PANDALens: Towards AI-Assisted In-Context Writing on OHMD During Travels

Runze Cai
runze.cai@u.nus.edu
Synteraction Lab
School of Computing, National University of Singapore
Singapore

Nuwan Janaka*
nuwanj@u.nus.edu
Synteraction Lab
Smart Systems Institute, National University of Singapore
Singapore

Yang Chen*
cyang@u.nus.edu
College of Design and Engineering, National University of Singapore
Singapore

Lucia Wang
luciajw@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, United States

Shengdong Zhao†
shengdong.zhao@cityu.edu.hk
Synteraction Lab
School of Creative Media & Department of Computer Science, City University of Hong Kong
Hong Kong, China
National University of Singapore
Singapore

Can Liu†
canliu@cityu.edu.hk
School of Creative Media, City University of Hong Kong
National University of Singapore
Hong Kong, China

ABSTRACT

While effective for recording and sharing experiences, traditional in-context writing tools are relatively passive and unintelligent, serving more like instruments rather than companions. This reduces primary task (e.g., travel) enjoyment and hinders high-quality writing. Through formative study and iterative development, we introduce *PANDALens*, a **P**roactive **A**I **N**arrative **D**ocumentation **A**ssistant built on an Optical See-Through Head Mounted Display that supports personalized documentation in everyday activities. *PANDALens* observes multimodal contextual information from user behaviors and environment to confirm interests and elicit contemplation, and employs Large Language Models to transform such multimodal information into coherent narratives with significantly reduced user effort. A real-world travel scenario comparing *PANDALens* with a smartphone alternative confirmed its effectiveness in improving writing quality and travel enjoyment while minimizing user effort. Accordingly, we propose design guidelines for AI-assisted in-context writing, highlighting the potential of transforming them from tools to intelligent companions.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Empirical studies in interaction design**.

*Both authors contributed equally to this research.
†Corresponding Authors.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642320

KEYWORDS

HMD, smart glasses, AI, large language model, multimodal information, Human-AI collaborative writing, in-context writing, travel blog

ACM Reference Format:
Runze Cai, Nuwan Janaka, Yang Chen, Lucia Wang, Shengdong Zhao, and Can Liu. 2024. *PANDALens*: Towards AI-Assisted In-Context Writing on OHMD During Travels. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3613904.3642320

1 INTRODUCTION

Documenting life experiences has long been an essential need of our lives [15, 19, 44], and various solutions have been developed to facilitate this process. Using mobile devices, users can take pictures on phones or use lifelogging cameras to record daily activities [10, 48]. One can also leverage AI technology to capture moments of significance automatically [10, 19]; however, relying solely on AI makes it difficult to fully understand human intentions and accurately extract the most interesting moments in a meaningful form [48]. Another approach is in-context multimodal authoring, exemplified by tools like *LiveSnippets* [44], which allow users to document their experiences with photos and voice comments as they unfold. However, interactions in such systems are passive, requiring explicit efforts with hands-occupied and heads-down interaction [39, 95] from the author. Additionally, the content generated often appears mundane, typically presented as a simple chronological listing of events rather than a compelling and engaging narrative [44].

With the introduction of Large Language Models (LLMs) and other technological advancements, there is now an opportunity for proactive engagement and the generation of longer, more detailed content using in-situ documentation of life experiences. In particular, LLMs can assist users in generating rich and expressive narratives by providing context-related guidance and suggestions.

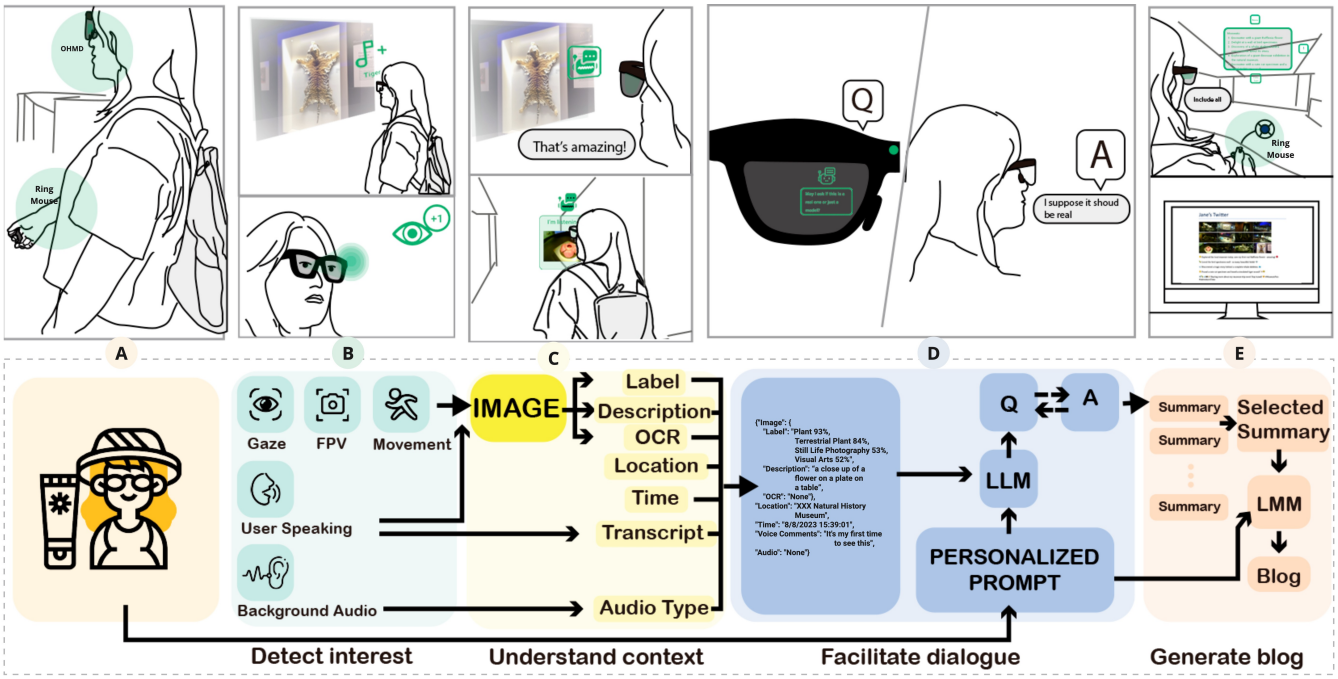


Figure 1: (A) A user travels with *PANDALens*, an AI-assisted in-context writing tool equipped with an Optical See-Through Head-Mounted Display (OHMD) and a ring mouse. (B) The system leverages various modalities to detect the user's interests during travel, such as potential interesting audio (e.g., tiger sounds in a museum) and gaze patterns (e.g., looking at flowers). (C) Detecting interests, the system displays icons (e.g., with auto-captured images) and prompts the user to comment verbally. It then transcribes this comment and combines it with other data such as image, audio, time, and location to assemble the contextual data. (D) Using the contextual information, the system formulates context-specific questions in the user's preferred style with a Large Language Model (LLM). The user can then respond to these questions. LLM also creates a summary of the moment, which can be refined based on the user's feedback. (E) Post-trip, the user can activate *PANDALens* using the ring mouse to automatically generate travel blogs. A list of recorded moments is displayed for the user to choose from. Once selected, the system drafts a travel blog that mirrors the user's unique style.

This means individuals can embed their immediate and personalized feelings and reflections in the instant moments. Proactive engagement through in-context writing allows for more immersive and authentic life document experiences, avoiding the decay of memories over time [44]. Moreover, with the assistance of LLMs, the need for post-editing efforts can be significantly reduced, as the generated content is more coherent, well-structured, and ready for sharing or preservation [67, 88]. These technological advances open up new possibilities for capturing and preserving meaningful moments, enabling individuals like Jane, a travel enthusiast in Figure 1, to create more engaging and detailed accounts of their experiences.

This leads us to our research question and design goal: *How can we support high-quality, personalized documentation in everyday activities (e.g., travels) but with seamless interaction during users' primary tasks (e.g., travels) and minimum post-editing efforts?*

We introduce *PANDALens* (**P**roactive **AI** **N**arrative **D**ocumentation **A**ssistant), an AI-assisted in-context writing system on Optical See-Through Head Mounted Displays (OST-HMD, OHMD, augmented-reality smart glasses). The wearable heads-up platform [95] reduces the efforts in moment capture by leveraging AI to observe multimodal context information[1] [26, 68] from user natural behaviors (e.g., gaze, movement, voice) and environments (e.g., objects in egocentric view and ambient audio), subsequently offering moment capture suggestions proactively. Users can respond to these suggestions via natural voice dialogue or subtle ring interactions [17, 72]. To elicit detailed user expressions and facilitate intelligent dialogues, the Large Language Model (LLM) is used to interpret the multimodal contextual information of the captured moment and generate context-related questions. To enhance the quality of the final documentation, the integrated LLM utilizes contextual information with detailed user expressions to craft the narratives progressively, minimizing user editing efforts.

Compared to previous approaches (e.g., [44]), which often only satisfy part of the design goals, *PANDALens* is the first of its kind we are aware of to largely satisfy all the design goals specified

[1] In our context, multimodal information refers to visual, audio, spatial, and temporal data of the user and environment.

in the research question. This is achieved by introducing multimodal sensing, AI interpretation, and carefully designed mixed-initiative interfaces and interactions. Through this integrated approach, *PANDALens* enables the generation of rich and personalized content, particularly longer documents, through implicit, natural, and context-aware inputs. Acting as a proactive AI with LLMs, *PANDALens* discerns documentation intentions within live, multimodal contexts, thus reducing user effort in capturing moments and providing context-rich descriptions. By transforming mentally demanding documentation into a secondary background activity, *PANDALens* enables users to concentrate on their primary tasks, such as travel, with minimum distraction.

PANDALens was compared with a smartphone in-context writing application [44] in realistic travel scenarios in a local museum involving 16 participants, assessing its overall capability to generate high-quality narratives with high user experience during travel. Our findings indicated that *PANDALens* could effectively improve travel enjoyment, evoke more profound reflections, and produce high-quality narrative documentation with reduced effort.

Our contributions are threefold: 1) We introduce a design space for multimodal context information naturally occurring in travel scenarios, demonstrating how this information can enhance interaction and writing. 2) We present an AI-assisted writing approach that transfers the passive mobile tool to a proactive wearable assistant. It is accompanied by a proof-of-concept artifact, *PANDALens*, designed for seamless in-context writing on OHMD. 3) We provide an empirical study validating this approach in realistic scenarios, offering further design implications.

2 RELATED WORK

To develop a wearable AI assistant that can help users document their experiences with minimal interference in their primary tasks (e.g., travels), we consider the following areas.

2.1 Documenting Life Experiences

Capturing and recording users' experiences and activities is an active area of research in Human-Computer Interaction (HCI), with lifelogging being a prominent focus [48, 66, 73]. Such experience documentation serves various purposes, including Recollecting, Reminiscing, Retrieving, Reflecting, and Remembering intentions (the 5Rs) [73], as well as sharing experiences [66]. Recent research in experience documentation methods [44, 47] highlights the advantages of in-context documentation (i.e., documenting the experience and personal thoughts in real-time) over traditional post-context/retrospective documentation (i.e., recalling past events and recording them). In-context documentation results in more detailed descriptions of experiences with lower memory decay and recall biases and reduces barriers to capturing [44, 47]. Nevertheless, the frequent manual actions required for in-context documentation can be a barrier to documenting everyday life experiences [47].

2.2 AI-Assisted Experience Capturing and Documentation

With the advancement of Machine Learning and Artificial Intelligence (AI) systems, barriers such as the manual effort to capture and filter data to find important moments and post-processing have been reduced [10, 19, 45, 73]. Current research has explored two avenues in this regard. The first is supporting automatic prominent moment detection and extraction. Such techniques use egocentric video alone [9, 23, 53, 79] to capture visually appealing moments or highlight events; or they are combined with other modalities (such as audio, motion, gaze, location, etc.) to understand fine-grained context and activities [10] and are adaptive to users' real-time attention [19]. The second avenue is to simplify post-processing on captured moments to craft high-quality narratives or stories [11] (e.g., Day One Journal App[2]). However, there are still challenges for in-context documentation of experiences with AI assistance. AI-initiated automatic capturing can result in incorrect captures due to a lack of user feedback or the user's mental state [10, 48]. Similarly, although AI helps in post-processing, including editing [75, 88], a lack of proper user guidance can lead to misleading information [7, 38, 85] and templated output [49, 65, 84] that deviates from user expectations. Moreover, the integration of AI for both in-context capturing and post-processing has received less attention in the literature. Therefore, we conduct a formative study to understand the most suitable moments for capturing experiences and user expectations of AI for in-context experience documentation.

To address the issues with AI-initiated systems that interact with humans, mixed-initiative interactions have been introduced [3, 4, 33], allowing humans and intelligent agents to collaborate efficiently to obtain the expected output. Mixed-initiative interactions have been applied to collaborative documentation and writing tasks [21, 71, 75]. In contrast to traditional Human-AI collaborative documentation, where users primarily focus on writing tasks, in-context experience documentation presents challenges because users primarily focus on the experience, with documentation as a secondary goal. Creating a good user experience when interacting with advanced AI (e.g., LLMs), especially when uncertain about its capabilities, is also challenging [91]. These challenges require both AI and humans to iteratively document user experience and refine the co-created document [20, 24, 71] without distracting users from their primary activities [61].

2.3 Heads-up Wearable AI Assistant for In-Context Documentation

While there are mobile phone applications to support in-context documentation [44], they can interfere with the travel experience due to the constant need for hands-occupied and heads-down posture [39, 95]. As an alternative, the emerging wearable platform, Optical See-Through Head-Mounted Display (OHMD, AR smart glasses) [37] and the heads-up computing interaction paradigm [95] show potential due to increased situational awareness [39, 60] and support for non-intrusive interactions with primary tasks [17, 40]. However, how to utilize heads-up computing for in-context documentation with AI assistance and minimal interference with the primary task is underexplored.

We introduce *PANDALens*, a system designed for OHMDs. *PANDALens* leverages mixed-initiative interactions to reduce interference and utilizes LLMs for document co-creation. It employs a multimodal context analyzer to detect user interests [69, 74] and initiate AI interactions when users' attention isn't occupied [5, 32],

[2]https://apps.apple.com/us/app/day-one-journal-private-diary/id1044867788

offering non-intrusive suggestions [17, 29]. In *PANDALens*, LLMs play a key role, benefiting from their text-processing capabilities and ease of customization for content organization [14, 25, 57, 67]. Unlike traditional Human-AI collaborative documentation, where users explicitly provide intentions [21, 75], *PANDALens* captures multimodal context implicitly through user behaviors, facilitating more natural and personalized co-created documentation [71]. To overcome challenges like proper prompting [94] and avoid generating unrelated information [7], *PANDALens* employs an LLM pipeline aligned with user expectations. Additionally, LLMs, armed with multimodal context information, ask context-related in-situ questions [52] to enrich the documentation, a departure from traditional systems offering fixed questions [44].

3 STUDY OVERVIEW

Our research began with a formative study to understand the design space of harnessing multimodal information for AI-assisted in-context writing, focusing on the travel scenario. Following this, we iteratively developed the proof-of-concept system, *PANDALens*. We then compared *PANDALens* with *LiveSnippets* [44], a smartphone-based in-context writing system in real-world travel settings. Note that all studies were approved by our university's institutional review board (IRB), and participants were compensated ~7.5 USD per hour, a standard rate for user studies in the local context.

4 FORMATIVE STUDY: HARNESSING MULTIMODAL CONTEXT INFORMATION FOR NATURAL HUMAN-AI COLLABORATIVE WRITING

4.1 Research Questions

We envision the development of a wearable AI assistant collaborating with travelers to enhance their travel experience documentation. However, realizing this vision necessitates addressing several key issues. These include identifying the challenges users face with current documentation methods, recognizing user behaviors that can accurately discern the user's interests and desire to document a particular moment, and understanding the user's preferences for AI assistance. To explore these intricacies, we conducted a formative study with the following three research questions.

- RQ1: What are the challenges users face when capturing and documenting interesting moments during travel using existing tools?
- RQ2: What behavior do users have to indicate their interests and intentions when they aim to capture and/or share specific moments during travel?
- RQ3: What are users' expectations and preferences regarding the behavior and interactions of the AI assistant during travel to facilitate in-context writing?

4.2 Participants

We recruited twelve volunteers (P1-P12, 5 females, 7 males, mean age = 24.3 years, SD = 3.9 years) from the university community. To ensure the accuracy of our eye-tracking equipment, we selected participants with normal or corrected vision, excluding those wearing spectacles. Our participant pool primarily consisted of eleven frequent travel sharers on social media, but included one participant who did not share travel experiences to provide an alternative perspective. Our main goal was to observe behaviors from first-time visitors (10) to identify behaviors associated with interests that accompany first-time exposure, but we included two re-visitors to explore different perspectives during revisit experiences.

4.3 Apparatus

To better capture users' natural behavior during travel, we utilized portable devices to record what users see, hear, and do during the trip, aligning with prior lifelogging and travel research [10, 16].

Participants wore a backpack containing a laptop (Acer Swift Go 14, 1.2kg), which collected all the recordings. Their visual experiences, inclusive of gaze patterns over first-person view (FPV), were recorded using a Pupil Core[3] eye-tracker (World Camera: 30Hz, 1080p, FoV: 139°×83°; Eye Cameras: 120Hz) that connected to the laptop. Audio experiences comprising verbal interactions and ambient sounds were captured via a microphone attached to the participants. As with a conventional trip, participants could use their smartphones to record moments as they desired. An accompanying experimenter, maintaining a distance to avoid interference, recorded user actions using a mobile phone (Pixel 6) from a third-person view (TPV). This TPV feed was streamed to the laptop in real-time through DroidCam[4]. The laptop synchronized and recorded the FPV, gaze, TPV, and audio data using ShareX[5]. This setup enabled instant playback for reviewing recorded experiences later.

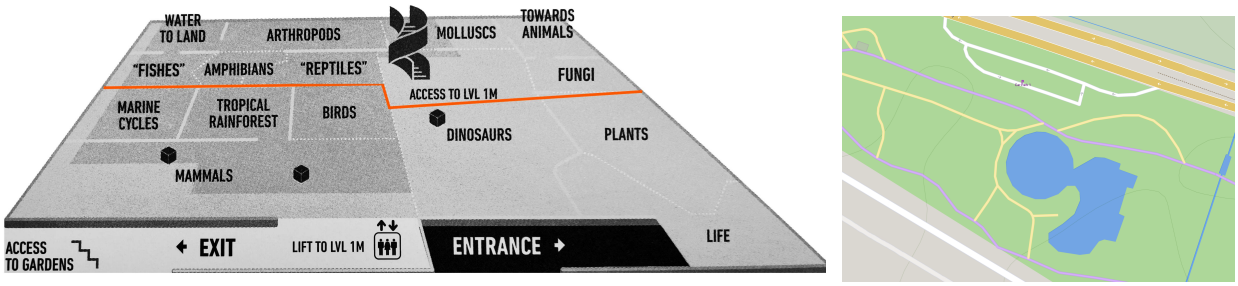
4.4 Study Design

To examine the user behavior across travel contexts, we included two types of travel experience: educational exploration (i.e., visiting a natural history museum) [50] and recreational exploration (i.e., visiting a park) [36]. The visit to the museum offers an information-rich setting where users can deeply explore various specimens and fossils (as shown in Figure 2a) and thematic displays. In contrast, in the park, participants can relax and immerse themselves in nature (as shown in Figure 2b). Participants were randomly allocated to one of two locations, which resulted in 12 travel sessions comprising six park and six museum experiences.

4.5 Task and Procedure

Our study includes two phases: a free-form travel exploration followed by a retrospective interview. In the exploration phase, participants were required to wear an eye tracker with the laptop in a backpack (sec 4.3) for recording and asked to freely explore the assigned destination for at least 30 minutes, accompanied by an experimenter. After this, participants were asked to review the recordings, including both FPV, Gaze, TPV, and Timestamp, with the audio of the user and their ambient environment at 1.5x speed (as shown in Figure 3). They were asked to pause the recording to indicate moments they considered interesting (e.g., those that influenced their emotions or engagement levels) during travel and the moments they wanted to document. They were also asked to

[3]https://pupil-labs.com/products/core/
[4]https://play.google.com/store/apps/details?id=com.dev47apps.droidcam
[5]https://getsharex.com/



(a)

(b)

Figure 2: Locations of Experiments. (a) A local natural history museum with exhibitions of specimens and fossils. Note: For Comparative Study (sec 6), we divided the museum into two areas with equal exhibitions, as indicated in the orange line. (b) A local park with greenery and a lake.

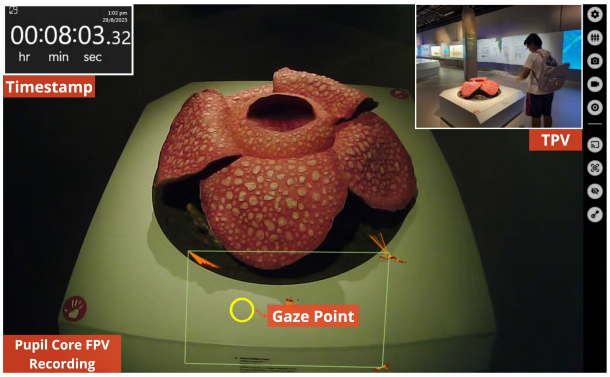


Figure 3: Recordings of users' natural travel exploration. After the trip, participants can review their travel behaviors with FPV, Gaze Trace, TPV, Timestamp, and Audio from the user speaking and environment.

share their challenges as they attempt to use the identified moments for documentation and their vision for how future AI may assist in such tasks. The entire procedure took approximately 60 minutes per participant.

4.6 Data Analysis

Upon consolidating 12 transcribed interview notes with observational notes detailing user behaviors and environments (i.e., participant-marked recording frames, as shown in Figure 3, with remarks, including why interesting, what behavior reveals interest, why documentation, challenges with documentation, envisioned AI functions), we employed a thematic analysis approach as outlined by Braun and Clarke [12]. The details of the analysis procedure can be found in Appendix A.

4.7 Findings

4.7.1 RQ1: What are the challenges users face when capturing and documenting interesting moments during travel using existing tools? Participants highlighted the significant challenge of temporal pressure when attempting to capture captivating moments during their travels. For instance, P1, when recalling an unexpected encounter with a butterfly, mentioned, "It's difficult to capture such a transient and unanticipated moment," pointing to the issue of not having enough time to take out a phone or camera in such fleeting situations.

For similar reasons, many moments that evoked strong emotional responses went uncaptured and unrecorded. Participants expressed their desire for these moments to be documented in real-time, as they recognized that "capturing my emotions as they occurred could have provided more vivid details and a deeper replay of my experience compared to solely relying on memory (P2)". However, the fleeting nature of these emotions made it impractical to use their phones to record them constantly. Instead, participants often had to depend on retrospection to document these moments. Unfortunately, this reliance on retrospection often resulted in the omission or forgetfulness of many of these valuable experiences due to the natural decay of memory [44].

4.7.2 RQ2: What behavior do users have to indicate their interests and intentions when they aim to capture and/or share specific moments during travel? Aligning with the existing literature [19, 77, 81, 86], our observations from users' free travel exploration (duration: M = 32.7 mins, SD = 4.3 mins) reveal that specific behaviors such as prolonged gaze, the increased pitch of tone, and sudden change in movement (e.g., slowing down, moving closer to target), emotion change, facial cues, indicates users' heightened interest. For example, "Witnessing a butterfly literally jump from one petal to another was astonishing. I stopped walking and wished to record it (P1)". These behaviors are valuable indicators of what captures users' attention and curiosity and subsequently predict the events that users want to document.

However, it is worth noting that not all show-interest moments of users lead to writing desire. Participants reported two primary motivations for documenting travel experiences: sharing experiences and preserving memories, aligning with prior research [82, 83]. Thus, participants are more likely to document interesting moments that resonate with these motivations. For instance, P8 remarked,

"This moment [touching dinosaur skeletons] is quite engaging and certainly worth sharing with my close friends," while P9 stated, "I prioritize unique moments that evoke cherished memories for documentation." Other moments, although attracting interest (e.g., viewing a local map depicting crustal evolution or finding someone skateboarding quickly in the park), are not considered worthy of documentation. Among all the show-interest moments we recorded, roughly 49.7% are regarded as worth documenting, while the remaining 50.3% are considered not. Notably, the participant who focused on preserving memories without sharing needs recorded more interesting moments (75%) than the group average (49.7%), showing a preference for comprehensive personal logs without judging what's interesting to others. Additionally, two re-visitors documented interesting moments at a rate (54.3%) similar to first-time visitors. Although the sample size is too small to generate meaningful inferences, we see instances where participants' familiarity with a location doesn't lessen their interest in documenting revisiting experiences, as "I can always find something new and worth recording even in the same place. (P12)"

Two Types of Interests that Influence Writing Intentions. Interestingly, we discerned that different types of interest during travel influenced documentation intentions. Building upon previous literature, we categorized these interests as situational and personal [10, 69]. Situational interest refers to transient cognitive and emotional shifts, often triggered by external events or surprises [69]. In contrast, personal interest represents more enduring motivations connected to individual preferences and values [69] (e.g., encountered personally favored audio and visual experiences during travel).

In our study, moments driven by situational interest are more linked with social sharing, allowing participants to share "interesting surprises" with others, "travel differs from daily routines with its unexpected surprises. That's the beauty of journeys (P3)". These sudden, unexpected events enhance content engagement, as noted by "I'd like to share about the 'ugly' turtle specimen I found, which makes my sharing humorous (P7)". In contrast, personal interests are less frequently documented, mainly serving as personal recollections rather than shared experiences, "I plan to keep the adorable dog as a precious memory in my personal collection. But I have no intention of sharing it with my friends as I have already posted many pictures of my dogs on social media. (P2)". Yet, these personal interests, while in the minority (11.2%), play an indispensable role in the narrative. They introduce embellishing elements that reflect the author's individuality and preferences. P2 added, "I will keep the encounter violin practice [in the park] in sharing, as it reminded me of practicing that same melody in my childhood. It's still such a beautiful sound."

Thus, considering the unique value each type of interest brings to the documentation of travel experiences, both situational and personal interests should be incorporated, although they may hold different weight in the final documentation composition.

4.7.3 *RQ3: What are users' expectations and preferences regarding the behavior and interactions of the AI assistant during travel to facilitate in-context writing?*

Proactive Travel Assistant. Participants desired an AI system capable of understanding their preferences, identifying interesting topics and experiences, and predicting their interests based on past travels. One individual hoped the AI would function like "a well-informed travel buddy (P1)". Additionally, they appreciated an AI engaging proactively, asking questions related to previous experiences and encouraging them to elaborate on moments for detailed travel blogs. As P9 expressed, "I'd appreciate it if the AI could capture my verbal reactions to personalize my blog. When detecting my happiness, it can ask me questions for more details." Another participant added, "It can ask about my past experiences. This will make the trip more enjoyable. (P2)"

Despite the enthusiasm for an interactive AI, participants highlighted the necessity for the AI to maintain a non-intrusive presence, especially during moments of immersion in travel. As outlined by P6, "When immersed in nature's sounds in the garden, I don't wish AI to talk to me... It should display relevant information subtly, allowing me to act quickly without detracting from my experience." When asked what happens if AI fails to take the initiative to execute desired tasks or perform wrong tasks, participants expressed the need to regain control, using verbal commands (12/12) to gestures (9/12) and controllers (6/12).

Auto Blog Generation with User Authenticity. Post-travel, participants envisioned AI to draft a narrative automatically, reducing the editing workload, a major burden in limiting participants from creating such documents, "One reason I don't share [my experience] is, it takes a lot of effort to create a nice blog (P11)". Nonetheless, users emphasized the importance of personal style, "I still prefer to rephrase the final writing in my own style; it should reflect me (P7)".

Our findings illustrate a comprehensive perspective on AI assistant design, emphasizing proactive engagement while respecting and aligning with users' preferences. To achieve this, various contextual cues for guiding AI interactions and content creation can be utilized, as detailed in the next section.

4.8 Design Space of Context Information for AI-Assisted In-Context Writing During Travels

Considering users' challenges faced when using smartphones during travel and their expectations for an AI assistant, we summarize the design space for utilizing multimodal context information in AI-assisted In-Context Writing, consisting of two parts: interaction and content generation, as presented in Figure 4.

4.8.1 *Interaction.*

Utilizing various contextual information can assist in AI-involved interactions (i.e., the left branch in Figure 4), making it a proactive travel companion by identifying interesting moments to capture, asking questions, and guiding the optimal interaction methods. For interest detection, both situational and personal interest [77, 81] need to be discerned, as they trigger user documentation intentions. Specifically, in the travel context, gaze patterns and movement changes (e.g., approaching an entity) can reveal both situational and personal interests [19, 77, 86]. Additionally, speaking tone, gestures (e.g., interaction with physical objects),

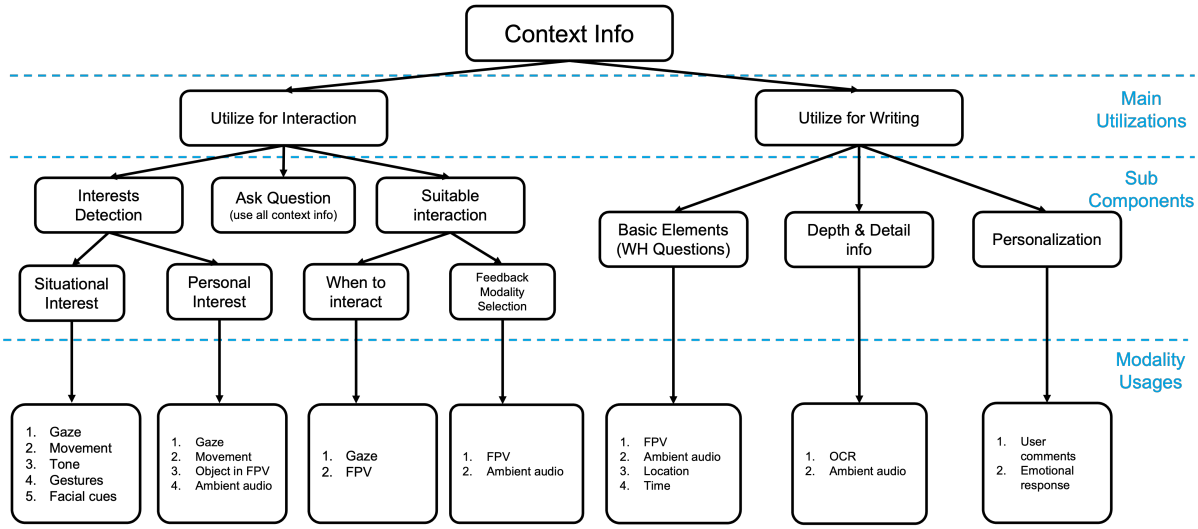


Figure 4: The design space describes how to utilize multimodal context information for both interaction and content generation for in-context writing.

and facial cues can pinpoint situational interests, while detecting environmental cues can identify personal interests, such as objects in FPV matched preferred visual elements (e.g., animals) and audio (e.g., music) associated with memories.

After detecting such interests, the system can automatically capture moments and leverage all the available context information of the moment to pose questions to users if they confirm interest. These questions can reduce travel boredom and capture users' fleeting in-the-context emotions. Furthermore, context information can help decide the optimal time of interaction to avoid interruption. Using gaze information, it can recognize users' engagement in primary tasks (e.g., enjoying a scenery); thus, certain AI suggestions should be deferred until users are more receptive [5]. Such deferrals can be aligned with transitions in activities [32], as indicated by significant changes in FPV. Additionally, the feedback modality should be tailored to the user's engagement, such as employing visuals when auditory stimuli dominate their experience, to prevent sensory saturation [90].

4.8.2 Content Generation. Utilizing context information from interesting moments can reduce the writing workload (i.e., the right branch in Figure 4). Writing covers various levels of abstraction, as suggested by various literature [22, 34], ranging from basic event descriptions to intricate details. Contextual data associated with these "interesting" moments can aid in crafting both the structural and detailed elements of a narrative. Specifically, sensory (such as FPV and audio), spatial, and temporal modalities (such as location and time) can help structure the narrative, addressing the WH questions (i.e., what, when, where, who, and how) of user experiences. Introducing more precise details into the narrative, e.g., by recognizing background elements like information boards or nearby tour guide speeches in a museum, can enrich the content and enhance its depth. In addition to these relatively objective writing components, participants emphasized the importance of personalizing narratives. To achieve this, participants recommended AI capturing in-the-moment emotional responses (e.g., expressions of astonishment) and encouraging them to reflect (e.g., by connecting with past experiences) through elicitation questions.

5 *PANDALENS* SYSTEM

We introduce *PANDALens*, a proof-of-concept system that aligns with the design considerations highlighted by the formative study. In this section, we first depict the usage scenarios of *PANDALens*, then detail the primary features of the *PANDALens* system, their underlying rationale through iterative design, and its implementation. Notably, while our current version addresses most user requirements, as a prototype, it has the potential for further refinement, especially when technologies such as OHMDs and AI become more powerful and practical.

5.1 Usage Scenarios of *PANDALens*

Consider Jane, the previously mentioned traveler, at the local natural history museum. Upon entering the museum, Jane is immediately captivated by a giant Rafflesia flower. She moves closer to the flower and takes a careful look. Detecting Jane's prolonged gaze, *PANDALens* automatically takes a photo of the flower. However, as Jane is engrossed with the flower, *PANDALens* delays prompting her with comment suggestions until Jane has finished enjoying the exhibit and moves on to the next one. When seeing the captured photo, Jane is satisfied and tells *PANDALens* that this is her first time seeing this flower. Analyzing Jane's comments along with the captured image and location information (as shown in **Figure 1**), *PANDALens* understands the context and then asks Jane questions for details, "May I ask if this is a real one or just a model?" To which Jane replies, "I suppose it should be real, as there was a 'do not touch' sign."

As a bird lover, when Jane passes by a wall full of bird specimens, *PANDALens* detects the birds in Jane's FPV and, considering Jane's personal interests, takes a picture and shows a 'like icon' to Jane,

indicating an invitation for comments. Jane expresses her delight at seeing so many birds on one wall, and *PANDALens* subsequently asks what species is her favorite. However, Jane finds something even more interesting at that moment—a complete whale skeleton. Thus, she ignores the question, which gradually fades away, and tries to find a good angle to take a picture. After finding an ideal spot, Jane uses the subtle ring interaction to capture a photo and converses with *PANDALens* about the story behind the whale (which died from a collision with a cargo ship). After the whale visit, when Jane goes to another exhibition and passes by the bird area again, *PANDALens* avoids auto-capture and commenting suggestions to avoid repeated suggestions. Later, when Jane finds a cute cat specimen and moves closer to it, *PANDALens* takes photos for Jane. A commenting invitation also shows up when detecting the simulated tiger sound in the museum. Similarly, Jane continues to enjoy the museums with her '*PANDALens* companion'.

After the visit, Jane asks *PANDALens* to generate a blog using the ring interaction. *PANDALens* first compiles a list of recorded moments for Jane to select the highlights to include in narratives. After Jane's selection, *PANDALens* creates a travel blog detailing Jane's experience at the museum with her reflections and interesting moments. Although Jane appreciates narratives with details, she also wants to share the trip on Twitter. Therefore, Jane asks *PANDALens* to convert the writing into a Twitter-style by condensing the content and adding emojis. Upon receiving verbal revision requests, *PANDALens* revises and presents the content on the OHMD. Once Jane confirms the final version, *PANDALens* outputs the narrative along with the captured images to Jane's laptop, enabling her to share it on Twitter.

5.2 Key Features of *PANDALens* System through Iterative Design

As demonstrated in the above usage scenarios (sec 5.1), the interaction flow of *PANDALens* encompasses three stages: (1) Capturing of Interested Moments: Using a mixed-initiative interface that seamlessly merges AI-driven and human-initiated actions. (2) Context-Related Questions Generation: *PANDALens* presents context-related queries by leveraging the multimodal information extracted from the captured moments. (3) Final Narratives Generation: After travel, *PANDALens* offers users the autonomy to select their favored captured moments. It then generates a draft document and enables revisions according to the user's preferences. In the following, we will introduce the functions of three major components in the ***Final System***. We will then elucidate how each component was refined through a one-round iterative design process.

5.2.1 Participants. Eight users (P1-P8) from the university were involved in the iterative design to test our *Initial System* and give feedback. We sought a varied participant group for comprehensive insights. Three regularly posted travel stories online, while five rarely did, finding editing too labor-intensive. Besides, the participant group comprised a UI designer and two HCI researchers.

5.2.2 Mixed Initiation for Moment Capture: AI Initiation. We incorporated a set of modalities tailored for travel scenarios from our formative study (sec 4.8) as a proof of concept to detect users' situational and personal interests [69, 74]. We also designed strategies to mitigate false positive suggestions and information overwhelming from the AI assistant.

Multimodal Analyzer for User Interest Detection. The system processes various modalities in real-time and concurrently to detect the two types of interests. To identify situational interests, the system recognizes positive sentiments, including joy and surprise, in user verbal expressions, given travelers mainly report positive experiences in travel blogs [18].

For personal interest detection, the system monitors two types of context information from the environment that match user preferences, objects within the FPV and background audio, to discern visual and auditory preferences. To quickly assess user visual and auditory preferences from a wide range of categories, we ask LLM to "Create an interactive questionnaire to narrow down two lists related to interest detection for the COCO dataset and MediaPipe Audio classification, based on the user's travel preferences." Based on users' answers, the LLM consequently formulates two lists of potential options, allowing users to narrow down their choices further.

Additionally, two triggers are utilized to detect both situational and personal interest, with optimization based on our pilot testing results. The first, Gaze Fixation, is detected when eyes remain focused on a small area, deviating no more than 4.91 degrees for at least 1 second. The second trigger, named "Zoom-In", activates when users approach an object closely while looking at it. This intent is identified by the target object size increases by 10% in two consecutive FPV frames.

Interaction Design. As depicted in Figure 5 (AI Initiative-Detect Interest), the AI suggests moment capture upon detecting user interest. Users can confirm such interest by verbally commenting, which triggers the system to auto-record. To overcome the uncertainty in AI decisions, we follow the mixed-initiative guidelines [4, 33]. If the user ignores the suggestion, it gradually vanishes, or users can dismiss it manually (by pressing the center button on the ring mouse).

Moreover, to mitigate distraction from AI suggestions, we adopted the following designs: (1) We utilized the principles of "matching attentional draw with utility" [29, 90] for notifications. For instance, audio notifications attract more attention [27]. Thus, initial invitations of moment commenting are only conveyed through subtle visual feedback (e.g., icons [17, 41]), and audio notifications are only enabled after confirming the user's interest. (2) To facilitate user concentration on primary activities while still being subtly aware of digital alerts, we situated the visual notifications within peripheral vision [17, 40]. Additionally, we employed higher inter-line text spacing [96] to enhance the text readability during mobility, as shown in Figure 1 and Figure 5. (3) To prevent users from being bombarded with notifications, we limit the frequency of sending the same type of notification. Specifically, we set a minimum interval for the same type of suggestions within a similar FPV ($threshold = base\_threshold\ (15s) + (FPV\_similarity)^2 \times threshold\_factor\ (200s)$). (4) To ensure users remain engaged in the present experience, certain notifications are deferred [5] to less interesting moments. For instance, gaze fixation-based suggestions are deferred until a transition [32] during the trip.

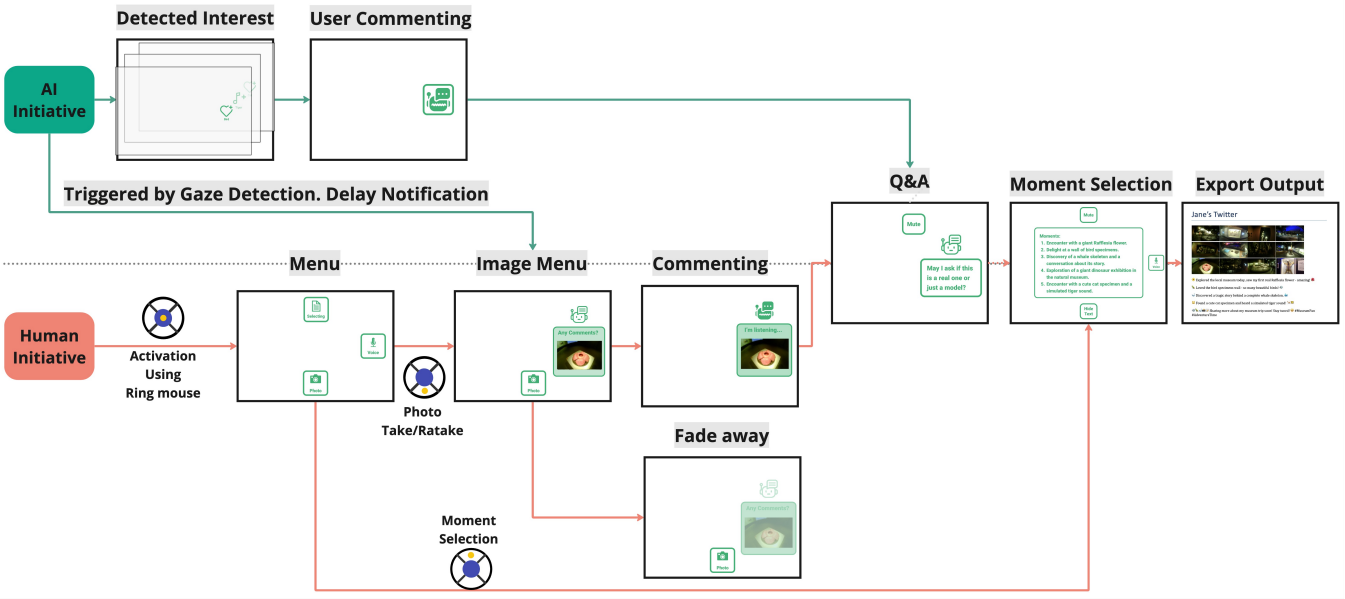


Figure 5: Interaction flow of *PANDALens* system. It includes both AI-initiative and Human-Initiative interactions. The ring mouse for human-controlled interaction is also shown (yellow dots presenting button clicks). Note: Icons are re-scaled to make the figure clear.

Iterative Design. There is no difference in interest detection triggers between the *Initial System* and the ***Final*** *System*. However, after detecting potential interest, the *Initial System* prompts both visual (e.g., photo) and audio notifications (e.g., saying, "Do you have any comments?") to users. Participants (P1, P2, P3) found this sometimes distracting, particularly when context analysis models generated false-positive errors. Thus, we adopted the "matching attentional draw with utility" principle (omit audio feedback for commenting invitations) and limited the frequency of similar notifications to minimize distraction from AI suggestions.

5.2.3 Mixed Initiation for Moment Capture: Human Initiation. We incorporate human initiation using subtle ring interaction [17, 72] to complement AI initiation, especially when AI might not detect user interest. Adopting the attention-maintaining interface design of *ParaGlassMenu* [17], our design enables users to remain engaged in their travel activities while leveraging their peripheral vision for menu manipulation. By default, the menu is hidden to reduce disruptions. As illustrated in Figure 5 (Human Initiative-Menu), users can activate the menu by pressing the center button on the ring mouse. Following natural spatial mapping guidelines [58] to minimize cognitive effort, users can utilize the up, down, and right buttons to generate final writing, take photos, or record voice comments, respectively. Moreover, proficient users can snap photos directly via the ring mouse's down button as a shortcut, bypassing menu activation. Options for photo retakes are provided, enabling refining captures. Once a photo is taken, the system displays a notification consisting of an image and comment invitation. Mirroring the AI initiation process (sec 5.2.2), the system anticipates user voice commenting, and fades the notification if left unattended after 8 seconds.

Iterative Design. Two differences exist between the *Initial System* and the ***Final*** *System* regarding human initiation. First, the *Initial System* required prior activation for any action, deemed cumbersome by users (P4, P5, P6). Thus, the ***Final*** *System* introduced shortcuts for photo-taking, bypassing menu activation. Second, the *Initial System* did not support capturing new moments while processing previous requests, causing missed fleeting moments (P3, P6). The ***Final*** *System* thus allows photo-taking during processing and adds them to a pending list for later commenting.

5.2.4 Processing of Interested Moments. Upon user confirmation of interesting moments via comments, *PANDALens* transcribes the user's voice into text. As shown in Figure 6 (Contextual Information), these transcriptions are then sent to the LLM, enriched with additional contextual modalities in textual formats using various AI models (detailed in Table 1), including image descriptions, identified objects' labels with confidence scores, text recognized from images (OCR), and information about time, location, and background audio category. This facilitates 1) presenting context-relevant questions to users for inspiration and 2) creating a concise moment summary that eventually contributes to the final narratives.

Context-Related Questions for Inspiration. Leveraging the aforementioned multimodal context, the LLM employs a predefined prompt to pose context-specific questions tailored to the user's preferred style (e.g., 'direct and specific to the current moment,' 'question links to memories,' etc.). User preferences regarding question formats are pre-configured (Figure 6-green highlighted parts) and summarized by another LLM model, which first queries users for their preferred style and offers several examples for decision support when user preferences are unclear (e.g., one of the question style examples provided by the LLM is: *Specific and Detailed: "Can*

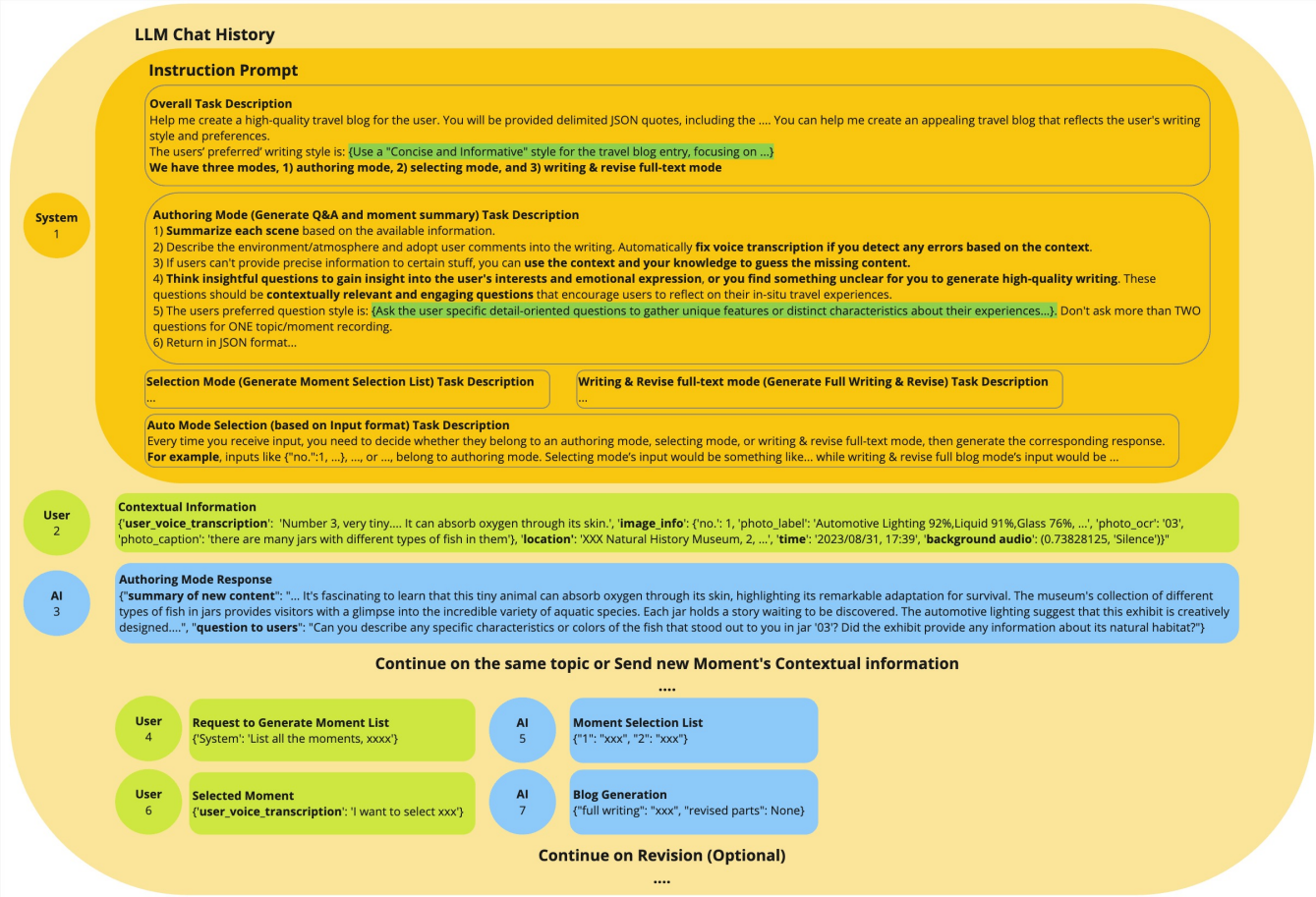


Figure 6: LLM Chat History for the *PANDALens* system. 'System' represents the initial prompts directing the LLM's tasks. In the Instruction Prompt, sections highlighted in green are customized parts tailored to individual preferences and generated by another LLM. 'User' and 'AI' signify the inputs and outputs within the LLM dialogue, respectively, with message sequencing indicated numerically. Note: Some details are redacted to conserve space. Details can be found in Appendix B.2.

you describe the flavors and aromas of your coffee? How did they contribute to your overall experience?"). It operates using a separate prompt, as detailed in Appendix B.3.

To balance inspiration and potential distraction, we limit the number of questions posed for each moment to two, as suggested by users. Regarding the notification modality of these questions, our system integrates both automatic and manual toggling between auditory and visual feedback to ensure a balance between noticeability and minimal distractions. Automatic modality toggles are environmentally dependent; for instance, a scene with many nearby individuals in the FPV prompts auditory rather than visual feedback to preserve the user's visual focus. Concurrently, manual modality adjustments using the ring mouse, such as muting or unmuting notifications, are also available.

Prompt Design: Processing Interesting Moments for High-Quality Questions and Final Narratives. Figure 6 demonstrates the interaction flow with LLM to generate questions and final narratives. To ensure a comprehensive understanding of user travel experiences, interactions with LLM maintain the chat memories, including previous contextual and Q&A details in the same travel session. However, two primary issues were encountered during LLM data processing: 1) the LLM asked irrelevant questions due to overlooking important context that contains unclear or erroneous information (e.g., voice comments with errors like 'Seeshell Potoms' ['Seashell Patterns']), and 2) it produced unsatisfactory final narratives from lengthy, unstructured chat histories (e.g., voice transcription errors preserving in final narratives while user elaborations on questions are missing). To mitigate these challenges, we iteratively refined (Appendix B.1) the Instruction Prompt for LLM, as shown in Figure 6.

To address the first challenge, the refined prompts require the LLM to correct inaccuracies using multimodal information before generating context-relevant questions (detailed in the Authoring Mode Task Description in Figure 6). This approach reduced unsatisfied questions and enabled a more accurate understanding of the environment and user intentions. For example, instead of ignoring 'Seeshell Potoms,' the refined prompt enabled the LLM to accurately

understand it with the museum's multimodal context and inquire about captivating aspects of the seashell pattern.

To address the second challenge, we adopt an approach similar to Chain-of-Thought [87]. Rather than prompting the LLM to generate final narratives directly from an unstructured chat history, the prompt first instructs the LLM to craft a summary for each distinct moment, accompanied by every in-situ question generation. These summaries can be dynamically enriched or corrected based on users' responses regarding specific moments. For example, a moment summary first accurately recorded a plant name as 'Rafflesia' instead of 'Raising' from voice transcription and then updated details on how the plant's structure enables its regeneration after a fire disaster using user responses to questions. Ultimately, the LLM model generates the final narrative using these refined momentary summaries.

Iterative Design. In the ***Final*** *System*, we improved both the interaction experience and question quality based on user feedback. The *Initial System* showcased live voice transcriptions to users, but this feature proved distracting, especially with transcription errors. Many users (P1, P2, P3, P4, P8) spent more time correcting errors than enjoying their travel experience. Given that participants (P2, P3, P7, P8) observed minor transcription errors that didn't hinder LLM's comprehension, especially with other rich multimodal contexts available, we removed the live voice transcription display in the ***Final*** *System*. Furthermore, the inability to mute and constantly ask questions until the user stops responding annoyed several participants (P3, P5, P7). To address this, we introduced a 'mute' feature in ***Final*** *System* and limited the number of questions per moment to two, as recommended by the users. As for question quality, two participants (P4, P7) reported that the questions were too general and didn't reflect their preferences. Thus, we introduced personalized prompts in the ***Final*** *System* to allow users to customize their preferred question style.

5.2.5 Generation of Final Writing. Post-travel, users can compose their final narrative by selecting which captured moments to incorporate (see Figure 5, Moment Selection and Export Output). The LLM provides a concise summary for each recorded moment, facilitating users in choosing different moment combinations for diverse narratives. After moment selection, the LLM crafts the complete narrative based on a predetermined personalized prompt (Figure 6-green highlighted parts)[6]. Recognizing that preferences may change over time, users can modify the writing style or make other narrative adjustments through voice commands. Ultimately, the system offers the final draft narrative in Microsoft Word format, facilitating various sharing options, including social media posts. In addition, to satisfy the comprehensive reviewing needs, the system attaches all the moment summaries to the end of the documentation.

Iterative Design. Two major changes were implemented in the ***Final*** *System*. Firstly, while the *Initial System* generated narratives based on the entire chat history, two participants (P1, P8) requested the ability to create multiple blog variations with different moments to suit various sharing purposes. Thus, we introduced moment selection in the ***Final*** *System*. A full log of all moment summaries was also added to the documentation for comprehensive reviewing needs (P7). Secondly, the *Initial System* did not accommodate users' preferences for generating styled narratives. As a result, two participants (P4, P7) noted that the final draft narratives, while formal, sometimes appeared templated and did not reflect their personal styles. Therefore, similar to customizable question styles, we incorporated personalized prompts for writing style into the LLM pipeline in the ***Final*** *System*.

[6]Mirroring the approach for setting question preferences, user preferences for writing styles are preconfigured using an LLM with a separate prompt, detailed in Appendix B.3.

5.3 Implementations

The *PANDALens* system is developed with the OHMD, XREAL Air[7], for a near-eye display and uses the Pupil Core add-on for gaze detection and FPV streaming. Built on a TKinter-based UI and a Python backend, it seamlessly handles the real-time capture and concurrent processing of various context data and user interactions. Due to computational constraints, our choice of context analysis models aimed to balance performance and efficiency, especially in mobile scenarios without constant power sources. We employed the GPT3.5-Turbo-16K model as the primary LLM to generate context-specific questions and structure narratives. Few-shot prompts (i.e., Auto Mode Selection in Figure 6) enabled LLM to discern whether to generate questions, compile a moment selection list, or create a full blog based on the input format, and additional prompt engineering techniques [70, 87] were utilized to enhance its output. To address the LLM's token limitations, our system compresses chat history into summaries, facilitating longer documentation sessions. Detailed prompt information can be referenced in the Appendix B. Comprehensive implementation details can be referred to in Table 1 and Appendix C.

6 COMPARATIVE STUDY

To evaluate the overall efficacy and user experience of *PANDALens* for in-context writing during travels, we conducted a comparative study with *LiveSnippets*, a smartphone application for in-context writing, during a realistic museum visit.

Introduced in 2020, *LiveSnippets* [44] is an application that allows users to capture photos and record verbal comments on their smartphones. Unlike the traditional approach, *LiveSnippets* encourages in-context content creations using voice-based multimedia input and stores them in multi-modal "snippets" (i.e., images with voice transcriptions, locations, and timestamps). The snippets can be rearranged to form multimedia articles and published with light copy-editing.

LiveSnippets shares many common features with *PANDALens* in leveraging voice-based input and automatically tagging multimodal metadata when users take photos. However, the primary distinction is that it does not infer users' interests by observing their behavior; instead, it employs a user-initiative approach to moment capturing rather than a mixed-initiative approach. Furthermore, *LiveSnippets* lacks the assistance of LLMs. Despite these differences, *LiveSnippets* stands out as the closest in-context writing tool we have identified, making it an ideal candidate for exploring the impact of these feature variations on the user experience.

[7]https://www.xreal.com/air

Table 1: System Components and Associated Technologies. The links to these tools are in Appendix C.

Component	Description	Associated Technologies/Tools
PANDALens	Main system developed for the application.	Python 3.9
OHMD UI	Interfaces built on a laptop for near-eye display.	Tkinter
Pupil Core	Facilitates gaze detection and FPV video streaming.	Socket connection in Python with Pupil Capture App
Multimodal Analyzer	Analyzes multimodal context data concurrently and integrates contextual information in JSON format.	1. Object Detection & OCR: YOLO v8, Google Cloud Vision API 2. Image Description: BLIP-large on Hugging Face 3. FPV Similarity: OpenCV 4. Audio Classification: MediaPipe 5. Voice Transcription: Whisper 6. Tone Analysis: Emotion English DistilRoBERTa-base model 7. Location: Geopy, Geocoder. 8. Time: Python's Datetime.
LLM Model	Processes context data to provide questions and assist writing.	GPT3.5-Turbo-16K (temperature value: 0.3)
Prompt Engineering	Ensures efficient task performance and seamless integration.	1. Clear and Specific Instructions, 2. Few-shot prompts, 3. JSON formatted responses, 4. Chain-of-Thought approach

6.1 Participants

We recruited 16 participants (8 females, 8 males; mean age = 23.4 years, SD = 2.8) from the university community with self-reported professional English proficiency. None of them had participated in our prior two studies. To ensure compatibility with OHMD, all participants had no visual or auditory impairments (contact lenses were allowed). Their travel-sharing habits varied: eight participants regularly shared experiences more than twice monthly on social media or personal blogs, five shared 1-2 times per month, and three infrequently shared. This allowed us to assess how users with different sharing experiences perceive the *PANDALens* system. Additionally, twelve participants were first-time visitors, while four had visited before, providing insights from both novel and revisit experiences.

6.2 Apparatus

For the *PANDALens* system, as depicted in Figure 1, participants wore XREAL Air glasses with the Pupil Core addon for eye tracking. The XREAL Air glasses were connected to a laptop (Macbook Pro 14-inch, M2 Pro chip) running the AR system and logging user interactions. They held a Sanwa ring mouse (400-MA077) in their dominant hand for menu control. Participants also carried a light backpack to house the laptop during the trip. For the *LiveSnippets* system, participants used a phone (Mi A1) and a light backpack with the same laptop to mitigate confounding factors.

An accompanying experimenter, maintaining a distance to avoid interference, recorded participants' behaviors from a third-person perspective (TPV) using a mobile phone (iPhone 12).

6.3 Study Design

We employed a repeated-measures within-subject design for this study. Each participant experienced two travel sessions using the *PANDALens* and *LiveSnippets* systems. The order in which they used these systems was counterbalanced. The museum was divided into two areas, each containing an equal number of exhibitions (as illustrated in Figure 2a). The visiting order for these two areas was fixed for all participants, ensuring that each system was used in both areas.

6.4 Tasks and Procedure

The primary task for participants was to immerse themselves in the travel experience with no restrictions (e.g., no time limits). The secondary task required them to record any interesting moments based on their personal preferences. After the trip, participants could edit and refine the generated writing drafts from both systems until they were satisfied.

Following prior research [44], we divided the experience associated with each device into two phases: 1) travel while capturing interesting moments and 2) post-trip editing. After signing the consent forms, participants received a briefing about the experiment. They then spent approximately 10 minutes on the pre-study setup. This included specifying question and writing style preferences and selecting interested audio/visual object types. Subsequently, they underwent a 15-minute training session for each system, ending with creating a sample writing. Once participants were familiar with the system, they spent 45-60 minutes exploring specified locations using the designated interface. Post-exploration, they completed questionnaires about their travel and system interaction experiences. After a 5-10 minute break, participants continued on a second travel session with the alternate system for another 45-60 minutes.

Upon completing both travel sessions, participants went to a lab room with computer access for the post-editing. In the lab room, they generated writing drafts for the two systems and then proceeded to edit the content further, either on the two systems or on the computer, with the same order of exploration. For both systems, participants were also free to leverage any tools to assist their writing [88], including AI tools (e.g., ChatGPT, Grammarly), which helps us understand if the merits of *PANDALens* were not simply due to using LLM for polished narratives. Although there was no strict time constraint, participants typically took 5 to 20 minutes for editing. Afterward, they completed a questionnaire reflecting on their writing experience, followed by a 15-minute semi-structured interview discussing their overall travel and writing experiences.

The entire experiment took approximately 180 minutes, including two 5-10 minute mandatory breaks to avoid fatigue.

6.5 Measures

6.5.1 Quality of Content Generation. For objective measures, we collect the data of post-editing time (min) and *Word Count*. For subjective measures, self-rated *Post-Editing Effort*, quality of *Language*, *Creativity & Appeal*, and meeting preferred *Writing Style* for final writing (7-point Likert scale) [49] were collected. Additionally, we measure self-rated scores on *Control over Content* and *Trust in Content* [20] to evaluate authenticity (7-point Likert scale). Moreover, *Self-Rated Writing Score* for overall final writing satisfaction (out of 100) was collected to encompass several aspects of content quality from a holistic view.

6.5.2 Quality of Travel And Moment Capture. For objective measures, system usage counts, including the count for photos and comments, are collected. Specifically for the *PANDALens*, we track how often users find AI suggestions helpful and the number of suggestions that lead to user comments.

For subjective measures, we collected *Travel Enjoyment* scores using a 7-point Likert scale. Recognizing that smartphones currently offer more advanced hardware than smart glasses, which might influence the present travel experience, and anticipating that OHMD will be improved in the future [1], *Travel Enjoyment* scores with both existing and anticipated enhanced devices (for example, more lightweight devices) were collected. For the effectiveness of moment capture (i.e., whether photo-taking and commenting with the system can well support documentation needs), we evaluate *Writing Productivity* aided by the system's moment capture ability and efficacy in *Ideation Support* (7-point Likert scale). Interaction *Distraction* and *Naturalness* to primary task (i.e., travel) [17] (7-point Likert scale), usability (*SUS* [13]), and perceived task load (*RTLX* [30]), were collected to evaluate interaction quality of moment capture in the travel. Lastly, we collected users' *Familiarity* with both devices using a 7-point Likert scale.

6.5.3 Analysis. A paired-sample T-test or Wilcoxon signed-rank test was used depending on the data's characteristics. Descriptive statistics were also used to analyze the results. The interview recordings were transcribed and thematically analyzed following Braun and Clarke's guidelines [12].

6.6 Results

Users spent an average duration of 47.3 minutes (SD=5.7) in the museum with *PANDALens*, while they spent an average of 45.4 minutes (SD=3.3) with *LiveSnippets*. Overall, *PANDALens* significantly improved the quality of content generation with reduced user efforts. Regarding travel and moment capture experience, *PANDALens* enabled a higher travel experience with more moment capture and comments made, significantly improving users' in-context documentation. Notably, *PANDALens*' interaction quality of moment capture was comparable to *LiveSnippets*, although participants were significantly less familiar with OHMD than smartphones (3.13 vs. 6.50 out of 7, $p < 0.001$). Participants' blog-sharing and revisiting experiences didn't influence interaction behavior, and they all valued high-quality documentation.

6.6.1 Quality of Content Generation. As illustrated in Figure 7, *PANDALens* significantly outperformed ($p < 0.05$) *LiveSnippets* in terms of *Language*, *Creativity & Appeal*, *Writing Style*, and achieved a significantly higher *Self-Rated Writing Score* (82.19 vs. 60.88, $p < 0.001$). Moreover, it significantly reduced ($p < 0.001$) post-editing time (5.26 min vs. 17.93 min) and effort (2.50 vs. 5.88 out of 7), while *Word Count* between the two systems did not exhibit any significant differences.

Effectiveness of PANDALens' LLM Pipeline. The better performance of *PANDALens* can be largely attributed to its integrated LLM pipeline, including personalization and multimodal context information transformation, which empowered participants to receive well-crafted blogs from the system via bite-sized interactions during their travel. For *LiveSnippets*, while participants were also allowed to use AI tools, including LLM (e.g., ChatGPT), for post-editing, only six participants used such tools and encountered challenges. These ranged from uncertainties about using prompts to receiving unsatisfied output when merely inputting their voice transcriptions from the *LiveSnippets* system to LLM.

Authenticity of Generated Content. Although there was no significant difference in *Control over Content* between the two systems, participants had a significantly higher ($p < 0.05$) *Trust in Content* with *PANDALens* compared to *LiveSnippets*. This is largely due to more transcription errors shown in *LiveSnippets*' output document.

6.6.2 Quality of Travel And Moment Capture. As illustrated in Figure 8, *PANDALens* outperformed *LiveSnippets* in terms of overall travel enjoyment (especially with envisioned future enhanced OHMD with lightweight and transparent lenses ($p < 0.05$)). *PANDALens* also significantly ($p < 0.05$) enabled users to capture more moments (34.13 vs. 28.63) and in-context comments (30.00 vs. 17.38), as shown in Table 2, with comparable interaction quality (*Distraction*, *Naturalness*, *SUS*, *RTLX*) with *LiveSnippets*. Additionally, participants rated that capturing moments using the *PANDALens* system significantly improved the quality ($p < 0.001$) of *Writing Productivity* and *Ideation Support*.

Mixed Initiative Interaction for Moment Capture. As demonstrated in Table 2, *PANDALens* facilitated a higher total moment capture count (increase 19.2%), yet with fewer manual interaction initiations (reduce 31.7%) compared to *LiveSnippets*. This suggests that AI-driven initiations significantly reduce user efforts required for moment capture initiation, providing more seamless interactions during travel. Regarding AI initiation, suggestions activated by gaze-related actions, such as Zoom-in (approaching a specific object while looking at it) and Gaze Fixation, emerge as the most effective triggers, as evidenced by their highest acceptance rates. As elaborated by P2, most scenery locations are rich in visual attractions, maximizing the benefits of gaze modality, where users predominantly direct their gaze to indicate what they find most exciting.

It's worth noting that although not all AI suggestions were accepted (overall acceptance rate: 67%) during the journey, most participants (14/16) found unused AI suggestions non-intrusive due to the ease of ignoring the irrelevant notification. Intriguingly, four participants noted that AI suggestions prompted them to reexamine moments they initially dismissed, leading to newfound interest.

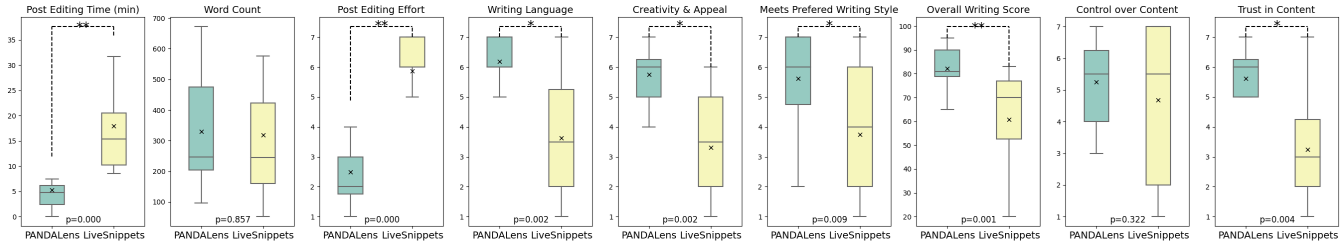


Figure 7: Boxplots for both Objective and Subjective measures of the quality of content generation. × inside box plot represents the mean value point. * indicates significance of $p < 0.05$, and ** indicates significance of $p < 0.001$. See the statistics detail in Appendix D-Table 3.

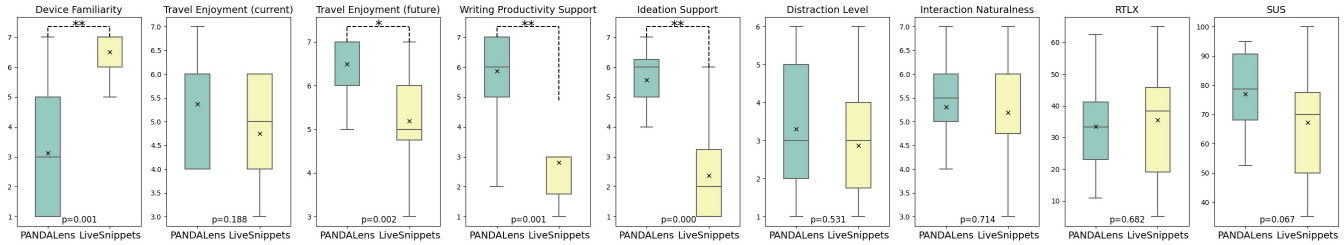


Figure 8: Boxplots for Subjective measures of the quality of travel and moment capture. × inside box plot represents the mean value point. * indicates significance of $p < 0.05$, and ** indicates significance of $p < 0.001$. See the statistics detail in Appendix D-Table 4.

Table 2: Mean system usage counts per participant for *PANDALens* ([P]) and *LiveSnippets* ([L]). For the *PANDALens* system, we present interaction counts initiated by both AI and User. Specifically for AI-initiation, we present Acceptance (rate), indicating how many AI suggestions were perceived as useful by participants (either for helping with taking photos (final kept) or leading to making comments). For *LiveSnippets*, we present interaction counts by User, and the Acceptance (rate) represents how many manual photo-taking actions resulted in final kept photos or making comments. We also present Comments (rate), indicating how many interactions lead to making comments.

PANDALens Interaction	Total	Acceptance	Comments	Acceptance Rate	Comments Rate
[P] AI: Zoom-in	2.69	2.56	1.56	0.95	0.58
[P] AI: Gaze Fixation	4.13	3.75	1.94	0.91	0.47
[P] AI: Object in FPV	9.06	6.75	2.31	0.74	0.26
[P] AI: Ambient Audio	5.06	0.81	0.81	0.16	0.16
[P] AI: Positive Tone	0.94	0.69	0.13	0.73	0.13
[P] AI-initiation	21.88	14.56	6.75	0.67	0.31
[P] Manual Photo	19.56	19.56	12.31	1.00	0.63
[P] Total Initiation	41.44	34.13	19.06		0.46

PANDALens Interaction	Total	Acceptance	Comments	Acceptance Rate	Comments Rate
[P] AI: Context-Related Questions	23.06	10.94	10.94	0.47	0.47

LiveSnippets Interaction	Total	Acceptance	Comments	Acceptance Rate	Comments Rate
[L] Manual Photo	29.25	28.63	17.38	0.98	0.59

Most participants (14/16) also valued the extra automated photos for post-trip review, noting the potential to find unnoticed highlights during the travel (P7, P9).

Experience (Passive Tool vs. Proactive Mate). Participants also made more comments when using *PANDALens* with increased travel enjoyment, as "it [*PANDALens*] converts the travel experience from a passive tool to proactive interactions (P14)", and makes it more like a "friend (P15)", which validates the user expectation of "travel mate" in the formative study. As P3 elaborates, this was

largely due to the context-awareness suggestions and dialogue-based Q&A of the system, "it's like somebody who asks me what you have seen in the museum. I would have forgotten to say this before [with my smartphone], but now it's natural to speak out my feelings". P14 further addressed these benefits for solo travelers, "I think AI-initiated conversations are actually a good way for solo visitors. If I visit a museum alone, I want to chat with someone. This [*PANDALens*] is a very good medium to record your thoughts at that moment."

Emerging Technology vs. Status Quo. While most participants (15/16) preferred using *PANDALens* for in-context writing during travels, it is important to note that one user (who frequently uses smartphones for Vlog creation) still favored the traditional smartphone interaction. This preference was largely due to the immature hardware of OHMD and their familiarity with phones. Moreover, participants highlighted that smartphones provide a more intuitive method for framing photos, yielding comparable interaction quality in metrics such as *Distraction*, *Naturalness*, *SUS*, and *RTLX*.

Despite their relative unfamiliarity with OHMD, participants who preferred *PANDALens*' moment capture experiences emphasized its natural and less demanding feel, as "It's more convenient and natural for me to simply click the ring mouse to take a photo or directly accept the AI suggestion and comment. But when I use a phone, it's tiring to always have it in hand [to prepare to take photos] (P5)", and "It's awkward for me to bring the phone close to my mouth and speak to its microphone in public. But *PANDALens* offers a more natural speaking experience (P3)". This indicated ease of *PANDALens* usage, leading to a higher average *SUS* than *LiveSnippets* (77.03 vs. 67.19), indicating 'Good' usability [8].

6.7 Discussion

6.7.1 Balancing AI Suggestions with Distraction Management. Detecting personal and situational interests alleviated participants' moment capture burden significantly. This approach, proven in museum explorations, is also seen as beneficial in other travel scenarios, "When watching the Air Show, using my phone to take pictures was NOT convenient as I was holding something else. I envision the glasses can be the perfect mate who knows [what] I like [in] Air Shows, like jets' sounds (personal interest), and auto-record when I look at the sky (situational interest). (P1)"

However, interest detection inevitably introduces false positive suggestions (as evidenced in Table 2), leading to potential distractions. Our ***Final*** *System* significantly improved this aspect when comparing its feedback with our *Initial System*. This aligns with literature suggestions [4, 5, 29, 40, 46, 90], highlighting the need for AI suggestion designs on OHMD to balance notification delivery with user status. Firstly, suggestions should be minimized and carefully managed during primary tasks through frequency control and position and output modality selection. Secondly, the information delivery has to adapt to the user's attention level to primary tasks (e.g., delay notifications during immersive observation), interesting level to the side notification (e.g., less potentially interesting notifications are non-intrusive), and user-situated environment (e.g., switch output modality considering the complexity of environment). Such designs (sec 5.2.2 and 5.2.4) ensure a balanced interaction, minimizing distractions while maintaining user engagement.

6.7.2 Improvement of Content Quality from In-Situ Human-AI Interaction and Multimodal Information. Our study demonstrated that in-situ human-AI interaction and effective multimodal context utilization significantly enhanced content richness (moment capture and in-situ comment quantity) and expressiveness (language, creativity & appeal, and preferred writing style). In particular, in-situ human-AI interaction, incorporating mixed-initiative interaction and in-situ Q&A, increased moment capture by 19.2% and comment amount by 72.6% with more details. For instance, P7's comments with *LiveSnippets* were more basic: "There are different types of moss, but they look very similar. It's very interesting." In contrast, prompted by context-related questions from *PANDALens*, P7 responded more elaborately: "The fossil's sheer size was astonishing, as we never get a sense of dinosaur fossils' scale from books. I'll share this with the audience to give them a sense of scale, like how the dinosaur's torso reached the top of my head."

Moreover, results validated that utilizing multimodal context information with LLM offloaded participants' burden of context description, shifting their focus to personal and deep expression that makes blogs unique. This necessitated prompt designs that effectively manage and interpret multimodal data, even in the presence of potential errors. (sec 5.2.4). For instance, P16 commented on seashells arranged in a circle: "This is an example of interspecific variation." Despite the image description model mislabeling the exhibit as a clock display with keys, *PANDALens* effectively directed the conversation towards a biological theme by analyzing multimodal context. This conversation prompted P16 to discuss "species adaptations and chronological gradation", enhancing the accuracy and descriptiveness of the moment summary. Transforming the enriched summary into the user's preferred style further increased content expressiveness.

Thus, *PANDALens* improved overall *Self-Rated Writing Score* with rich and expressive content, in contrast to merely utilizing *LiveSnippets*' voice transcription with LLMs, which lacked detailed content and effective error correction.

6.7.3 Balance between AI-Augmented Narratives and Content Authenticity. While *PANDALens* resulted in more trustworthy and cohesive narratives, more attention is still needed to these narratives' authenticity.

(1) *User Preferences and AI Interpretation.* While personalized prompts are used to tailor *PANDALens* narratives to user preferences, occasional dissatisfactions with initial drafts were noted (3/16), mainly due to two reasons. Firstly, the challenge of articulating precise needs in a single attempt leads to misalignment between user expectations and AI interpretation. For example, participants seeking a "direct and concise" style sometimes received "travel reports" without emotional depth. Secondly, user preferences change with context; while some initially favor detailed narratives when few moments are involved, they may later opt for a more succinct style suitable for platforms like Twitter, especially when encompassing many moments. This highlights the importance of human intervention in post-editing, as users often identify their exact needs only after reviewing the draft. In our study, participants

found adjusting the style through post-travel voice commands effective for refining the narrative to their latest preferences.

(2) *Dichotomy of Hallucination.* As expected, hallucinations showed in *PANDALens*' generated content, although they were minor reported by participants (8/10 who met such issues), compared to *LiveSnippets*. Interestingly, hallucinations in *PANDALens* exhibit a dichotomous nature. 'Intrinsic' hallucinations that contradict the fed context source [42] can distort the user's authentic voice, with inaccuracies in voice transcription and image interpretation exacerbating this issue. Conversely, some participants retained 'external' hallucinations, which cannot be verified against the fed context source [42], because they accurately predicted unreported but real experiences (e.g., "visited a dinosaur skeleton") and emotions (e.g., "want to eat seafood" when mentioning crabs). These 'external' hallucinations might originate from the LLM's external training data [42] when contextual information (e.g., location) is provided. To balance the two sides of hallucination, 'moment selection' could effectively filter out unsatisfactory recordings, preventing inaccuracies from compounding during full blog generation. Participants also recommended enabling local editing of each entry with LLMs through ring and voice interaction for finer control in the moment selection list.

6.7.4 Supporting Users with Different Writing Purposes. *PANDALens* effectively supports two documentation purposes, experience sharing and memory preservation, as identified in the Formative Study. For the fifteen users with experience-sharing needs, *PANDALens* meets their requirements well by allowing the generation of content with a format and tone suitable for various social media platforms. Non-native English speakers also noted, "it could facilitate global interactions by sharing good-quality writings on English social media (P8)". Three participants with only personal memory needs, who prioritize comprehensive recording with highlighted moments over writing quality, highly appreciated *PANDALens* could export all moment summaries. This fully logged moment summary complements the travel blog that focuses on cohesive content and may occasionally overlook details important to users.

7 OVERALL DISCUSSION

With the current results, *PANDALens* has successfully created an initial version of our vision: a wearable AI assistant that collaborates with travelers to enhance in-context travel experience documentation with improved travel enjoyment. This success is primarily attributed to two factors: 1) the mixed-initiative interactions that transform the passive tool experience into a proactive travel companion, and 2) the LLM pipeline with multimodal context analysis significantly reducing the user effort required to create high-quality documentation.

To achieve our ultimate vision, a wearable AI assistant that proactively utilizes various context information and generates documents across various scenarios, we highlight the features that future systems should emphasize and further enhance, including the need to transform tools into companions, support bite-sized interactions to reduce user efforts, and effectively utilize multimodal context information. Consequently, we discuss the next steps required to achieve an ideal wearable AI assistant that can enhance in-context content creation during daily activities.

7.1 From Passive Tool to Proactive Travel Companion

Unsurprisingly, by introducing AI to observe contextual information from user behaviors and the environment, *PANDALens* provided proactive suggestions, prompting users to capture more moments and provide additional comments. However, the goal of *PANDALens* is not to create a machine that constantly nudges users during their travels. Instead, it is essential to carefully design this proactivity to feel organic, akin to a "travel companion," as described by participants, ensuring that users enjoy their journey while more effortlessly documenting rich experiences.

7.1.1 Providing Suggestions Like a Companion Who Knows the User Well. Previous research has highlighted the importance of human travel companions, as they assist solo travelers in overcoming internal constraints, such as accessibility to information and the discovery of interests [92]. Additionally, these companions play a role in positively influencing a tourist's emotions, and higher supportive ability can further amplify these emotional benefits [78]. Compared with uni-interest detection, by analyzing both situational and personal interests based on multimodal contextual information, *PANDALens* appears as an intelligent "digital companion" that is well familiar with the user. This offers more effective support in overcoming constraints, such as assisting with photography and pinpointing intriguing moments during travel. In particular, personal interest detection positions the system like a close friend, aware of the user's preferences. On the other hand, situational interest detection behaves like a travel buddy, observing the user's behaviors and resonating with their in-the-moment reactions.

7.1.2 Enabling Engaging Dialogue with a Companion. Besides being a friendly companion assisting users in overcoming internal constraints during travel, most participants (13/16) appreciated engaging in dialogue with *PANDALens*. This interaction not only alleviated the monotony often experienced by solo travelers but also naturally elicited deeper reflection. We observed two strategies in *PANDALens*' usage that help to enhance the dialogue experiences.

(1) *Engagement from Bidirectional Voice Communication.* While *LiveSnippets* enabled users to voice out comments during photo capture on the phone, users still felt it "awkward and unnatural," especially in public settings. In contrast, using voice-based interaction in Q&A dialogues with *PANDALens*, users could naturally build a bidirectional communication with a "travel companion", creating a more interactive atmosphere than voice-out thoughts in a one-sided manner using phones. In an environment with many visual attractions, vocal prompts from *PANDALens* tend to elicit more spontaneous responses and reduce cognitive load compared to text-based prompts [55, 80].

(2) *Engagement from Tailored Questioning.* Contrary to tools that frequently suggest templated questions [44], *PANDALens* adjusts its inquiries using LLMs, resonating more like a "companion" based on various contexts and user preferences. This approach aligns with Micro-Phenomenology [54, 64], a method that deeply explores users' micro-moments in intricate detail. Such an approach has proven to help users find ordinarily inaccessible dimensions of lived experience with high accuracy and reliability [63]. For example, *PANDALens* spotlighted P15's mention of a dry leaf, connecting

it to childhood leaf preservation projects, increasing user interest and willingness to share. Moreover, the adaptive questions make interactions fresh and unpredictable, further enhancing the travel's engagement (P3).

7.1.3 Avoiding Distraction and Enhancing User Autonomy. Although *PANDALens* can act as a supportive companion that identifies interest and reduces boredom through context-relevant dialogues, careful designs are needed to avoid its potential annoyance to users.

Beyond attention management strategies that consider information filtering, presentation, and output modality selection based on user attention and context awareness (sec 5.2.2 and 5.2.4), future systems can introduce more advanced algorithms that interpret fine-grained human intent to increase system reliability, making it a more intelligent "companion". Although gaze is identified as an effective modality to disclose interests in the comparative study results (sec 6.6.2), it failed to distinguish between the interest to "gain knowledge" and the intent for "documentation and sharing." For instance, two participants observed information boards in a museum to acquire knowledge but had no intention to record. In such scenarios, gaze fixation would erroneously suggest recording due to its inability to discern these subtle differences, making participants reject such recommendations. To address this, a potential remedy is integrating multiple information modalities [10, 77]. For example, merging gaze data with semantic content from FPV [19] and Electrodermal Activity (EDA) signals [81], combined with user interaction history, could refine intention predictions.

7.2 Enhancing In-Context Writing with LLM

Although existing literature recognizes the efficacy of LLM in content creation [57, 88], our research indicated the importance of carefully managing prompts and content generation for optimal outcomes.

7.2.1 Leveraging Bite-Sized Interaction with LLM for Efficient Content Creation. Overall, a micro-task-based approach [35, 43] that gradually feeds LLM with chunked information [89] instead of providing it all at once not only enhances the quality of LLM output but also reduces user effort in post-editing and increases engagement.

Sole reliance on LLM for post-event editing with *LiveSnippets* presented challenges in practical scenarios. They were not only rooted in users' unfamiliarity with framing effective prompts [94], but also attributable to LLM's tendency to generate unsatisfied or hallucinatory content [7] when minimal context is provided but extensive output is expected.

In contrast, *PANDALens*' approach, aligning with the 'Task Salami Technique' [76], breaks content creation into smaller tasks, and each task automatically contributes to the final narratives. Simple yet engaging interactions during the trip, such as collaboratively capturing moments and Q&A dialogues, help *PANDALens* understand both the user's interest and the context accurately and deeply in-situ, cumulatively constructing more precise and comprehensive narratives with significantly reduced (by 70.7%) post-editing time.

Through this approach, *PANDALens* transforms traditional AI-assisted writing tools into a platform where users can seamlessly and effortlessly create higher-quality content during the primary tasks, making in-context content creation less intimidating and more accessible.

7.2.2 Effective Multimodal Contextual Prompts in Mobile Scenarios. The results show that using multimodal contextual information with effective instruction prompts in LLMs tackles two main challenges in mobile scenarios. Firstly, it alleviates the issue of low-quality voice recordings [59]. For example, visual and spatial information can help LLMs overcome errors (e.g., misheard names) in voice transcription and understand the context accurately. Secondly, it offsets reluctance or difficulty in using voice input in public [56]. The use of multimodal data simplifies event descriptions and enables users to focus instead on personal feelings or deep stories, which are vital for creating unique and personalized blog content.

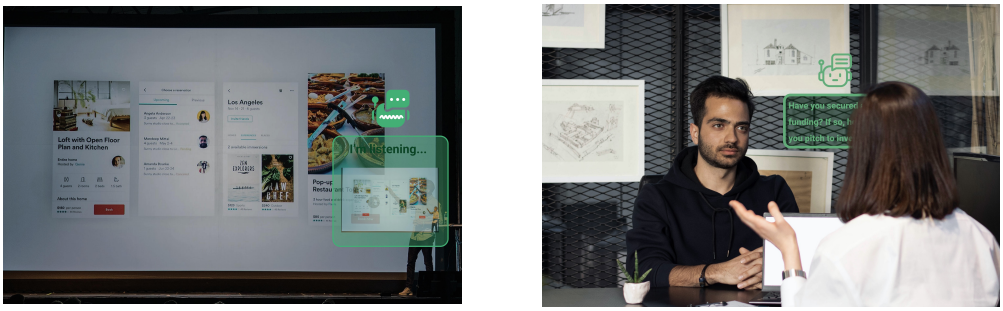
However, compared to stationary settings, mobile scenarios still present unique challenges for LLMs, such as multiple modalities producing errors simultaneously. For instance, a voice transcription mistake of 'crab' for 'grab' with the image model misidentifying crabs as shells. Such significant misalignment obscured the actual context. To address this, *PANDALens* employs Q&A dialogues, using questions to make users aware of and correct potential errors. Future improvements can enhance voice transcription with advanced models [28] and analyze image sequences instead of a single potential blurring image with tools like GPT-4V [93]. Additionally, LLM prompts can dynamically prioritize the most reliable modalities based on the context. For example, it can prioritize image or location data when voice comments are brief and erroneous, while focusing on detailed voice transcriptions when the confidence of image recognition is low.

7.3 Beyond Travel Blog: Applying *PANDALens* to Daily Activities

7.3.1 Towards In-context Information Processing in General. While our focus in this research has primarily been on travel blogs, we envision a general AI assistant that proactively takes multimodal context information and generates multimedia content to satisfy users' context-specific needs. For instance, at an academic symposium, *PANDALens* could transform into a "Scholar's Assistant" (Figure 9a), meticulously capturing keynote speeches, central debates, and invaluable networking moments. It could also assist journalists (Figure 9b) in providing interview question hints, summarizing interview scripts with brief descriptions of the environment, and automatically generating news reports. Participants also expressed diverse expectations for *PANDALens*' output formats, including Vlogs, Comics, and Graphic Narratives based on various needs. Such adaptability equips *PANDALens* to offer content creation across an expansive spectrum of media formats, each tailored to user context and preferences.

7.3.2 Additional Design Considerations. Beyond enhancing the AI's ability to comprehend context and user needs and improving OHMD's portability, participants raised two additional suggestions/concerns.

Dynamic Dialogue with AI. Our research indicates that interactions with *PANDALens* can evoke feelings [38] of companionship in the travel setting. Nevertheless, optimizing Q&A sessions could better serve diverse needs. The system can first modulate dialogue



(a) (b)

Figure 9: Extended Application Scenarios for *PANDALens*. (a) Conference Assistant to record slides and presenter's speech. (b) Journal Assistant to provide interview hints and record the interviewee's answer.

frequency based on the context. For instance, during outdoor activities, participants favor more questions (P10), while during recording lectures, participants prefer either no interruptions or post-class questions (P13). Dialogue with AI also needs to be tailored to individual habits. In travel contexts, immersive explorers prefer posing questions and receiving AI answers (P5, P6). Conversely, sightseers aiming to visit multiple locations quickly might prefer fewer Q&A interactions. Notably, a user's preferences might shift across different scenarios (P15), suggesting tailoring dialogue frequency to users' real-time responses.

Privacy Consideration. Addressing privacy concerns for both users and bystanders is vital [2]. P2 expressed concerns about "voicing inner thoughts in public, especially indoors." To mitigate this, future systems could employ silent speech detection [51]. P11 recommended giving users flexibility in determining the modality data (particularly biometric as gaze) that AI accesses. On bystanders' privacy, most participants (10/16) observed a general acceptance of public photography, such as selfies. However, incorporating a signal light in OHMDs for photo alerts and using local face-blurring before cloud processing can alleviate related concerns.

8 LIMITATIONS AND FUTURE WORK

Evaluation Limitations and Improvements. Our comparative study affirmed the effectiveness of *PANDALens*, but some limitations remain. First, participants were limited to non-frame spectacle wearers for accurate eye tracking. This restriction might bias preferences, especially considering the current OHMD model's weight and semi-transparency, which could potentially diminish the user experience in dimly-lit museum settings. Second, we targeted tech-savvy university-affiliated users, anticipating them as potential OHMD early adopters. However, in-context writing can interest various users, so future research should incorporate diverse demographics for broader applicability. Third, unfamiliarity with the *PANDALens* interface could have affected initial user interactions, notably the ring mouse interactions with OHMD. Moreover, we only compared the efficacy of *PANDALens* with *LiveSnippets* holistically without dissecting the impact of specific system components, such as the in-situ human-AI interaction design or the LLM pipeline, on content generation effectiveness. Future work could conduct ablation studies to assess the contribution of each component separately, such as isolating the impact of in-situ human-AI interaction from the summarizing capabilities in the final content generation stage.

System Limitations and Improvements. Moment capture can be improved in the following three areas. First, in social interactions (e.g., users engaging in conversation while looking at others), integrating GPT-4V can enhance understanding of user documentation intentions through extended video analysis [93], aiding in more accurate interest detection. However, this approach necessitates personalized recording rules tailored to diverse preferences [6]. Secondly, in noisy environments, speaker diarization techniques [62] can help differentiate voices to prevent unintended recordings in current systems. Thirdly, to address issues related to detection latency or user movements, future photo-taking can capture short videos and automatically select the most aesthetically pleasing frame, with an option to change the highlighted frame, similar to Live Photo[8]. Additionally, advanced context information can be provided in future systems to improve content generation quality, e.g., integrating GPT-4V for detailed FPV descriptions and advanced audio models for music recognition. Moreover, expanding language support beyond English is crucial for increasing global accessibility.

9 CONCLUSION

We explored the integration of OHMD interactions with a proactive AI assistant, equipped with a multimodal context analyzer and the LLM pipeline. This facilitates in-context writing during travel, transforming a passive tool into a travel companion. Our comparative study of the proposed *PANDALens* against an in-context writing smartphone application in real-world travel situations confirmed that *PANDALens* effectively captures interesting moments and enriches user expressions. As a result, it helps to create high-quality travel blogs with reduced user effort, enhancing travel enjoyment. We have open-sourced this project at: https://github.com/Synteraction-Lab/PANDALens, and welcome contributions from the community to expand its usage scenarios. Looking forward, future work could focus on developing a general AI assistant capable of processing multimodal contexts and auto-generating documents

[8] https://support.apple.com/en-us/104966

in various application scenarios. These scenarios should incorporate context-specific designs and a wider output modality spectrum to cater to diverse in-context information processing needs.

ACKNOWLEDGMENTS

This research is partially supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). It is also partially supported by the Ministry of Education, Singapore, under its MOE Academic Research Fund Tier 2 programme (MOE-T2EP20221-0010). Additionally, the CityU Start-up Grant also provides partial support. We extend our gratitude to the Lee Kong Chian Natural History Museum for their invaluable assistance with our user studies and to all members of the Synteraction Lab (formerly NUS-HCI Lab) for their help in completing this project. We also thank the reviewers for their valuable feedback.

REFERENCES

[1] Evan Ackerman. 2021. Bosch Gets Smartglasses Right With Tiny Eyeball Lasers. https://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/bosch-ar-smartglasses-tiny-eyeball-lasers Retrieved February 06, 2021.

[2] Fouad Alallah, Ali Neshati, Yumiko Sakamoto, Khalad Hasan, Edward Lank, Andrea Bunt, and Pourang Irani. 2018. Performer vs. observer: whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display?. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3281505.3281541

[3] J.E. Allen, C.I. Guinn, and E. Horvtz. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14, 5 (1999), 14–23. https://doi.org/10.1109/5254.796083

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[5] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–27. https://doi.org/10.1145/3214261

[6] Michiel Bakker, Martin Chadwick, Hannah Sheahan, Michael Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, Amelia Glaese, John Aslanides, Matt Botvinick, et al. 2022. Fine-tuning language models to find agreement among humans with diverse preferences. *Advances in Neural Information Processing Systems* 35, 38176–38189. https://doi.org/10.48550/arXiv.2211.15006

[7] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, et al. 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023* (2023).

[8] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction* 24, 6 (July 2008), 574–594. https://doi.org/10.1080/10447310802205776

[9] Uttaran Bhattacharya, Gang Wu, Stefano Petrangeli, Viswanathan Swaminathan, and Dinesh Manocha. 2021. Highlightme: Detecting highlights from human-centric videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8157–8167.

[10] M. Blum, A. Pentland, and G. Troster. 2006. InSense: Interest-Based Life Logging. *IEEE MultiMedia* 13, 4 (2006), 40–48. https://doi.org/10.1109/MMUL.2006.87

[11] Marc Bolaños, Mariella Dimiccoli, and Petia Radeva. 2017. Toward Storytelling From Visual Lifelogging: An Overview. *IEEE Transactions on Human-Machine Systems* 47, 1 (2017), 77–90. https://doi.org/10.1109/THMS.2016.2616296

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

[13] John Brooke. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 7.

[14] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared 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 M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher 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. https://doi.org/10.48550/arXiv.2005.14165

[15] Vannevar Bush et al. 1945. As we may think. *The atlantic monthly* 176, 1 (1945), 101–108.

[16] Daniela Buzova, Amparo Cervera-Taulet, and Silvia Sanz-Blas. 2020. Exploring multisensory place experiences through cruise blog analysis. *Psychology & Marketing* 37, 1 (2020), 131–140. https://doi.org/10.1002/mar.21286 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.21286

[17] Runze Cai, Nuwan Nanayakkarawasam Peru Kandage Janaka, Shengdong Zhao, and Minghui Sun. 2023. ParaGlassMenu: Towards Social-Friendly Subtle Interactions in Conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 721, 21 pages. https://doi.org/10.1145/3544548.3581065

[18] Lalith Chandralal, Jennifer Rindfleish, and Fredy Valenzuela. 2015. An Application of Travel Blog Narratives to Explore Memorable Tourism Experiences. *Asia Pacific Journal of Tourism Research* 20, 6 (2015), 680–693. https://doi.org/10.1080/10941665.2014.925944

[19] Yuhu Chang, Yingying Zhao, Mingzhi Dong, Yujiang Wang, Yutian Lu, Qin Lv, Robert P Dick, Tun Lu, Ning Gu, and Li Shang. 2021. MemX: An attention-aware smart eyewear system for personalized moment auto-capture. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–23.

[20] Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel Tetreault, and Alejandro Jaimes-Larrarte. 2022. Mapping the Design Space of Human-AI Interaction in Text Summarization. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 431–455. https://doi.org/10.18653/v1/2022.naacl-main.33

[21] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 329–340. https://doi.org/10.1145/3172944.3172983

[22] Didier Coste and John Pier. 2009. *Narrative levels*.

[23] Ana Garcia Del Molino, Cheston Tan, Joo-Hwee Lim, and Ah-Hwee Tan. 2016. Summarization of egocentric videos: A comprehensive survey. *IEEE Transactions on Human-Machine Systems* 47, 1 (2016), 65–76.

[24] Zijian Ding and Joel Chan. 2023. Mapping the Design Space of Interactions in Human-AI Text Co-creation Tasks. https://doi.org/10.48550/arXiv.2303.06430

[25] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. 2023. PaLM-E: An Embodied Multimodal Language Model. arXiv:2303.03378 [cs.LG]

[26] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, Cheng Peng, Chris Sweeney, Cole Wilson, Dan Barnes, Daniel DeTone, David Caruso, Derek Valleroy, Dinesh Ginjupalli, Duncan Frost, Edward Miller, Elias Mueggler, Evgeniy Oleinik, Fan Zhang, Guruprasad Somasundaram, Gustavo Solaira, Harry Lanaras, Henry Howard-Jenkins, Huixuan Tang, Hyo Jin Kim, Jaime Rivera, Ji Luo, Jing Dong, Julian Straub, Kevin Bailey, Kevin Eckenhoff, Lingni Ma, Luis Pesqueira, Mark Schwesinger, Maurizio Monge, Nan Yang, Nick Charron, Nikhil Raina, Omkar Parkhi, Peter Borschowa, Pierre Moulon, Prince Gupta, Raul Mur-Artal, Robbie Pennington, Sachin Kulkarni, Sagar Miglani, Santosh Gondi, Saransh Solanki, Sean Diener, Shangyi Cheng, Simon Green, Steve Saarinen, Suvam Patra, Tassos Mourikis, Thomas Whelan, Tripti Singh, Vasileios Balntas, Vijay Baiyya, Wilson Dreewes, Xiaqing Pan, Yang Lou, Yipu Zhao, Yusuf Mansour, Yuyang Zou, Zhaoyang Lv, Zijian Wang, Mingfei Yan, Carl Ren, Renzo De Nardi, and Richard Newcombe. 2023. Project Aria: A New Tool for Egocentric Multi-Modal AI Research. arXiv:2308.13561 [cs.HC]

[27] Niaja Farve, Tal Achituv, and Pattie Maes. 2016. User Attention with Head-Worn Displays. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, Santa Clara, California, USA, 2467–2473. https://doi.org/10.1145/2851581.2892530

[28] Sanchit Gandhi, Patrick von Platen, and Alexander M. Rush. 2023. Distil-Whisper: Robust Knowledge Distillation via Large-Scale Pseudo Labelling. arXiv:2311.00430 [cs.CL]

[29] Jennifer Gluck, Andrea Bunt, and Joanna McGrenere. 2007. Matching Attentional Draw with Utility in Interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/1240624.1240631

[30] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006),

904–908. https://doi.org/10.1177/154193120605000909
[31] Jochen Hartmann. 2022. Emotion English DistilRoBERTa-base. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/.
[32] Joyce Ho and Stephen S. Intille. 2005. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. ACM Press, Portland, Oregon, USA. https://doi.org/10.1145/1054972.1055100
[33] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. https://doi.org/10.1145/302979.303030
[34] Kuo-Lun Hsiao, Hsi-Peng Lu, and Wan-Chin Lan. 2013. The influence of the components of storytelling blogs on readers' travel intentions. *Internet Research* 23, 2 (2013), 160–182.
[35] Shamsi T. Iqbal, Jaime Teevan, Dan Liebling, and Anne Loomis Thompson. 2018. Multitasking with Play Write, a Mobile Microproductivity Writing Tool. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin Germany, 411–422. https://doi.org/10.1145/3242587.3242611
[36] Seppo E Iso-Ahola. 1983. Towards a social psychology of recreational travel. *Leisure studies* 2, 1 (1983), 45–56.
[37] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2021), 120:1–120:36. https://doi.org/10.1145/3453157
[38] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3544548.3581196
[39] Nuwan Janaka, Jie Gao, Lin Zhu, Shengdong Zhao, Lan Lyu, Peisen Xu, Maximilian Nabokow, Silang Wang, and Yanch Ong. 2023. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023). https://doi.org/10.1145/3610931
[40] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and near-peripheral visualizations: Towards attention-maintaining secondary information presentation on OHMDs during in-person social interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3502127
[41] Nuwan Nanayakkarawasam Peru Kandage Janaka, Shengdong Zhao, and Shardul Sapkota. 2023. Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 575, 23 pages. https://doi.org/10.1145/3544548.3580891
[42] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12, Article 248 (mar 2023), 38 pages. https://doi.org/10.1145/3571730
[43] Bumsoo Kang, Chulhong Min, Wonjung Kim, Inseok Hwang, Chunjong Park, Seungchul Lee, Sung-Ju Lee, and Junehwa Song. 2017. Zaturi: We Put Together the 25th Hour for You. Create a Book for Your Baby. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 1850–1863. https://doi.org/10.1145/2998181.2998186
[44] Hyeongcheol Kim, Shengdong Zhao, Can Liu, and Kotaro Hara. 2020. LiveSnippets: Voice-Based Live Authoring of Multimedia Articles about Experiences. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 31, 11 pages. https://doi.org/10.1145/3379503.3403556
[45] Rafal Kocielnik, Fabrizio Maria Maggi, and Natalia Sidorova. 2013. Enabling self-reflection with LifelogExplorer: Generating simple views from complex data. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. 184–191. https://doi.org/10.4108/icst.pervasivehealth.2013.251934
[46] Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. 2020. Effects of Proactive Dialogue Strategies on Human-Computer Trust. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 107–116. https://doi.org/10.1145/3340631.3394840
[47] Rebecca Krosnick, Fraser Anderson, Justin Matejka, Steve Oney, Walter S. Lasecki, Tovi Grossman, and George Fitzmaurice. 2021. Think-Aloud Computing: Supporting Rich and Low-Effort Knowledge Capture. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445066
[48] Amel Ksibi, Ala Saleh D. Alluhaidan, Amina Salhi, and Sahar A. El-Rahman. 2021. Overview of Lifelogging: Current Challenges and Advances. *IEEE Access* 9 (2021), 62630–62641. https://doi.org/10.1109/ACCESS.2021.3073469
[49] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
[50] Ping Li and Huimin Liang. 2020. Factors influencing learning effectiveness of educational travel: A case study in China. *Journal of Hospitality and Tourism Management* 42 (2020), 141–152.
[51] Richard Li, Jason Wu, and Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3311823.3311831
[52] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents in the Post-ChatGPT World. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3452–3455. https://doi.org/10.1145/3539618.3594250
[53] Behrooz Mahasseni, Michael Lam, and Sinisa Todorovic. 2017. Unsupervised video summarization with adversarial lstm networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 202–211.
[54] Elke Mark and Lindsey French. 2021. In Formation: Micro-Phenomenology as a Technology of Memory. *Algorithmic and Aesthetic Literacy: Emerging Transdisciplinary Explorations for the Digital Age* (2021), 135.
[55] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.
[56] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human–Computer Interaction* 31, 4 (2015), 307–335. https://doi.org/10.1080/10447318.2014.986642
[57] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of GPT-4 on Medical Challenge Problems. arXiv:2303.13375 [cs.CL]
[58] Donald A. Norman. 2013. *The design of everyday things* (revised and expanded edition ed.). Basic Books.
[59] Caroline Nowacki, Anna Gordeeva, and Anne-Hélène Lizé. 2020. Improving the usability of voice user interfaces: a new set of ergonomic criteria. In *Design, User Experience, and Usability. Design for Contemporary Interactive Environments: 9th International Conference, DUXU 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*. Springer, 117–133.
[60] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review* 18, 2 (June 2014), 20–31. https://doi.org/10.1145/2636242.2636246
[61] Sanghun Park and Carla Almeida Santos. 2017. Exploring the Tourist Experience: A Sequential Approach. *Journal of Travel Research* 56, 1 (2017), 16–27. https://doi.org/10.1177/0047287515624017 arXiv:https://doi.org/10.1177/0047287515624017
[62] Tae Jin Park, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J Han, Shinji Watanabe, and Shrikanth Narayanan. 2022. A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language* 72 (2022), 101317.
[63] Claire Petitmengin. 2017. Uncovering the dynamics of lived experience through micro-phenomenology. (2017).
[64] Claire Petitmengin, Martijn Van Beek, Michel Bitbol, Jean-Michel Nissou, and Andreas Roepstorff. 2019. Studying the experience of meditation through micro-phenomenology. *Current opinion in psychology* 28 (2019), 54–59.
[65] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3580907
[66] Blaine A. Price, Avelie Stuart, Gul Calikli, Ciaran Mccormick, Vikram Mehta, Luke Hutton, Arosha K. Bandara, Mark Levine, and Bashar Nuseibeh. 2017. Logging You, Logging Me: A Replicable Study of Privacy and Sharing Behaviour in Groups of Visual Lifeloggers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 22 (jun 2017), 18 pages. https://doi.org/10.1145/3090087
[67] Alec Radford, Jeff Wu, Rewon Child, D. Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners.
[68] Valentin Radu, Catherine Tong, Sourav Bhattacharya, Nicholas D. Lane, Cecilia Mascolo, Mahesh K. Marina, and Fahim Kawsar. 2018. Multimodal Deep Learning for Activity and Context Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 157 (jan 2018), 27 pages. https://doi.org/10.1145/3161174
[69] K Ann Renninger and Suzanne Hidi. 2015. *The power of interest for motivation and engagement*. Routledge.
[70] Laria Reynolds and Kyle McDonell. 2021. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 314,

7 pages. https://doi.org/10.1145/3411763.3451760
[71] Jeba Rezwana and Mary Lou Maher. 2022. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interaction* (Feb. 2022). https://doi.org/10.1145/3519026
[72] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. 2021. Ubiquitous Interactions for Heads-Up Computing: Understanding Users' Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 36, 15 pages. https://doi.org/10.1145/3447526.3472035
[73] Abigail J. Sellen and Steve Whittaker. 2010. Beyond total capture: a constructive critique of lifelogging. *Commun. ACM* 53, 5 (May 2010), 70–77. https://doi.org/10.1145/1735223.1735243
[74] Paul J Silvia. 2006. *Exploring the psychology of interest*. Psychology of Human Motivation.
[75] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* (feb 2022). https://doi.org/10.1145/3511599
[76] Branislav L Slantchev. 2005. Introduction to International Relations Lecture 8: Deterrence and Compellence. *Lecture, Department of Political Science, University of California–San Diego* (2005).
[77] Mohammad Soleymani and Marcello Mortillaro. 2018. Behavioral and physiological responses to visual interest and appraisals: Multimodal analysis and automatic recognition. *Frontiers in ICT* 5 (2018), 17.
[78] Lujun Su, Jin Cheng, and Scott R. Swanson. 2020. The impact of tourism activity type on emotion and storytelling: The moderating roles of travel companion presence and relative ability. *Tourism Management* 81 (2020), 104138. https://doi.org/10.1016/j.tourman.2020.104138
[79] Min Sun, Ali Farhadi, and Steve Seitz. 2014. Ranking domain-specific highlights by analyzing edited videos. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer, 787–802.
[80] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.
[81] Aik Lim Tan, Robyn Gillies, and Azilawati Jamaludin. 2021. A Case Study: Using a Neuro-Physiological Measure to Monitor Students' Interest and Learning during a Micro:Bit Activity. *Education Sciences* 11, 8 (2021). https://doi.org/10.3390/educsci11080379
[82] Tony SM Tse and Elaine Yulan Zhang. 2013. Analysis of blogs and microblogs: A case study of Chinese bloggers sharing their Hong Kong travel experiences. *Asia Pacific Journal of Tourism Research* 18, 4 (2013), 314–329.
[83] Vincent Wing Sun Tung, Pearl Lin, Hanqin Qiu Zhang, and Aimin Zhao. 2017. A framework of memory management and tourism experiences. *Journal of Travel & Tourism Marketing* 34, 7 (2017), 853–866.
[84] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K. Kane. 2023. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3581560
[85] Niels Van Berkel, Jeremy Opie, Omer F. Ahmad, Laurence Lovat, Danail Stoyanov, and Ann Blandford. 2022. Initial Responses to False Positives in AI-Supported Continuous Interactions: A Colonoscopy Case Study. *ACM Trans. Interact. Intell. Syst.* 12, 1, Article 2 (mar 2022), 18 pages. https://doi.org/10.1145/3480247
[86] Tina Caroline Walber, Ansgar Scherp, and Steffen Staab. 2014. Smart Photo Selection: Interpret Gaze as Personal Interest. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2065–2074. https://doi.org/10.1145/2556288.2557025
[87] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.
[88] IpKin Anthony Wong, Qi Lilith Lian, and Danni Sun. 2023. Autonomous travel decision-making: An early glimpse into ChatGPT and generative AI. *Journal of Hospitality and Tourism Management* 56 (2023), 253–263.
[89] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582
[90] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–30. https://doi.org/10.1145/3544548.3581500
[91] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301
[92] Rongcan Yang and Vincent Wing Sun Tung. 2018. How does family influence the travel constraints of solo travelers? Construct specification and scale development. *Journal of Travel & Tourism Marketing* 35, 4 (2018), 507–516. https://doi.org/10.1080/10548408.2017.1363685
[93] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). arXiv:2309.17421 [cs.CV]
[94] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388
[95] Shengdong Zhao, Felicia Tan, and Katherine Fennedy. 2023. Heads-Up Computing Moving Beyond the Device-Centered Paradigm. *Commun. ACM* 66, 9 (aug 2023), 56–63. https://doi.org/10.1145/3571722
[96] Chen Zhou, Katherine Fennedy, Felicia Fang-Yi Tan, Shengdong Zhao, and Yurui Shao. 2023. Not All Spacings Are Created Equal: The Effect of Text Spacings in On-the-Go Reading Using Optical See-Through Head-Mounted Displays. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 720, 19 pages. https://doi.org/10.1145/3544548.3581430

A FORMATIVE STUDY: DATA ANALYSIS

We employed a thematic analysis approach outlined by Braun and Clarke [12] on 12 sets of transcribed interview notes with observational notes detailing user behaviors and environments. Initially, two co-authors reviewed four interview scripts from a random selection of four participants, two from indoor and two from outdoor sessions. They independently formulated initial codes and clustered them around shared themes rooted in the original research questions. After achieving an initial agreement of 85%, subsequent discussions addressed, interpreted and resolved any differences encountered during this process. One of the co-authors then continued to code the remaining data, refining the themes as necessary until data saturation was reached. Finally, both co-authors examined the textual data and video footage to identify specific quotes relevant to each identified theme.

B PROMPTS FOR LARGE LANGUAGE MODEL

B.1 Iterative Prompt Development

We followed the guidelines provided by OpenAI and DeepLearning.AI [9] to iteratively develop our prompts in addition to the specific technique (e.g., chain-of-thoughts) mentioned in the paper. We summarize the high-level steps as follows.

(1) **Scoping:** Define the system scope and role
e.g., AI chat assistant for writing travel blogs with multimodal inputs
(2) **Personalization:** Configure the system to personalize the writing style
- (user given) provide some examples to follow
- (user given) explicitly mention what style is preferred
- (user selection) request system to suggest some styles to choose

[9] https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson/2/guidelines

(3) **Task detailing:** Instruct the system about expected inputs, outputs, and interactions in detail.
- e.g., input types and formats
- e.g., output types and formats
- e.g., authoring modes (data feeding, final outputs)
- e.g., user role (providing content for authoring)

(4) **Iterative Refinement:**
(a) **Task refinement:** Refine the system task based on output e.g., add capabilities to the system
(b) **Output refinement:** e.g., improve the output (quality, creativity, interest, accuracy, style, etc.)
(c) **Input refinement:** e.g., improve the input format and make sure all inputs are taken when writing; if not, ask to prioritize
(d) **Interaction refinement:** e.g., check users to follow their role (e.g., providing clear content for authoring), else enable to proactively engage users to do so (ask questions)

B.2 Prompt for *PANDALens* System's Moment Capture and Content Generation

Figure 10 shows the main prompt of *PANDALens* system for moment analysis and content generation.

B.3 Prompt for Customizing Context-Related Question Style and Final Writing Style

Figure 11 shows the prompts used for customizing Context-Related Question and Final Writing Style in the *PANDALens* system to satisfy different users' preferences.

B.4 Prompt for Concise Chat History

Figure 12 shows the prompt for compressing chat history with the GPT model.

C DETAILS OF SYSTEM IMPLEMENTATION

The *PANDALens* is developed using Python 3.9. The OHMD (XREAL Air) with a laptop serves as the near-eye display, while the Pupil Core add-on, also attached to the OHMD, facilitates gaze detection and offers FPV video streaming. Tkinter is utilized for the OHMD's graphical user interface.

On the software back end, several processes concurrently capture context data and handle user commands. Communication with the Pupil Core for acquiring gaze positions, fixation detection, and FPV frames is realized through a socket connection with the Pupil Capture App. The YOLO v8 model[10] operates in real-time to analyze FPV data, detecting potential objects of interest and tracking user gaze interactions with these objects. The OpenCV's 'compareHist'[11] function gauges the similarity between sequential frames, identifying transitions during trips. Google Cloud Vision API[12] analyses user-captured photos, recognizes image labels, and performs OCR operations. The BLIP large model[13], hosted by Hugging Face, delivers image descriptions. Audio categorization is handled by MediaPipe, voice transcription by Whisper, and the Emotion English DistilRoBERTa-base model [31] undertakes sentiment analysis of users' speech. Geolocation data is sourced from Geopy[14] 2.3.0 and Geocoder[15] 1.38.1, while timestamps are derived from Python's Datetime package[16].

For contextual data processing and writing assistance, we employed the GPT3.5-Turbo-16K model[17] as our LLM, chosen for its faster responses and larger token capacity, aiding in retaining chat history. We applied various prompt engineering techniques to ensure the LLM could perform tasks efficiently and integrate seamlessly into *PANDALens*. Techniques encompassed crafting clear and specific LLM instructions, adopting few-shot prompts with demonstrative examples, formatting LLM responses as JSON, and employing methodologies like the Chain-of-Thought approach [70, 87]. Detailed prompt information can be referenced in Appendix B.2. When approaching the set token length threshold (10k tokens for the current configuration), the system instructs the LLM to condense the preceding chat into a succinct yet inclusive summary (detailed prompts in Appendix B.4). This ensures users can document many moments (tested for more than 4 hours of continuous travel) without worrying about the LLM's token limitations (16k tokens).

D STATISTICS FOR COMPARATIVE STUDY

Table 3 and Table 4 present the statistics of quality of content generation and quality of travel and moment capture, respectively.

[10] https://ultralytics.com/yolov8
[11] https://docs.opencv.org/3.4/d8/dc8/tutorial_histogram_comparison.html
[12] https://cloud.google.com/vision
[13] https://huggingface.co/Salesforce/blip-image-captioning-large
[14] https://pypi.org/project/geopy/
[15] https://geocoder.readthedocs.io/
[16] https://docs.python.org/3/library/datetime.html
[17] https://platform.openai.com/docs/models/gpt-3-5

```
Help me create a high-quality travel blog for the user. You will be provided delimited JSON quotes, including the number of images of interesting moments, image descriptions/labels, and OCR. I may also send the user's thoughts or other comments on their experiences. Background context, e.g., user behaviors and background audio, may also be sent to you. You can help me create an appealing travel blog that reflects the user's writing style and preferences.
The users' preferred' writing style/example is:\n
{Use a "Concise and Informative" style for the travel blog entry, focusing on straightforward descriptions and relevant details without excessive embellishments.}
We have three modes, 1) authoring mode, 2) selecting mode, and 3) writing & revise full-text mode.

The authoring mode focuses on each moment of user travel. To achieve authoring mode, perform the following actions:
1) Summarize each picture and recognize the scene based on the available information.
2) Describe the environment/atmosphere when I send the background context of the user and adopt user comments into the writing if any. Btw, automatically fix users comments (i.e., voice transcription) if you detect any errors based on the context.
3) If users can't provide precise information to certain stuff, you can use the context and your knowledge to guess the missing content and add it to the full blog.
4) Think insightful questions to gain insight into the user's interests, purposes, and emotional expression, or you find something unclear for you to generate high-quality writing. These questions should be contextually relevant and engaging questions that encourage users to reflect on their in-situ travel experiences. Focus on capturing unique moments, interactions, and sensory details that make the travel experience memorable and personal.
The users preferred question style is: {Ask the user specific detail-oriented questions to gather unique features or distinct characteristics about their experiences.}
5) Ask interesting questions. And don't ask more than TWO questions for ONE topic/moment recording.
6) Return the response **ONLY** in JSON format, with the following structure: ```json{"mode": "authoring", "response":{"summary of new content": "[snippet of the travel blog content preview in first person narration]", "question to users": "[Question to help them provide deeper and more interesting content *if necessary*, return 'None' when no question you want to ask. (Put all questions here.)]" } }```

The selecting mode helps users to select the moments they want to include in the final travel blog.
To achieve writing & revise full blog mode, perform the following actions:
1) Consider previous interesting moments in authoring mode and summarize them.
2) Send the summary of each moment in the list and ask users to select their favorite moments that they want to include in the final writing.
3) Return the response **ONLY** in JSON format, with the following structure: ```json{"mode": "selecting," "response": "List:\n no.1, [One sentence summary for moment1] \n" }```


The writing & revise full blog mode focuses on writing and revising the final full travel blog when I give you instructions for writing a full blog. To achieve writing & revise full blog mode, perform the following actions:
1) Consider user's selected moments (i.e., ONLY use the moment(s) the user mentioned in "user_voice_transcription") in previous selecting mode.
2) Adopt the user's preferred writing style.
3) Revise the content & structure when users ask you to do so.
4) Return the response **ONLY** in JSON format, with the following structure: ```json{"mode": "full", "response": {"full writing": "[full travel blog content in first person narration]", "revised parts": "[the newly added or revised content, return 'None' when no revision.]"}}```

Every time you receive input, you need to decide whether they belong to an authoring mode, selecting mode, or writing & revise full-text mode, then generate the corresponding response. For example, inputs like {"no": 1, "photo_label": "Food: 97.82%, Tableware: 96.79%, Pizza: 95.05%", "photo_caption": "a large pizza sitting on top of a table", "audio": "Crowded people", "user_behavior": None, "user_voice_transcription": "This is our lunch after my first CHI presentation. We went to a very good restaurant."} or {"user_voice_transcription": "[Users' answer to the question you asked.]"} belong to an authoring mode. Selecting mode's input would be something like {"User Command": "List all the moments' summary."}, while writing & revise full blog mode's input would be something like {"User Command": "Write a full blog based on the previous chat history."}, or {"user_voice_transcription": "Just help me to shorter the writing. I want to make it more like a Twitter style and add emojis to it"}.

Note: **Only return the necessary response in JSON format** to save tokens. No other conversation content is needed. Let's start with authoring mode.
```

Figure 10: Main Prompt of *PANDALens* system for moment analysis and content generation. Note: The highlighted parts are customized prompts for different users' preferences.

```
You are a travel blog helper. Based on my input, you need to help me write an interesting and personalized travel blog.
An example input is "{
 "no": 1,
            "label": "Coffee 94% Cafe 92% Beverage 90% Table 88% Interior 86% Latte 84% Morning 82% Breakfast 80% Food 78% Relaxation 76%",
            "Caption": "A cup of coffee on a wooden table in a cozy cafe",
            "Audio": "soft music, low chatter, espresso machine noises",
            "Comment": "Starting my day with a delicious cup of coffee at this lovely cafe!",
            "User Behaviors": "The user is holding a coffee cup"
        }".

To better understand users preferred writing styles, you need to try your best to figure it out. You can achieve this task by performing the following steps. First, ask the users if they have any preferred writing style or tone that they would like you to incorporate into their travel blog. This will help you create a writing style that aligns with their expectations.

If users have no specific preference, you will provide the user with several writing samples that you think are good and in different styles and ask them to choose a preferred one. The  format is <[Style Type]: [Example]>. For example, "Descriptive and Imaginative: This style focuses on creating vivid and immersive descriptions to help the reader imagine themselves in the scene. For example, "The aroma of freshly roasted coffee beans filled the air, blending with the soft background music and the gentle chatter of patrons. The rustic wooden table exuded warmth, inviting me to sit down and savor the rich flavor of the latte."  This will help you to understand the user's taste and writing preferences.

Once you have identified the user's preferred writing style, you will summarize the writing style in a prompt so that other GPTs can understand how to help the users better to write in the future.

After summarizing the writing style, you also need to think about a good question style that you can ask users after they give you their life moments to dig into users' comments to help you better write the blog. Follow the same approach as the writing style, such as providing some examples in different styles, i.e.,  <[Style Type]: [Example]>, and then summarizing the question style in a prompt.
```

Figure 11: Prompt for customizing Context-Related Question Style and Final Writing Style in *PANDALens* system for different users' preferences.

```
Summarize the chat history, ensuring that key details, user preferences, and interesting moments are preserved, while keeping as many details as possible to provide sufficient material for writing.
Focus on maintaining the essence of the conversation without sacrificing important information.
Create a concise version of the chat history by prioritizing the following aspects: main activities, location highlights, unique experiences, user emotions, and personal reflections.
This summary should provide a comprehensive and engaging overview of the user's travel experience, serving as a rich foundation for creating a captivating travel blog.
```

Figure 12: Prompt for compressing chat history with GPT model.

Table 3: Objective and Subjective measures for the quality of content generation.

	Post-Editing Time (min)		**Word Count**		***Post-Editing Effort***		***Language***		***Creativity & Appeal***		***Writing Style***		***Self-Rated Writing Score***		***Control over Content***		***Trust in Content***	
	PANDALens	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***
Mean	5.26	17.93	329.13	317.75	2.50	5.88	6.19	3.63	5.75	3.31	5.63	3.75	82.19	60.88	5.25	4.69	5.63	3.25
SD	5.15	9.72	176.83	272.49	1.46	1.45	0.75	1.93	1.13	1.66	1.50	2.08	8.35	21.83	1.48	2.41	1.41	1.73
Median	4.75	15.40	246.50	245.00	2.00	6.00	6.00	3.50	6.00	3.50	6.00	4.00	81.00	70.00	5.50	5.50	6.00	3.00
25th percentile	2.38	10.28	204.00	160.75	1.75	6.00	6.00	2.00	5.00	2.00	4.75	2.00	78.75	52.50	4.00	2.00	5.00	2.00
75th percentile	6.18	20.50	474.50	423.50	3.00	7.00	7.00	5.25	6.25	5.00	7.00	6.00	90.00	77.00	6.25	7.00	6.25	4.25
Statistics	$t(15) = -4.87,\ p = 0.0002$		$t(15) = 0.18,\ p = 0.8567$		$Z = 6.0,\ p = 0.0004$		$Z = 0.0,\ p = 0.0021$		$Z = 0.0,\ p = 0.0021$		$Z = 11.0,\ p = 0.0088$		$t(15) = 4.21,\ p = 0.0008$		$Z = 26.5,\ p = 0.3225$		$Z = 7.0,\ p = 0.0040$	

Table 4: Subjective measures for the quality of travel and moment capture.

	Device Familiarity		***Travel Enjoyment* (current)**		***Travel Enjoyment* (future)**		***Writing Productivity***		***Ideation Support***		***Distraction***		***Naturalness***		***RTLX***		***SUS***	
	PANDALens	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***	***PANDALens***	***LiveSnippets***
Mean	3.13	6.50	5.38	4.75	6.50	5.19	5.88	2.81	5.56	2.38	3.31	2.88	5.31	5.19	33.49	35.47	77.03	67.19
SD	2.03	0.73	1.09	1.18	0.89	1.22	1.31	1.52	1.46	1.54	1.62	1.59	1.01	1.64	15.56	19.27	14.41	20.65
Median	3.00	7.00	6.00	5.00	7.00	5.00	6.00	3.00	6.00	2.00	3.00	3.00	5.50	6.00	33.33	38.33	78.75	70.00
25th percentile	1.00	6.00	4.00	4.00	6.00	4.75	5.00	1.75	5.00	1.00	2.00	1.75	5.00	4.75	23.12	19.17	68.13	50.00
75th percentile	5.00	7.00	6.00	6.00	7.00	6.00	7.00	3.00	6.25	3.25	5.00	4.00	6.00	6.00	41.25	45.83	90.63	77.50
Statistics	$Z = 0.0,\ p = 0.0009$		$Z = 22.5,\ p = 0.1875$		$Z = 0.0,\ p = 0.0019$		$Z = 1.5,\ p = 0.0008$		$Z = 2.0,\ p = 0.0001$		$Z = 26.0,\ p = 0.5312$		$Z = 29.0,\ p = 0.7139$		$t(15) = -0.4173,\ p = 0.6824$		$t(15) = 1.98,\ p = 0.0668$	