 Summary of the finding

Our study on the usage of IntrospectusAI reveals its impact on enhancing self-reflective abilities and influencing the acceptance of generative AI through long-term use. Regarding reflection, significant changes primarily manifested under long-term conditions. In particular, after long-term use, the need for self-reflection, engagement in the reflection process, and insight all showed notable improvements. These findings suggest that personalized reviews of experiences supported by generative AI can be a valuable tool for deepening self-awareness and promoting reflective practices [6, 104]. Introspectus AI demonstrated significant positive effects on enhancing self-reflection abilities during long-term interventions. However, most of our participants stated that they did not already regularly engage in self-reflection habits. While this indicates that Introspectus AI might be a suitable tool that support people's ability to engage in self-reflection practice and build new habits, those who already have establish routine practices with whom they are satisfied, might not find the tool as useful, or even less effective than the strategies they already employ. Regarding the acceptance of generative AI, we also observed a positive trend among participants in the experimental condition. Although significant changes were present from the initial day, the effects were more pronounced towards the end of the study, indicating a stronger impact compared to short-term use. While significant changes following the initial use were less prominent, these findings nonetheless underscore the importance of a long-term intervention to achieve the desired outcomes. Participants' attitudes toward facilitating conditions showed positive effects, but these were not statistically significant. This outcome may be due to challenges in the video recording process, such as issues with recording equipment and event selection. Despite the easy-to-use interface, these factors highlight the need for a more integrated application and improved user experience.

6.2 Implications on Generative AI in Enhancing Self-Capabilities and Long-term Integration

Our research contributes to the literature on the impact of generative AI on self-capabilities. Beyond the increasingly mature instrumental applications, generative AI's high analytical capabilities and personalized interventions can effectively influence human introspection, guiding users to better understand their consciousness and behavior, promoting better decision-making, and fostering positive behavioral change [21, 69, 69, 109]. Our study emphasizes the long-term potential of such interventions. Simply put, we extend the advantages of generative AI by leveraging its powerful summarization and inference capabilities to facilitate interaction between individuals. In contrast to prior work in the CSCW community that has examined users' tendencies to redact or modify their disclosures to chatbots[27, 68], this study employs long-form videos as an intervention to support the ongoing development of users' own reflective experiences and to enhance the objectivity of their self-observation. The unique feature of Introspectus AI lies in providing individuals with a low-cost way to "experience" this conceptual functionality, even as long video analysis is still being perfected. Building on existing discussions about the short-term effects of generative AI on self-capabilities [32, 94, 104], we extend the potential for long-term impacts. We demonstrate the long-term interventions of Introspectus AI on self-reflection capabilities, highlighting the significance of long-term interventions in shaping the future impact of generative AI on human self-development [77]. This could serve as a reference for how to use generative AI for therapeutic applications and well-being promotion in the future.

6.3 Implications for Long-Term Adoption of Generative AI in Daily Life

The effectiveness of generative AI for long-term use hinges on sustained user acceptance of its applications. Therefore, in this study, we also discuss the factors and prerequisites influencing the long-term use of generative AI. Overall, user trust is a prerequisite for generative AI applications to become integrated into daily life and to facilitate long-term use, which aligns with prior work in the CSCW community on trust and adoption of AI systems [42, 49, 67]. These studies emphasize transparency, fairness, and explainability as key to long-term acceptance. User trust can be divided into several dimensions. Beyond ensuring the "outstanding functionality" and "simple, low-barrier interaction" commonly recognized, our study also emphasizes the importance of transparent data processing policies and effective privacy protection measures [19, 54]. The primary barrier preventing generative AI from becoming a part of everyday life lies in societal trust crises or rejection, often caused by issues like malicious deepfake scams and the inability to secure copyright for creative materials such as images and music in the generative AI context [3, 9, 41]. In our experimental discussions, we further emphasized data limitations, privacy assurances, subsequent processing measures, and timely feedback on related issues. Against this backdrop, through quantitative and qualitative analysis, we found a significant decrease in distrust during user interaction, and even observed an increase in participants' perception of social acceptance of generative AI within their social circles over long-term use. However, we did not entirely eliminate trust concerns, as many participants still expressed reservations about long-term use of such applications, particularly outside a research context where there is knowledge of specific ethical standards. We demonstrate the elements of Introspectus AI, highlighting how these elements can have positive or negative impacts on the acceptance of generative AI in both short-term and long-term scenarios. This serves as an important reference for the further development of generative AI and for its potential to become a transformative technology in society.

6.4 System Sustainability: Adaptability and Data Growth Management

Although our deployment was temporally bounded, the system was conceptually designed to support long-term, co-adaptive use [4], where both users and agents evolve through interaction. This view echoes recent CSCW research on human–AI collaboration [17, 38], which emphasizes the importance of transparency, task understanding, and reflective tools in enabling human users to meaningfully co-adapt with AI systems. As participants engaged in reflective interactions, they began to articulate new expectations—such as the ability to revisit earlier sessions or resume unfinished dialogues with their past selves. To address these emerging demands, we introduced a user-specific indexing mechanism in which each participant was assigned a persistent identifier linked to a personalized data repository. This structure enabled individuals to access previously uploaded videos and reactivate prior conversations, thereby facilitating continuity across reflective episodes.

In parallel, we implemented a series of backend optimizations to enhance system performance under sustained use, including asynchronous video processing and improved server concurrency.

These refinements contributed to maintaining responsiveness and scalability as user-generated content accumulated.

To further support the long-term integration of Introspectus AI, we adopted a hybrid memory strategy combining episodic summarization with lightweight context preservation. Instead of storing entire video archives or dialogue logs, the system retains abstracted behavior summaries and reflective cues, enabling continuity while minimizing storage overhead. Additionally, we introduced a periodic memory compression process that distills interaction history into high-level insights or behavioral trajectories—consistent with memory abstraction models in reflective technologies [26].These developments expand on prior CSCW work by supporting co-adaptation not only through system transparency or shared task goals, but also through open-ended, personally meaningful reflection grounded in everyday lived experience.

7 ETHICAL CONSIDERATIONS

Introspectus AI undoubtedly raises privacy concerns with its use of personalized audio and video recordings. However, this issue is not unique to our system, as any generative AI that processes user uploaded information containing personal voices, images, or videos could potentially misuse such data for forgery or identity theft and cause issues such as reputational damage or financial loss [3, 9, 70]. This is particularly true for systems that generate hyper-realistic versions of individuals. To address these issues, numerous studies have focused on audio, image, and video generation tampering detection methods, as well as digital watermarking techniques that prevent information from being used for generative AI training [41, 121]. In our system, while we demonstrated the benefits of using AI to review and reflect on video content, we remained aware of risks such as system is manipulation for malicious purposes, which could result in Introspectus AI providing false information or misleading guidance. Given that application such as Introspectus AI involve reflecting on one's own experiences, there is a risk that users may develop absolute trust in the feedback, making them susceptible to malicious influence. Therefore, before starting the experiment, we repeatedly informed participants that the AI system's feedback was not authoritative and should be questioned if it seems incorrect. Maintaining one's judgment and critical thinking during the process of receiving feedback is also an essential part of enhancing self-reflection. Additionally, participants were informed prior to recording to avoid including any private information in the video and to ensure that all individuals appearing in the video were aware of its intended use. After the experiment, all data, including videos and logs, were be removed from the cloud and stored only temporarily on local devices for experimental records, after which they have been permanently deleted to prevent any potential privacy breaches due to server intrusions.

8 LIMITATION AND FUTURE WORK

Video processing duration. Although we selected a server plan with relatively high computational power and bandwidth, and minimized the computational time complexity, the processes of video slicing, processing, audio extraction, and feedback from OpenAI and iFlytek APIs still took a considerable amount of time. Additionally, to ensure server stability, we had to add redundancy and concurrency measures to prevent crashes caused by high processing loads. As a result, while managing to keep the processing time for a single video to around five minutes was considered manageable for the study, it could still have led to negative user experiences due to the long waiting time.

Restricted Emotion Parameters for Chatbot. We set the temperature parameter of the chatbot to 0.8 to provide gentle and friendly suggestions to users. However, we later found that not everyone prefererd gentle advice—some felt that overly gentle suggestions lacked persuasiveness and that even slight counterarguments easily made the chatbot retract its previous stance. In future studies, we will consider making the temperature parameter adjustable or dynamically changing it based on the user's preferences.

Recognition accuracy. While our system achieves practical efficiency in visual data processing, it still faces limitations in recognition accuracy. Initially, we employed equidistant frame slicing to reduce computational load. Later we integrated a keyframe extraction technique to supplement the equidistant frames. By identifying and including visually salient or transitional frames, the system could preserve important temporal cues and improve semantic coverage of the video content. Nevertheless, it is still difficult to fully capture nuanced behaviors or emotions conveyed through micro-expressions or fine-grained movement. Balancing computational efficiency with recognition precision remains an ongoing challenge.

Recording Devices. We identified certain limitations in the recording devices for long-term use. Participants often needed to capture video while engaging in daily activities. Since some solely used smartphones for documention, portability of the device presented a considerable challenge. To minimize interference with users' natural activities, we plan to provide more suitable recording devices in the future to allow smoothier interactions.

Bridging First-Person Experiences. Although Introspectus AI dedicates to enable individual self-reflection and personal growth, the same tool could be shared by multiple individuals, provided all consent to reciprocal video analysis and reflection. This could potentially foster new types of perspective-taking activities between friends, families, trainers and trainees, or even between owner and their pet. Such implementation could support mutual understanding and dialogue mediated by AI supported video analysis.

Scenario Preferences and Adaptive Design for Future Use. Introspectus AI was designed for open-ended daily reflection, however, users predominantly engaged with it in task-oriented scenarios including working, cooking, or skill practice (Fig. 14). These settings provided clear structure, making AI feedback easier to interpret and apply. In future iterations, the system could benefit from adaptive mechanisms that align with preferred usage contexts, offering tailored prompts and feedback strategies for structured, high-engagement scenarios. Such adaptations may enhance the effectiveness and long-term appeal of video-based self-reflection.

Long-Term Deployment and Scalability. While our study demonstrates high-frequency use over a one-week period, to support real-world use and future deployment beyond academic settings, realizing long-term integration requires addressing behavioral drift, engagement fatigue, and evolving user needs. Although features like user-specific indexing and episodic summarization were implemented, they remain untested in for longer-term use. Future versions should adopt adaptive interaction strategies, such as adjusting tone or question complexity based on usage history and introduce re-engagement mechanisms like milestone-based feedback. As a web-based system, Introspectus AI is device-accessible, but variations in hardware and usage context may affect experience quality. To ensure sustainable deployment, the system must support secure data management through privacy-preserving summarization, temporal compression, and user-controlled retention, along with scalable strategies for model adaptation and performance stability. These considerations offer a roadmap toward future deployment of Introspectus AI in real-life environments such as coaching, education, and wellness applications. To support this broader applicability, we are working towards the creation of an open GitHub repository which will be accessible to researchers and coders interested in furthering its development.

9 CONCLUSION

We present Introspectus AI, a generative AI system designed to enhance self-reflection and support behavior change. By leveraging multimodal information from everyday recordings, it provides personalized feedback on user behavior. Our study examines the impact of long-term interactions with Introspectus AI, focusing on its potential benefits in inducing participants' reflective practices, self-awareness, and generative AI acceptance. The results indicate that AI-supported interventions significantly improve engagement in self-reflection, the need for reflection, and insight. Additionally, user acceptance of generative AI tools increased gradually during use.

We then discussed the broader challenge of using generative AI products effectively in everyday contexts, especially against the backdrop of trust crisis often associated with such technologies. Ultimately, we envision Introspectus AI as an accessible, adaptable tool that supports reflective decision-making and promotes ongoing self-improvement, integrating generative AI into daily life as an indispensable part of personal reflection and behavioral growth.

Acknowledgments

This work is supported by JST COI-NEXT "Minds1020Lab" Project(Grant Number JPMJPF2203), JST SPRING(Grant number JPMJSP2123) and University of Auckland Faculty of Science Research Development Fund (Grant number 3731533)

References

[1] Eleni Adamopoulou and Lefteris Moussiades. 2020. An Overview of Chatbot Technology. In *Artificial Intelligence Applications and Innovations*, Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis (Eds.). Springer International Publishing, Cham, 373–383.

[2] Nahdatul Akma, Mohamad Hafiz, Azaliza Zainal, Muhammad Fairuz, and Zuraidy Adnan. 2018. Review of Chatbots Design Techniques. *International Journal of Computer Applications* (2018). https://api.semanticscholar.org/CorpusID:70077174

[3] Ivan A. Aleshkovski. 2022. Social Risks and Negative Consequences of Diffusion of Artificial Intelligence Technologies. *ISTORIYA* (2022). https://api.semanticscholar.org/CorpusID:251891429

[4] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *AI magazine* 35, 4 (2014), 105–120.

[5] Rob Anderson and Kenneth N. Cissna. 2008. Fresh Perspectives in Dialogue Theory. *Communication Theory* 18 (2008), 1–4. https://api.semanticscholar.org/CorpusID:142601019

[6] Riku Arakawa and Hiromu Yakura. 2020. INWARD: A computer-supported tool for video-reflection improves efficiency and effectiveness in executive coaching. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[7] Riku Arakawa and Hiromu Yakura. 2024. Coaching Copilot: Blended Form of an LLM-Powered Chatbot and a Human Coach to Effectively Support Self-Reflection for Leadership Growth. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*. 1–14.

[8] Mr.DHANUSH B. 2024. CHATBOT USING LARGE LANGUAGE MODEL. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT* (2024). https://api.semanticscholar.org/CorpusID:269835077

[9] Mojtaba Barari, Lars-Erik Casper Ferm, Sara Quach, Park Thaichon, and Liem Ngo. 2024. The dark side of artificial intelligence in marketing: meta-analytics review. *Marketing Intelligence & Planning* (2024). https://api.semanticscholar.org/CorpusID:270293530

[10] James Bennett-Levy and Nicole K. Lee. 2014. Self-practice and self-reflection in cognitive behaviour therapy training: what factors influence trainees' engagement and experience of benefit? *Behavioural and cognitive psychotherapy* 42 1 (2014), 48–64. https://api.semanticscholar.org/CorpusID:9351245

[11] James Bennett-Levy, Frederick H. Turner, Taryn Beaty, Michelle Smith, Beth Paterson, and Sarah Farmer. 2001. THE VALUE OF SELF-PRACTICE OF COGNITIVE THERAPY TECHNIQUES AND SELF-REFLECTION IN THE TRAINING OF COGNITIVE THERAPISTS. *Behavioural and Cognitive Psychotherapy* 29 (2001), 203 – 220. https://api.semanticscholar.org/CorpusID:38570549

[12] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting reflection in hci: Four design resources for technologies that support reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[13] Vic Boyd, Nikki N Woods, Arno K. Kumagai, Anne A. Kawamura, Angela Orsino, and Stella L. Ng. 2022. Examining the Impact of Dialogic Learning on Critically Reflective Practice. *Academic Medicine* 97 (2022), S71 – S79. https://api.semanticscholar.org/CorpusID:251495520

[14] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[15] James Bridgeman and Andrea Giraldez-Hayes. 2024. Using artificial intelligence-enhanced video feedback for reflective practice in coach development: benefits and potential drawbacks. *Coaching: An International Journal of Theory, Research and Practice* 17, 1 (2024), 32–49.

[16] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[17] Ángel Alexander Cabrera, Adam Perer, and Jason I Hong. 2023. Improving human-AI collaboration with descriptions of AI behavior. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–21.

[18] Jack Cahn. 2017. CHATBOT: Architecture, design, & development. *University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science* (2017).

[19] Gerard Canal, Rita Borgo, Andrew Coles, Archie Drake, Dong Huynh, Perry Keller, Senka Krivic, Paul Luff, Quratul ain Mahesar, Luc Moreau, S. Parsons, Menisha Patel, and Elizabeth Sklar. 2020. Building Trust in Human-Machine Partnerships. *Comput. Law Secur. Rev.* 39 (2020), 105489. https://api.semanticscholar.org/CorpusID:227924372

[20] Mirko Casu, Sergio Triscari, Sebastiano Battiato, Luca Guarnera, and Pasquale Caponnetto. 2024. AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. *Applied Sciences* (2024). https://api.semanticscholar.org/CorpusID:271018624

[21] Si Chen, Haocong Cheng, and Yun Huang. 2024. *Emotion Recognition in Self-Regulated Learning: Advancing Metacognition Through AI-Assisted Reflections.* Springer Nature Switzerland, Cham, 185–212. https://doi.org/10.1007/978-3-031-64487-0_9

[22] Sifan Chen, Danyang Peng, Giulia Barbareschi, Dunya Chen, and Chihiro Sato. 2025. A Reflective Journey for Individuals with Intellectual Disabilities: Rediscovering Competency Through AI-Enhanced Iron Beads Crafting. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 583, 18 pages. https://doi.org/10.1145/3706598.3713984

[23] Gary Cheng and Juliana Chau and. 2009. Digital video for fostering self-reflection in an ePortfolio environment. *Learning, Media and Technology* 34, 4 (2009), 337–350. https://doi.org/10.1080/17439880903338614 arXiv:https://doi.org/10.1080/17439880903338614

[24] Janghee Cho, Tian Xu, Abigail Zimmermann-Niefield, and Stephen Voida. 2022. Reflection in theory and reflection in practice: an exploration of the gaps in reflection support among personal informatics apps. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.

[25] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 173–182.

[26] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1143–1152.

[27] Yichao Cui, Yu-Jen Lee, Jack Jamieson, Naomi Yamashita, and Yi-Chieh Lee. 2024. Exploring Effects of Chatbot's Interpretation and Self-disclosure on Mental Illness Stigma. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 52 (April 2024), 33 pages. https://doi.org/10.1145/3637329

[28] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL] https://arxiv.org/abs/1810.04805

[29] Simon D'Alfonso. 2020. AI in mental health. *Current Opinion in Psychology* 36 (2020), 112–117. https://doi.org/10.1016/j.copsyc.2020.04.005 Cyberpsychology.

[30] Vanessa Echeverria, Linxuan Zhao, Riordan Alfredo, Mikaela Milesi, Yuequiao Jin, Sophie Abel, Jie Fan, Lixiang Yan, Xinyu Li, Samantha Dix, et al. 2025. TeamVision: An AI-powered Learning Analytics System for Supporting Reflection in Team-based Healthcare Simulation. *arXiv preprint arXiv:2501.09930* (2025).

[31] Naveed Ejaz, Irfan Mehmood, and Sung Wook Baik. 2013. Efficient visual attention based framework for extracting key frames from videos. *Signal Processing: Image Communication* 28, 1 (2013), 34–44.

[32] Cathy Mengying Fang, Phoebe Chua, Samantha Chan, Joanne Leong, Andria Bao, and Pattie Maes. 2024. Leveraging AI-Generated Emotional Self-Voice to Nudge People towards their Ideal Selves. *arXiv preprint arXiv:2409.11531* (2024).

[33] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the new world of HCI. *Interactions* 24 (2017), 38 – 42. https://api.semanticscholar.org/CorpusID:207646037

[34] Christine Frith, Sallie M Cowan, and Clare Delaney. 2015. Reflection training as a form of professional development for physiotherapy clinical educators. *Focus on health professional education : a multi-disciplinary journal* 16 (2015), 88–91. https://api.semanticscholar.org/CorpusID:70481877

[35] Chaoyou Fu, Yuhan Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, et al. 2024. Video-mme: The first-ever comprehensive evaluation benchmark of multi-modal llms in video analysis. *arXiv preprint arXiv:2405.21075* (2024).

[36] Dr. V. Geetha, Dr.C K Gomathy, Mr.D.Sri Datta Vallab Yaratha Yagn, and Sai Praneesh. 2023. THE ROLE OF NATURAL LANGUAGE PROCESSING. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT* (2023). https://api.semanticscholar.org/CorpusID:265638856

[37] Adam Gerace, Andrew Day, Sharon Casey, and Philip Mohr. 2017. 'I Think, You Think': Understanding the Importance of Self-Reflection to the Taking of Another Person's Perspective. *Journal of Relationships Research* 8 (2017), e9. https://doi.org/10.1017/jrr.2017.8

[38] Katrin Glinka and Claudia Müller-Birn. 2023. Critical-Reflective Human-AI Collaboration: Exploring Computational Tools for Art Historical Image Retrieval. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33.

[39] Anthony M Grant, John Franklin, and Peter Langford. 2002. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: an international journal* 30, 8 (2002), 821–835.

[40] Igor Grossmann, Anna Dorfman, Harrison Oakes, Henri C Santos, Kathleen D Vohs, and Abigail A Scholer. 2021. Training for wisdom: The distanced-self-reflection diary method. *Psychological Science* 32, 3 (2021), 381–394.

[41] Parisa Hassanzadeh and Robert E. Tillman. 2022. Generative Models with Information-Theoretic Protection Against Membership Inference Attacks. *ArXiv* abs/2206.00071 (2022). https://api.semanticscholar.org/CorpusID:249240443

[42] Allyson I Hauptman, Wen Duan, and Nathan J Mcneese. 2022. The components of trust for collaborating with ai colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 72–75.

[43] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. 2018. Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173989

[44] Xufei Huang and Professor Mitch Marcus. 2021. CHATBOT: DESIGN, ARCHITECUTRE, AND APPLICATIONS. https://api.semanticscholar.org/CorpusID:239014231

[45] Ioanna Iacovides, Joe Cutting, Jen Beeston, Marta E Cecchinato, Elisa D Mekler, and Paul Cairns. 2022. Close but not too close: Distance and relevance in designing games for reflection. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–24.

[46] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. 2020. Privacy concerns in chatbot interactions. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3*. Springer, 34–48.

[47] Claus D. Jacobs and Loizos Heracleous. 2005. Answers for questions to come: reflective dialogue as an enabler of strategic innovation. *Journal of Organizational Change Management* 18 (2005), 338–352. https://api.semanticscholar.org/CorpusID:144942665

[48] Adrianna C. Jenkins and Jason P. Mitchell. 2011. Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience* 6 (2011), 211 – 218. https://api.semanticscholar.org/CorpusID:30255862

[49] Jun Li Jeung and Janet Yi-Ching Huang. 2023. Correct me if I am wrong: exploring how AI outputs affect user perception and trust. In *Companion publication of the 2023 conference on computer supported cooperative work and social computing*. 323–327.

[50] Camille S. Johnson. 2010. Reflection Versus Self-Reflection: Sources of Self-Esteem Boost Determine Behavioral Outcomes. *Cognitive Social Science eJournal* (2010). https://api.semanticscholar.org/CorpusID:145721799

[51] Patrik Jonell. 2019. Using social and physiological signals for user adaptation in conversational agents : Doctoral consortium. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* :. 2420–2422. http://aamas2019.encs.concordia.ca/ QC 20200324.

[52] Amit Kapoor and Vishal Verma. 2024. EMOTION AI: UNDERSTANDING EMOTIONS THROUGH ARTIFICIAL INTELLIGENCE. *International Journal of Engineering Science and Humanities* (2024). https://api.semanticscholar.org/CorpusID:269895972

[53] Pelin Karaturhan, Ecem Arıkan, Pelin Durak, Asim Evren Yantac, and Kemal Kuscu. 2022. Combining momentary and retrospective self-reflection in a mobile photo-based journaling application. In *Nordic Human-Computer Interaction Conference*. 1–12.

[54] Arnav Kartikeya. 2021. Examining correlation between trust and transparency with explainable artificial intelligence. In *Sai*. https://api.semanticscholar.org/CorpusID:236965949

[55] Dr. (CS) Bhavesh A. Kinkhabwala. 2020. Can Meditation Practices Be Elevated, for the Higher Level of Consciousness, Taking Help of Artificial Intelligence? *PRN: Indian Philosophy (Topic)* (2020). https://api.semanticscholar.org/CorpusID:237530086

[56] Kamilla Klefbeck. 2023. Navigating Self-Reflection for Aspiring Special Education Teachers: A Scoping Review on Inclusive Educational Practices and Their Insights for Autism Education. *Education Sciences* (2023). https://api.semanticscholar.org/CorpusID:265442083

[57] STANLEY B Klein. 2014. Evolution, memory, and the role of self-referent recall in planning for the future. *What is adaptive about adaptive memory* (2014), 11–34.

[58] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for workplace reflection: a chat and voice-based conversational agent. In *Proceedings of the 2018 designing interactive systems conference*. 881–894.

[59] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: a conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–26.

[60] Siu Cheung Kong. 2010. Using a web-enabled video system to support student-teachers' self-reflection in teaching practice. *Comput. Educ.* 55 (2010), 1772–1782. https://api.semanticscholar.org/CorpusID:11071048

[61] Martijn JL Kors, Erik D Van der Spek, Gabriele Ferri, and Ben AM Schouten. 2018. You; the Observer, Partaker or Victim. Delineating Three Perspectives to Empathic Engagement in Persuasive Games using Immersive Technologies.. In *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts*. 493–501.

[62] Ethan Kross, Madeline Ong, and Ozlem Ayduk. 2023. Self-Reflection at Work: Why It Matters and How to Harness Its Potential and Avoid Its Pitfalls. *Annual Review of Organizational Psychology and Organizational Behavior* (2023). https://api.semanticscholar.org/CorpusID:256224839

[63] Ramakrishna Kumar and Maha Mahmoud Ali. 2020. A review on chatbot design and implementation techniques. *Int. J. Eng. Technol* 7, 11 (2020), 2791–2800. https://api.semanticscholar.org/CorpusID:237502044

[64] Rama Moorthy Senthil Kumar and Maha Mahmoud Ali. 2020. A Review on Chatbot Design and Implementation Techniques. https://api.semanticscholar.org/CorpusID:237502044

[65] Suh Keong Kwon, Dongkwang Shin, and Yongsang Lee. 2023. The application of chatbot as an L2 writing practice tool. (2023). https://hdl.handle.net/10125/73541

[66] Anton-Rupert Laireiter and Ulrike Willutzki. 2003. Self-reflection and self-practice in training of Cognitive Behaviour Therapy: an overview. *Clinical Psychology & Psychotherapy* 10 (2003), 19–30. https://api.semanticscholar.org/CorpusID:143397388

[67] Tina B Lassiter. 2024. The Role of Algorithmic Audits and Other Soft Law Approaches in Informing Users' Calibrated Trust in Artificial Intelligence Tools. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. 54–56.

[68] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.

[69] Yoon Kyung Lee, Yongha Park, and Sowon Hahn. 2023. A Portrait of Emotion: Empowering Self-Expression through AI-Generated Art. *ArXiv* abs/2304.13324 (2023). https://api.semanticscholar.org/CorpusID:258331575

[70] Shengyin Li, Vibekananda Dutta, Xin He, and Takafumi Matsumaru. 2022. Deep learning based one-class detection system for fake faces generated by GAN network. *Sensors* 22, 20 (2022), 7767.

[71] Yunxin Li, Xinyu Chen, Baotain Hu, and Min Zhang. 2024. Llms meet long video: Advancing long video comprehension with an interactive visual adapter in llms. *arXiv preprint arXiv:2402.13546* (2024).

[72] Yifei Li, Junbo Niu, Ziyang Miao, Chunjiang Ge, Yuanhang Zhou, Qihao He, Xiaoyi Dong, Haodong Duan, Shuangrui Ding, Rui Qian, et al. 2025. OVO-Bench: How Far is Your Video-LLMs from Real-World Online Video Understanding? *arXiv preprint arXiv:2501.05510* (2025).

[73] Elizabeth D Liddy. 2001. Natural language processing. (2001).

[74] Eva Liesa, Paula Mayoral, Mireia Giralt-Romeu, and Salvador Angulo. 2023. Video-based feedback for collaborative reflection among mentors, university tutors and students. *Education Sciences* 13, 9 (2023), 879.

[75] Jiaying Liu, Yiheng Su, Praneel Seth, et al. 2025. Can Large Language Models Grasp Concepts in Visual Content? A Case Study on YouTube Shorts about Depression. *arXiv preprint arXiv:2503.05109* (2025).

[76] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*. Springer, 231–248.

[77] Afra Jahanbakhsh Mashhadi, Akhil Mathur, Marc Van den Broeck, Geert Vanderhulst, and Fahim Kawsar. 2016. Understanding the impact of personal feedback on face-to-face interactions in the workplace. *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (2016). https://api.semanticscholar.org/CorpusID:6426965
[78] Steven Montague, Aditi Khokhar, Brandy DeWeese, Jeananne Elkins, Morkel Jacques Otto, Orlando Ortiz, and Camila Marques. 2023. A Needs Assessment of Educator Self-reflection in Health Sciences Education. *Journal of Advances in Medicine and Medical Research* (2023). https://api.semanticscholar.org/CorpusID:265343820
[79] Joseph M. Moran, William M. Kelley, and Todd F. Heatherton. 2013. What Can the Organization of the Brain's Default Mode Network Tell us About Self-Knowledge? *Frontiers in Human Neuroscience* 7 (2013). https://api.semanticscholar.org/CorpusID:17450348
[80] Peter Nguyen, Javier Fdez, and Olaf Witkowski. 2024. AI-Driven Meditation: Personalization for Inner Peace. In *Artificial Intelligence in Music, Sound, Art and Design*, Colin Johnson, Sérgio M. Rebelo, and Iria Santos (Eds.). Springer Nature Switzerland, Cham, 296–310.
[81] Alandeom W. Oliveira, Adam Oliver Brown, Marissa Carroll, Elizabeth Blenkarn, Bradley Ellis Austin, and Tiffany Bretzlaff. 2021. Developing undergraduate student oral science communication through video reflection. *International Journal of Science Education, Part B* 11 (2021), 143 – 154. https://api.semanticscholar.org/CorpusID:233616400
[82] Kenneth Olmstead. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones. (2017). https://www.pewresearch.org/short-reads/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/
[83] Regina CHI'20: CHI Conference on Human Factors in Computing Systems Honolulu H. Bernhaupt. 2020. *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery.
[84] Anna V. Osipova, Brooke Prichard, Alison G. Boardman, Mary Theresa Kiely, and Patricia Elaine Carroll. 2011. Refocusing the Lens: Enhancing Elementary Special Education Reading Instruction through Video Self–Reflection. *Learning Disabilities Research & Practice* 26 (2011), 158 – 171. https://api.semanticscholar.org/CorpusID:146749997
[85] Eleftherios Papachristos, Dorte P Meldgaard, Iben R Thomsen, and Mikael B Skov. 2021. ReflectPal: exploring self-reflection on collaborative activities using voice assistants. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part IV 18*. Springer, 187–208.
[86] Sharmila Reddy Pappula and Sathwik Rao Allam. 2023. LLMs for Conversational AI: Enhancing Chatbots and Virtual Assistants. *International Journal of Research Publication and Reviews* (2023). https://api.semanticscholar.org/CorpusID:266461220
[87] Nandini Pasumarthy, Shreyas Nisal, Jessica Danaher, Elise van den Hoven, and Rohit Ashok Khot. 2024. Go-Go Biome: Evaluation of a Casual Game for Gut Health Engagement and Reflection. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
[88] N. Pavitha, Priyanka Bhatele, Sharmishta Suhas Desai, and Himangi Pande. 2022. Design and Implementation of Multipurpose Chatbot. *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (2022), 1332–1337. https://api.semanticscholar.org/CorpusID:247108652
[89] Jan Prasko, Petr Mozny, Miroslav Novotny, Milos Slepecky, and Jana Vyskocilova. 2012. Self-reflection in cognitive behavioural therapy and supervision. *Biomedical Papers of the Medical Faculty of Palacky University in Olomouc* 156, 4 (2012).
[90] Yue Qi. 2024. Pilot Quasi-Experimental Research on the Effectiveness of the Woebot AI Chatbot for Reducing Mild Depression Symptoms among Athletes. *International Journal of Human–Computer Interaction* 0, 0 (2024), 1–8. https://doi.org/10.1080/10447318.2023.2301256 arXiv:https://doi.org/10.1080/10447318.2023.2301256
[91] Mashfiqui Rabbi, Katherine Li, H Yanna Yan, Kelly Hall, Predrag Klasnja, and Susan Murphy. 2019. ReVibe: a context-assisted evening recall approach to improve self-report adherence. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 4 (2019), 1–27.
[92] Nina Rajcic and Jon Mccormack. 2020. Mirror Ritual: An Affective Interface for Emotional Self-Reflection. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). https://api.semanticscholar.org/CorpusID:216036185
[93] Bhavika R. Ranoliya, Nidhi Raghuwanshi, and Sanjay Singh. 2017. Chatbot for university related FAQs. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2017), 1525–1530. https://api.semanticscholar.org/CorpusID:33204447
[94] Bhupesh Rawat, Ankur Singh Bist, Muhammad Fakhrezzy, Regina Dinda Octavyra, et al. 2023. AI based assistance to reduce suicidal tendency among youngsters. *APTISI Transactions on Management* 7, 2 (2023), 102–109.
[95] Carol D Ryff and Burton H Singer. 2008. Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of happiness studies* 9 (2008), 13–39.
[96] Rahime Belen Sağlam, Jason R. C. Nurse, and Duncan Hodges. 2021. Privacy Concerns in Chatbot Interactions: When to Trust and When to Worry. *ArXiv* abs/2107.03959 (2021). https://api.semanticscholar.org/CorpusID:234599506

[97] Indriretno Setyaningrahayu, Pramusita Raring Widhi, and Elisabet Titik Murtisari. 2019. The use of video-based reflection to facilitate pre-service English teachers' self-reflection. *Indonesian JELT: Indonesian Journal of English Language Teaching* (2019). https://api.semanticscholar.org/CorpusID:210536080

[98] Ameneh Shamekhi and Timothy W. Bickmore. 2015. Breathe with Me: A Virtual Meditation Coach. In *International Conference on Intelligent Virtual Agents*. https://api.semanticscholar.org/CorpusID:16269226

[99] Olivia Sidoti and Jeffrey Gottfried. 2023. About 1 in 5 US teens who've heard of ChatGPT have used it for schoolwork. (2023). https://www.pewresearch.org/short-reads/2023/11/16/about-1-in-5-us-teens-whove-heard-of-chatgpt-have-used-it-for-schoolwork/

[100] Robert R. Stains. 2012. Reflection for connection: Deepening dialogue through reflective processes. *Conflict Resolution Quarterly* 30 (2012), 33–51. https://api.semanticscholar.org/CorpusID:144940415

[101] Anna Sutton. 2024. Memory recall can increase authenticity and engagement: understanding the role of work context and memory function in autobiographical memories. *Discover Psychology* 4, 1 (2024), 86.

[102] Yunlong Tang, Jing Bi, Siting Xu, Luchuan Song, Susan Liang, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, et al. 2023. Video understanding with large language models: A survey. *arXiv preprint arXiv:2312.17432* (2023).

[103] Petrus Te Braak, Theun Pieter van Tienoven, Joeri Minnen, and Ignace Glorieux. 2023. Data quality and recall bias in time-diary research: The effects of prolonged recall periods in self-administered online time-use surveys. *Sociological Methodology* 53, 1 (2023), 115–138.

[104] Gabriela A Torres. 2024. *Design considerations for an AI-prompted "Future-Self" video journaling tool to enhance self-efficacy.* Ph. D. Dissertation. Massachusetts Institute of Technology. https://hdl.handle.net/1721.1/154010

[105] Ba Tu Truong and Svetha Venkatesh. 2007. Video abstraction: A systematic review and classification. *ACM transactions on multimedia computing, communications, and applications (TOMM)* 3, 1 (2007), 3–es.

[106] A. M. TURING. 1950. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind* LIX, 236 (10 1950), 433–460. https://doi.org/10.1093/mind/LIX.236.433 arXiv:https://academic.oup.com/mind/article-pdf/LIX/236/433/30123314/lix-236-433.pdf

[107] Wayne Turnbull and Patricia M. Mullins. 2007. Socratic dialogue as personal reflection. *Reflective Practice* 8 (2007), 108 – 93. https://api.semanticscholar.org/CorpusID:143986454

[108] Lisette van der Meer, Sergi G. Costafreda, André Aleman, and Anthony S. David. 2010. Self-reflection and the brain: A theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience & Biobehavioral Reviews* 34 (2010), 935–946. https://api.semanticscholar.org/CorpusID:11526192

[109] Emma L van der Schyff, Brad Ridout, Krestina L. Amon, Rowena Forsyth, and Andrew J. Campbell. 2023. Providing Self-Led Mental Health Support Through an Artificial Intelligence–Powered Chat Bot (Leora) to Meet the Demand of Mental Health Care. *Journal of Medical Internet Research* 25 (2023). https://api.semanticscholar.org/CorpusID:258774301

[110] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[111] Ziyang Wang, Shoubin Yu, Elias Stengel-Eskin, Jaehong Yoon, Feng Cheng, Gedas Bertasius, and Mohit Bansal. 2024. VideoTree: Adaptive Tree-based Video Representation for LLM Reasoning on Long Videos. *arXiv preprint arXiv:2405.19209* (2024).

[112] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging large language models to power chatbots for collecting user self-reported data. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–35.

[113] Matthew Alexander Whitby, Ioanna Iacovides, and Sebastian Deterding. 2023. "Conversations with pigeons": Capturing Players' Lived Experience of Perspective Challenging Games. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (2023), 833–855.

[114] ShunYi Yeo, Gionnieve Lim, Jie Gao, Weiyu Zhang, and Simon Tangi Perrault. 2024. Help Me Reflect: Leveraging Self-Reflection Interface Nudges to Enhance Deliberativeness on Online Deliberation Platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–32.

[115] Fatma Gizem Karaoglan Yilmaz, Ramazan Yilmaz, and Mehmet Ceylan. 2023. Generative artificial intelligence acceptance scale: A validity and reliability study. *International Journal of Human–Computer Interaction* (2023), 1–13.

[116] Junjie Yin, Zixun Chen, Kelai Zhou, and Chongyuan Yu. 2019. A Deep Learning Based Chatbot for Campus Psychological Therapy. arXiv:1910.06707 [cs.AI] https://arxiv.org/abs/1910.06707

[117] Kam-shing Yip. 2006. Self-reflection in Reflective Practice: A Note of Caution. *British Journal of Social Work* 36 (2006), 777–788. https://api.semanticscholar.org/CorpusID:144596584

[118] Kam-shing Yip. 2007. Self-reflection in reflective practice: a Jaspers' orientation. *Reflective Practice* 8 (2007), 285 – 298. https://api.semanticscholar.org/CorpusID:144948030

[119] Xiaoyi Zhang, Laura R Pina, and James Fogarty. 2016. Examining unlock journaling with diaries and reminders for in situ self-report in health and wellness. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5658–5664.

[120] Guoying Zhao, Yante Li, and Qianru Xu. 2022. From Emotion AI to Cognitive AI. *International Journal of Network Dynamics and Intelligence* (12 2022), 65–72. https://doi.org/10.53941/ijndi0101006
[121] Zhengyue Zhao, Jinhao Duan, Xingui Hu, Kaidi Xu, Chenan Wang, Rui Zhang, Zidong Du, Qi Guo, and Yunji Chen. 2023. Unlearnable Examples for Diffusion Models: Protect Data from Unauthorized Exploitation. *ArXiv* abs/2306.01902 (2023). https://api.semanticscholar.org/CorpusID:259075262
[122] B. Zimmerman, Adam Moylan, John Hudesman, Niesha White, and Bert Flugman. 2011. Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. https://api.semanticscholar.org/CorpusID:15655596

A SYSTEM IMPLEMENTATION

A.1 Prompt for video understanding

A.1.1 How to understand the scene. "Here are some frames from a video. Describe, based on the timeline, what happened from a first-person perspective, including the environment and reactions from people around me. Make it as detailed as possible.",
"Below are several frames from a video. Please describe everything in as much detail as possible, focusing on the following aspects: ",
"- What is happening in the scene from a **first-person perspective**.",
"- The **environment**: including lighting, colors, objects, textures, weather, and the surroundings.",
"- The **environment and setting** (If it is indoor): describe the room, furniture, table arrangements, and any objects present (e.g., notebooks, dishes, utensils). Include lighting, colors, and background details.",
"- **People's interactions**: capture facial expressions, gestures, eye contact, and body language.",
"- **Actions and reactions**: focus on what people are doing (e.g., pouring drinks, passing food, taking notes). If someone makes a gesture or movement, describe how others respond.",
"- Highlight any **group dynamics**: does anyone seem to lead the conversation? Is someone quieter or more active? Are there any subtle social cues like nodding or leaning in?",
"- Pay attention to **small details**, such as words on items.",
"- Maintain **chronological order** to reflect the timeline accurately. If something changes between frames, be precise about what and when it changed.",
"- Avoid sensory details like smell or touch, but describe visible actions clearly, such as someone pointing, flipping a page, or glancing at their watch."

A.1.2 How to understand "My behavior" in the video. "Here are some frames from a video. Please summarize **from a first-person perspective** what actions I took. Focus not only on the actions but also on how they might have been perceived by others. Include both **positive and negative aspects** of my behavior, considering how my actions might have influenced the situation.",
"- Describe **non-verbal actions** (e.g., hand gestures, facial expressions, body language).",
"- Highlight actions that could be considered **helpful, polite, or constructive**.",
"- Identify any **potential mistakes, misunderstandings, or negative impacts** caused by my actions.",
"- If my actions changed during the interaction, provide a **sequential breakdown**.",
"- Include both **obvious** actions and more subtle ones (e.g., nodding, crossing arms, or avoiding eye contact).",
"- Summarize in a way that helps me reflect on what went well or bad and what could be improved."

A.1.3 How to understand "My interactions" with others. "Here are some frames from a video. Please focus on the **interactions** between me and other people. Summarize how I interacted with them and how they reacted, paying close attention to non-verbal communication.",
"- Focus on **non-verbal cues**: facial expressions, eye contact, gestures, and posture.",

"- Identify **patterns of interaction**: Was the interaction collaborative, tense, friendly, or neutral? Did the tone shift at any point?",
"- Highlight **key moments**: if a significant event or interaction occurred (e.g., a joke, a disagreement, or a shared laugh).",
"- Mention how my behavior may have influenced others' responses: Did someone seem more engaged or distant because of my actions?",
"- Provide a **chronological breakdown** if the interaction evolved across different moments.",
"- Ensure both subtle and explicit interactions are covered, including moments where no one spoke but body language conveyed meaning."

A.2 Prompt for keep memory

You are now the main character of a video described in four narratives. You are playing the role of "yesterday's me". Here's what happened:
1. Timeline-Based Description:timeline_description
2. Actions Summary:actions_summary
3. Interaction Summary:interaction_summary
4. Video Dialogue Content:
The following transcription captures the dialogue from the video, reflecting the verbal interactions among the characters. It provides insight into their emotional states and the context of their conversations video_transcription
In future conversations, please respond as if you were yesterday's version of me. Provide answers from the perspective of my experience at that time, only preserving **visual and auditory memories**. If further context or other details are needed, please ask for clarification.
Speak naturally, like a real person in their mid-twenties. Keep responses **concise**, under 150 words. Avoid mentioning videos directly or saying things like "you saw a video." The conversation will start with the user initiating directly.

A.3 Sample Multiple-Choice Video Understanding Questions(For Evaluation)

To evaluate the model's understanding of video content from a multimodal perspective, we constructed a standardized set of 12 multiple-choice questions for each tested video. Each question included four answer options, with only one correct answer. The questions were categorized into three types: perception, reasoning, and information synopsis.

The following two sets of questions are based on real-world video scenarios used in our system pipeline and illustrated in the video understanding flowchart (see Figure 3).

Example 1: Cooking Scene – Home Kitchen Preparation.

Perception.

(1) What items first appear in the sink in the video? **A.** Carrots and eggs **B.** Sponge and detergent **C.** Cutting board and spoon **D.** Onion and sliced beef
(2) What are the primary cooking tools used by the main character? **A.** Frying pan and rice cooker **B.** Knife and cutting board **C.** Oven and blender **D.** Microwave and induction cooker
(3) What kind of space is the person operating in? **A.** Shared kitchen **B.** Restaurant back kitchen **C.** Small home kitchen **D.** Outdoor temporary kitchen
(4) What is the likely identity of the white circular appliance shown in the video? **A.** Air fryer **B.** Rice cooker **C.** Multi-purpose blender **D.** Electric kettle

Reasoning.

(5) Why does the main character repeatedly check the pot? **A.** To take photos of the process **B.** To check if the food is fully cooked **C.** To have a remote conversation **D.** To prevent the food from burning
(6) Why does the person push carrots into the sink from the cutting board? **A.** For secondary washing **B.** For weighing **C.** For camera angle setup **D.** For later use
(7) What is the most likely substance added to the pot at the end of the video? **A.** Tomato sauce **B.** Cooking oil **C.** White sugar **D.** Salt or seasoning powder
(8) Why does the person stir the contents of the pot multiple times using chopsticks? **A.** To showcase stir-frying skills **B.** To avoid burning and ensure even heating **C.** To make the video more dynamic **D.** To demonstrate a cooking tutorial

Information Synopsis.

(9) What type of video is this most likely to be? **A.** Documentary segment **B.** Food vlog or short skit **C.** Cooking tutorial **D.** Promotional advertisement
(10) What is the core event of the video? **A.** Tasting various ingredients **B.** Demonstrating kitchen appliances **C.** Preparing a simple meal alone at home **D.** Teaching waste sorting techniques
(11) Who is the main character in the video? **A.** A middle-aged chef **B.** A young male home cook **C.** A food content creator **D.** A domestic helper
(12) What kind of atmosphere does the video convey? **A.** Warm and focused daily routine **B.** Tense and efficient work rhythm **C.** Entertaining and exaggerated **D.** Cold and technical instructional tone

Example 2: Convenience Store Parcel Pickup.

Perception.

(1) What is the first thing the person does in the video? **A.** Recording the exterior of a shop **B.** Waiting in line to make a payment **C.** Walking into a convenience store with a paper in hand **D.** Organizing items in a shopping bag
(2) What is the main location where the person is active? **A.** Ticket counter at a train station **B.** Home kitchen **C.** Checkout area of a convenience store **D.** Post office drop-off point
(3) What is the paper item that appears repeatedly in the video? **A.** A recipe **B.** A receipt or pickup slip **C.** A newspaper **D.** A discount flyer
(4) Which of the following items appears near the payment counter? **A.** A hand sanitizer bottle **B.** A coffee machine **C.** A black coin tray **D.** A mobile vacuum cleaner

Reasoning.

(5) What is the purpose of repeatedly opening the backpack? **A.** To showcase its structure **B.** To retrieve a wallet or place something inside **C.** To inspect the stitching **D.** To take out a water bottle
(6) Why does the person put a piece of paper into a red device? **A.** To recycle waste paper **B.** To submit an assignment **C.** To return a product **D.** To send a parcel label
(7) Why does the person repeatedly look at the same paper in the store? **A.** To check an address **B.** To verify a barcode **C.** To check a discount amount **D.** To fill out a questionnaire
(8) What does the presence of a coin tray at checkout suggest? **A.** Use of self-checkout systems **B.** Encouragement of cash payment **C.** Minimizing hand contact with the cashier **D.** Collecting point cards

Information Synopsis.

(9) What is the central event depicted in the video? **A.** Recording a queueing experience **B.** Demonstrating shopping hacks **C.** Completing a convenience store parcel pickup and payment **D.** A tutorial on unboxing a new gadget
(10) What is the main message the video conveys? **A.** How to pick up and pay for a package in a convenience store **B.** How to buy a train ticket at a station **C.** Beverage options at Japanese convenience stores **D.** Comparing different payment methods
(11) What is the relationship between the person and other characters? **A.** Customer and store staff **B.** Mentor and student **C.** Co-workers **D.** Parent and child
(12) What best describes the video's overall style? **A.** Emotional documentary **B.** Practical lifestyle tutorial **C.** News interview footage **D.** Entertaining comedy skit

B DESING WORKSHOPS GUIDE

B.1 Initial questions

(1) Do you usually engage in self-reflection?
(2) In what form do you typically reflect? Do you use any specific tools?
(3) What kinds of new reflection tools do you envision in today's technological landscape?
(4) Suppose there is a tool that creates an "AI version of yourself" based on your video recordings—would you be willing to try it?
(5) Would you feel comfortable sharing your personal video and behavioral data with such a system?
(6) What kind of data would you feel hesitant to share?

Introduction to Introspectus AI initial prototype

(1) After trying the prototype, what do you think of your experience with it?
(2) How can it support self-reflection?
(3) In what real-world scenarios do you think this tool could be useful?
(4) What aspects of the system do you think need improvement?
(5) Were there any features that felt confusing or difficult to use?
(6) Which features did you find most helpful for self-reflection?
(7) Did the interaction flow feel smooth and natural to you?
(8) What form do you think is the most reasonable for deploying this function?
(9) Do you have any concerns about using this system?

B.2 Representative Expert Feedback Extracts

Below are representative excerpts and paraphrased expert comments gathered during the design workshops. The feedback was grouped into four thematic categories through affinity diagramming.

Chatbot Tone and Communication Style

- "It would feel more natural if the AI spoke like a friendly peer,not like a therapist or a robot."
- "Sometimes the tone is too neutral. I want it to be encouraging, like someone I trust."
- "Avoid using judgmental language. The tone should be warm, even when pointing out mistakes."

Response Content

- "Don't just summarize what I did. Maybe it could ask 'why' or 'what should I do next?"'
- "Detailed suggestions would be helpful, especially when they are logically connected to my actions."

- “Can it give both positive and negative feedback? I’d like to hear what I did well too.”

AI Identity

- “I like the idea of the AI being a past version of myself, it makes the feedback easier to accept.”
- “If it’s framed as ‘me’, it feels more personal. I’d trust it more than a generic assistant.”
- “Don’t make it sound like a coach or someone else. Let it speak as ‘I’, that helps me relate to its perspective.”

Platform Design

- “I prefer not having to install anything complicated. Multi-platform access would be ideal.”
- “If it’s too complex, I wouldn’t use it often. A simple web-based tool might be best.”
- “Maybe add clearer instructions the first time to help new users get started will be better.”

C Guiding Questions for Reflective Dialogue

(1) **Scenario Recap and Understanding**
  - What happened during this scenario?
  - Did I overlook anything important during this activity?

(2) **Positive Behaviors and Strengths**
  - What did I do well?
  - What positive aspects of my behavior should I reinforce?
  - How can I maintain or enhance my good habits?

(3) **Areas for Improvement**
  - What could I have done better?
  - Are there areas where I could improve my behavior?
  - What suggestions do you have for correcting my negative behaviors?
  - How might I handle a similar situation differently in the future?
  - What long-term impact could this behavior have if repeated?

(4) **Emotional Awareness and Management**
  - How did I feel during this activity, and why?
  - Did my emotions influence my actions in this scenario?
  - How can I better manage my emotions in similar situations?

(5) **Broader Reflection and Patterns**
  - How does this behavior fit into my overall habits?
  - Are there any patterns in my actions that I should be aware of?

D USER STUDY QUESTIONNAIRE

D.1 Demographics

(1) Age
(2) Gender
(3) What is your highest professional qualification
(4) How familiar are you with: AIGC tools for creative writing (Such as stable difussion)
(5) How familiar are you with: Large language model for text dialogue(Such as Chatgpt))
(6) How familiar are you with: Using videos to document life&share life moments (such as vlogs)
(7) How familiar are you with: Using pictures to document life&share life moments (e.g., blogs, Instagram)
(8) Using text to document life&share life moments(such as diary))

D.2 Semi-Structured Interview Questions for Experimental Group

1. Challenges and Usability.

(1) Can you describe any challenges you faced while writing the daily reflection diary?
(2) Were the reflection structure and guiding suggestions helpful or confusing? Was there any part you found hard to follow?

2. Feedback Quality and Emotional Impact.

(1) While writing your reflections, did any emotions or unexpected thoughts arise?
(2) Did reflecting on your behavior through writing lead you to reconsider any actions or decisions? Could you give an example?
(3) Did the act of writing feel emotionally supportive or burdensome?

3. Understanding and Reflection Guidance.

(1) Did the daily reflection guide help you notice any behavior or emotion you usually wouldn't pay attention to?
(2) Did the written format help you process and organize your thoughts clearly?
(3) Did you find the structured questions (e.g., "what happened?", "what could I improve?") meaningful or repetitive?

4. Tool Acceptance and Long-Term Use.

(1) To what extent did the structured diary add value to your self-reflection? Could you provide an example?
(2) Did you find it easy or difficult to maintain the habit of daily writing? What factors affected your motivation?
(3) Do you think you would continue using structured journaling as a tool for reflection? Why or why not?
(4) What are your thoughts on reflecting through writing versus using an AI assistant?
(5) Do you see a role for journaling or reflection tools in your everyday routine?

5. Suggested Improvements and Reflection Process.

(1) If you could change anything in the reflection structure or format to better support your thinking, what would you change?
(2) Did the written prompts feel natural and helpful, or did they feel limiting? Why?

D.3 Semi-Structured Interview Questions for Control group

1. Challenges and Usability.

(1) Can you describe any challenges you faced while writing the daily reflection diary?
(2) Were the reflection structure and guiding suggestions helpful or confusing? Was there any part you found hard to follow?

2. Feedback Quality and Emotional Impact.

(1) While writing your reflections, did any emotions or unexpected thoughts arise?
(2) Did reflecting on your behavior through writing lead you to reconsider any actions or decisions? Could you give an example?
(3) Did the act of writing feel emotionally supportive or burdensome?

3. Understanding and Reflection Guidance.

(1) Did the daily reflection guide help you notice any behavior or emotion you usually wouldn't pay attention to?

(2) Did the diary help you process and organize your thoughts clearly?
(3) Did you use any tools (e.g., ChatGPT) to help structure your reflections? If so, how did it influence your writing or thought process?

4. Tool Acceptance and Long-Term Use.

(1) To what extent did the structured diary add value to your self-reflection? Could you provide an example?
(2) Did you find it easy or difficult to maintain the habit of daily writing? What factors affected your motivation?
(3) Do you think you would continue using structured journaling as a tool for reflection? Why or why not?
(4) What are your thoughts on reflecting through writing versus using an AI assistant?
(5) If you used GPT-based tools during the study, would you consider using them again for future self-reflection?
(6) Do you see a role for journaling or reflection tools in your everyday routine?

5. Suggested Improvements and Reflection Process.

(1) If you could change anything in the reflection structure or format to better support your thinking, what would you change?

Received October 2024; revised April 2025; accepted August 2025