Pixel Memories

Do Lifelog Summaries Fail to Enhance Memory but Offer Privacy-Aware Memory Assessments?

Elagroudy, Passant; Rzayev, Rufat; Machulla, Tonja Katrin; Le, Huy Viet; Dingler, Tilman; Lischke, Lars; Clinch, Sarah; Ward, Geoffrey; Schmidt, Albrecht

DOI
10.1145/3706598.3714145

Publication date
2025

Document Version
Final published version

Published in
CHI 2025 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems

Pixel Memories: Do Lifelog Summaries Fail to Enhance Memory but Offer Privacy-Aware Memory Assessments?

Passant Elagroudy
German Research Centre for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
RPTU Kaiserslautern
Kaiserslautern, Germany
passant.elagroudy@gmail.com

Rufat Rzayev
Interactive Media Lab Dresden
TUD Dresden University of Technology
Dresden, Germany
rufat.rzayev@tu-dresden.de

Tonja-Katrin Machulla
Institute for Media Research
TU Chemnitz
Chemnitz, Germany
tonja.machulla@phil.tu-chemnitz.de

Huy Viet Le
University of Stuttgart
Stuttgart, Germany
mail@huyle.de

Tilman Dingler
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

Lars Lischke
User Centric Data Science
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
lars.lischke@gmail.com

Sarah Clinch
Department of Computer Science
The University of Manchester
Manchester, United Kingdom
sarah.clinch@manchester.ac.uk

Geoffrey Ward
Department of Psychology
University of Essex
Essex, United Kingdom
gdward@essex.ac.uk

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

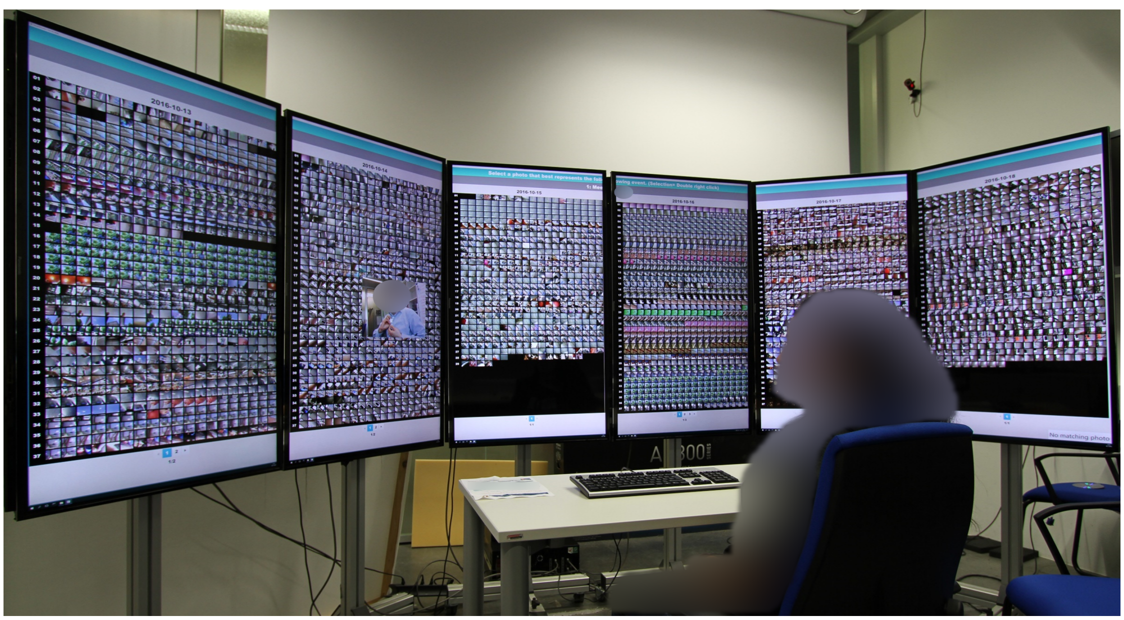

Figure 1: "Pixel Memories": A lifelog visualization prototype consisting of six 50-inch screens, each presenting a day in the life of a participant. Each entry in the grid is a photo captured every 30 seconds automatically by a clip-on camera. The screens show a maximum of 4218 photos at a glance (703 photos per screen).

Abstract

We explore the metaphorical "daily memory pill" concept – a brief pictorial lifelog recap aimed at reviving and preserving memories. Leveraging psychological strategies, we explore the potential of such summaries to boost autobiographical memory. We developed an automated lifelogging memory prosthesis and a research protocol (Automated Memory Validation "AMV") for conducting privacy-aware, in-situ evaluations. We conducted a real-world lifelogging experiment for a month (n=11). We also designed a browser "Pixel Memories" for browsing one-week worth of lifelogs. The results suggest that daily timelapse summaries, while not yielding significant memory augmentation effects, also do not lead to memory degradation. Participants' confidence in recalled content remains unaltered, but the study highlights the challenge of users' overestimation of memory accuracy. Our core contributions, the AMV protocol and "Pixel Memories" browser, advance our understanding of memory augmentations and offer a privacy-preserving method for evaluating future ubicomp systems.

CCS Concepts

• Human-centered computing → HCI theory, concepts and models; **• Applied computing → Psychology**.

Keywords

lifelogging, recall, memory research, privacy, case study

ACM Reference Format:

Passant Elagroudy, Rufat Rzayev, Tonja-Katrin Machulla, Huy Viet Le, Tilman Dingler, Lars Lischke, Sarah Clinch, Geoffrey Ward, and Albrecht Schmidt. 2025. Pixel Memories: Do Lifelog Summaries Fail to Enhance Memory but Offer Privacy-Aware Memory Assessments?. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3714145

1 Introduction

Memories shape our understanding of experiences and reactions in life. Thus, there is a rise in the research aimed at creating ubiquitous memory-altering prostheses that intrinsically change how we remember things or externalize our memories to be available on demand. Examples of intrinsic prostheses are micro-learning applications [15] and reminder features of social media applications of events that happened on the same day (e.g. "on this day" by Facebook). Examples of externalized memory prostheses are interventions for dementia patients [44, 47] and cloud services and reminders, such as calendars. On the other hand, recent research has also shown the potential of accidental memory alterations [1, 18]) or memory degradations (e.g. [14, 54, 56]) resulting from the usage of technological interventions. For example, Adams et al. [1] showed that eyewitnesses changed their true testimonies when they did not find supporting photos in their lifelogs from their chest-mounted cameras. Thus, there is a need in the UbiComp and HCI communities for efficient methods to gauge the impact of technological interventions targeting memories on the lives of people.

We explore in this work the notion of a metaphorical "daily memory pill", where a person looks at their day in a quick summary lasting less than a minute to revive and better retain their memories on the longer run. To build this pill, we rely on well-known psychological strategies such as repetitive review of content adopted by ubiquitous computing systems (UCSs) to reduce memory decay (e.g. [21, 28, 52]). We specifically chose to base our summaries on pictorial lifelogs because of their comprehensive nature of documenting one's life, their classical potential in literature to support memory tasks especially in health-related domains (e.g. [8, 29, 58]), their potential to support participants' reflection, emotional growth, and reasoning about past experiences (e.g. [27, 28, 53]), and their ambient nature in collecting the information without needing the users to explicitly collect the data. Wide adoption, however, has so far held off due to a combination of technological constraints, a lack of efficient reviewing techniques to summarize the sheer amounts of lifelogs (ca. 1200 photos per day) and privacy concerns. Our *goal* here is two-fold: 1) create those automated summaries and look at their impact on the formation and preservation of autobiographical memories, and 2) develop a research methodology that enables us to investigate those memory changes from technological interventions in a privacy-preserving manner. To this end, we specifically have *two* contributions in this paper:

(1) We built an in-situ lifelogging memory prosthesis. It is an automated time-lapse generator summarizing the daily events of the participant using lifelogs. We deployed the system and evaluated it for a month (n=11). We use it as a test bed to understand the impact of the "daily memory pill" on recall quality, quantity, and confidence (see hypotheses in Section 4.1).
(2) A novel research protocol, named *Automated Memory Validation (AMV)*, to use lifelogging for conducting privacy-aware in-situ evaluations of memory prostheses (see summary in Table 1). It addresses the key challenges: privacy and automated validation. Using this protocol, we support the ubiquitous collection and validation of human memories without exposing their content to the researchers. The protocol has four stages, namely: collecting participant's ground truth memories, collecting sensor ground truth data, subjective relevance of technological intervention on memories, and validating the participants' memories through the sensor data. We also built a novel system to browse large datasets of lifelogs (ca. 4000 photos at a glance) (see Figure 1 for the "Pixel Memories" prototype). The system showed promising results to support reflection as a standalone tool.

Our results show that consuming daily timelapse summaries of the day's activities (shorter than a minute) does not yield to the hoped for memory augmentation effects, neither in terms of remembering more or remembering more accurately. However, they

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714145

also do not lead to memory degradation or crystallization of events around their content. We also found that the participant's confidence in the accuracy of their recalled content is not affected by reviewing the summaries. Nevertheless, an interesting future problem to tackle is the users' massive overestimation of the accuracy of what they remember (over 75% of the time). The algorithm we developed was able to pick almost equally relevant and/or important events as well as irrelevant ones to the users that were forgotten. However, it did not succeed in making them remember those forgotten ones. Through these results, we show that the proposed research methodology can successfully automate the evaluation of interventions' effects on the autobiographical memories while being positively perceived by the users for protecting their privacy. Our work sheds the light on the need to rethink the designs of memory augmentation prostheses.

2 Background and Related Work

In this section, we review the benefits and common use cases for lifelogging technologies, specifically focusing on pictorial lifelogs to justify our choice for the technology. Next, we highlight common privacy concerns and approaches to overcome when using lifelogging solutions. Lastly, we review standard memory research methods in HCI and psychology justifying the extension of free recall interviews to support ubiquitous ground truth evidence.

2.1 Benefits of Lifelogging

Niche memory aids such as reminders and calendars focus on supporting prospective rather than retrospective memory problems and are widely adopted and integrated into the user's routine. Prospective memory failures are ones related to remembering to do future events such as going to a class in the evening or buying groceries. Retrospective memory failures are ones related to past information about locations and people such as forgetting someone's name or job. Chen and Jones [9] surveyed prior literature to identify user requirements for memory augmentation in daily life scenarios. They showed a need for circumventing retrospective memory problems using lifelogs because: 1) they are more commonly reported in prior work compared to prospective errors (e.g. [20]) and 2) users are more likely to benefit from a system covering retrospective memory problems as they are aware that they happen and would ask the system for help, unlike prospective memory issues that require intelligent proactive systems. Therefore, we built our lifelogging summaries to circumvent retrospective memory failures. Additionally, lifelogs can circumvent forgetting resulting from encoding problems [28]. Thus, a large number of works report on the benefits of existing lifelogging systems in augmenting the human cognition, specifically in the context of memory augmentation (see [10, 27, 28, 53] for examples). Example benefits include: supporting emotional growth, reflection and enhanced reasoning about past experiences, providing motivational cues to future actions (e.g. going to the gym), and supporting the recall of memories whether by providing cues to incidents (e.g. an old picture of a past trip with a friend) or by repetitively reviewing key incidents to better remember them. Thus, our experiment focuses on supporting retrospective memories. Lifelogs are not only used as memory prosthesis but also as a component of personal informatics systems. The difference is that personal informatics systems primarily focus on reflection to identify areas of improvement and shortcomings leading to behavioral change (see [10] for examples), while memory prostheses generically target altering memories for a variety of reasons such as reflection, behavioral change, trauma attenuation, and re-usage of past skills.

In this work, we log near-continuous photo captures and GPS location. We chose pictorial lifelogs and not other types such as audio recording as images promote more detail-rich recall than other types of data [24, 32] as they contain rich contextual information [36]. Specifically, episodic memories are sensitive to visual simulations [11, 28]. On the other hand, they provide a good trade-off for privacy protection in sensitive situations compared to other comprehensive formats such as videos and audio recordings. Thus, pictorial lifelogs have been extensively researched as a means to augment human memory (e.g., [6, 16, 26, 37, 52, 53]). Pictorial lifelogs have been extensively investigated for helping patients with medical memory impairments (e.g. [44, 47, 48]) and for supporting the memory of laymen (e.g. [32, 40, 42]). Early studies on lifelogging like the pioneer SenseCam ones were centered around enhancing autobiographical memories through showing raw daily reviews (e.g. [21, 29, 52]). However, as the research progressed in the area, domain-based memory augmentation scenarios were developed (e.g. to motivate running [7], supporting education [5], facilitating work productivity [13], and food logging [41]). The current body of work focuses on generically enhancing reminiscence rather than enhancing retrieval of specific forgotten information on demand when evaluating the memory augmentation benefits of lifelogging systems [28]. Our automated summaries deployment focuses on enhancing generic recall while the "Pixel Memories" promotes reminiscence and reflection as a side effect to participating in the experiment.

We chose using near-continuous pictorial lifelogs as a component of proposed research protocol for two reasons. First, it enables us to passively collect large datasets of relatively objective ground truth data to validate narrated memories. Second, the technology is still niche which can incentivize participation in memory studies to try it. This could be used to diversify the demographics of the participants. We chose to build the summarization system because it is among the most commonly used approaches for creating prosthesis for memory augmentation. Thus, the reader can see as a plus point an application for using the research protocol while focusing on the main contribution of the paper, understanding the memory impact of using daily summaries on people.

2.2 Privacy of the Bystanders in Pictorial Lifelogs

The continuous capture of pictorial lifelogs poses a significant challenge for the protection of the privacy of bystanders. Privacy infringements could happen through human consumption (others seeing private or uncomfortable content) or via computer vision attacks. This work focuses on the former, i.e. *infringements from human consumption*. The privacy of bystanders is primarily protected by social conventions allowing them to opt out of the wearable capturing space [22]. However, this is often a challenging task for the lifelogger to remember to disable their capturing device or to search for the violation after it happens. Hoyle et al. showed that lifeloggers are willing to apply propriety preferences to discard or modify photos infringing on the bystander's privacy if the detection of bystanders is automated [30]. Protection methods include: physically marking objects that should not be captured [50], automated activation of personalized capturing policies in specific contexts [2, 55, 57], automatic or manual deletion of content upon detecting certain cues [35], and obfuscating parts of the photos [35, 38, 39]. We recommend the reader to check [12] for an in-depth review of the privacy challenges associated with lifelogging.

Our work protects the privacy of the bystanders and the participants by prohibiting the researchers from seeing the captured ground truth lifelogs of the participants. However, as a limitation to our methodology, we still rely on the participant's discretion in respecting the capturing wishes of their bystanders and co-participants.

2.3 Standard Methods in Memory Research

We encourage interested readers to review [4, 49] for an in-depth historic account of memory research methods in cognitive psychology. In this section, we review two common approaches to studying the human memory: (1) recognition and recall tasks vs. (2) Ebbinghaus tasks. Recognition refers to verifying a detail about an incident after seeing a clear cue about it or associating a cue directly with a past incident. An example is recognizing the face of a criminal in the suspects' line up. Recall refers to retrieving details that are not directly present in the available cue but are mentally associated with it. An example is seeing a generic red t-shirt while shopping then remembering that the spouse's favorite type of t-shirt is red and that they wore one to the first anniversary. The generic red t-shirt is not the same as the spouse's but the color and cut similarity trigger an older memory. The metrics used in this approach are usually the number of correctly recalled events and the number and/or type of correctly recalled details of the events. The *recognition and recall* approach focuses on whether the information is found in our brain or not. The *Ebbinghaus* approach views episodic memories differently. In this approach, the memory is not only formed from information snippets but also from the associations between them. Thus, it focuses on detecting traces of such associations even if they do not directly lead to efficiently remembering the correct piece of information. Thus, the metrics used in this approach are usually the savings in the time to learn and relearn a memory till it is recalled with 100% accuracy. Both approaches were traditionally done on artificial lab stimuli consisting of lists of words. Later, the same tasks were replicated to other stimuli. The recognition and recall approach is more commonly used in psychology and HCI literature with a plethora of tasks to do it. Thus, our work mainly relies on it while partially incorporating Ebbinghaus's philosophy by also tracing and evaluating semi-correct answers to recalled memories.

Our work is largely inspired by free recall tasks and interviews promoting free recall of episodic memories as it is widely used in the HCI community to account for the complexity of investigating memory alterations in natural settings. Prior work standards (e.g. see [36]) ask participants to freely recall and narrate memories with as many details as they can. Afterwards, researchers code the recalled content to score it against a set of predefined memory elements such as locations and emotions. This scoring provides an indication of the quality of the recalled memory. The memory elements could be used to compare several memories and draw correlations with particular interventions. For example, participants tend to recall emotions with condition one in an experiment while they tend to recall locations in condition two. Our work uses similar details described by Le et al. (i.e. the recall of location and time) to score the recall quality and automate the data collection of ground truth. However, it contrasts it in that it allows free recall of events but cued recall of details as the system asks the participants to provide specific details such as location and time. We opted for the semi-cued recall of details to simplify the automated scoring process of the memory quality while preserving the participant's privacy. We specifically chose to extend this method as the rich naturally unverifiable narrative could benefit from the passive collection of ubiquitous evidence to verify it and better interpret it. Additionally, the coding task of the interviews is time-consuming and laborious. Thus, it could benefit from automating it.

3 Prototypes

We present here two prototypes: 1) a system for summarizing the daily photo lifelogs into a short timelapse and 2) a system for browsing the weekly lifelogs and searching for specific photos. Appendix B shows an overview of the system architecture for both prototypes.

3.1 Part 1 - Participant's Prototype: Timelapse Summary Generator

This part explains the lifelogging system used to capture the photos and summarize them into a timelapse video. We use the photos captured by a chest-mounted wearable camera (narrative clip) that automatically captures a photo every 30 seconds and the GPS location to generate the summaries.

3.1.1 Selection Algorithm. We tested previously proposed event segmentation algorithms (e.g. [17]) and found them ineffective for NarrativeClip[1] images. Unlike the stable images from the Microsoft SenseCam, which uses a fish-eye lens and is worn on a necklace, the NarrativeClip is clipped to users' shirts, causing unexpected shifts due to user movements (e.g., shirt becoming more loose when user is sitting). Therefore, we used a custom event segmentation approach that is resilient to short interference scenes, such as temporary camera movements. We define an event as a collection of images that capture visually similar scenes or were taken at the same location (e.g., a picnic, bike ride, or dinner). In short, our approach first clusters images using MPEG-7 descriptors and GPS location, ignoring their chronological order. This results in visually similar images taken at the same location being grouped together, but interference scenes from the same event might be split into separate clusters. To address this, a second step merges clusters that are close in both location and time. A heuristic is used to decide, based on cluster size, whether to merge clusters or keep them separate. The goal is to identify short spikes in sequential dissimilarity, indicating moments when the camera lens is temporarily occluded or shifted to another angle. For each cluster, up to three representative images are selected based on a score that combines several meaningful

[1] the wearable camera we use

features: the presence of a face, the amount of detail described by the MPEG-7 histogram, the illumination data provided by NarrativeClip metadata, and whether the images were manually captured by a double tap. It is important to note that the selection algorithm is not particularly novel but uses common concepts from existing literature as our main contribution is the in-the-wild aspect of the experiment rather than innovation in summarization techniques.

3.2 Part 2 - Lab Prototype: "Pixel Memories" Lifelogs Browser

As a part of a lab evaluation session, we needed a prototype that enables the participant to search for one photo within their weekly lifelogs (approximately 7000 photos) to represent an event they freely recalled. However, the browsing and search tasks are daunting because of the large data volumes (ca. 10,000 photos per week) and the repetitive nature of the task (on average 20 events per session). Additionally, we were not allowed to filter/group the photos by any metadata to avoid cueing participants during the free recall. Therefore, we built the prototype *"Pixel Memories"* to address this problem (see Figure 1). We will discuss later in Section 4.2.2 why those constraints were necessary. It is important to note that this experiment is mainly about evaluating the first prototype (see Section 3.1) rather than this one.

3.2.1 System Structure. It consists of six large displays arranged in a half circle surrounding the participant. Each screen was built using a 50-inch TV with 4k resolution. Each screen represents a day of the week arranged in ascending order from left to right. The participant sits in the middle of the dome at a distance allowing them to scan all screens at once (i.e. parallel reviewing) and is allowed to walk around to get a closer look (i.e. selective reviewing).

3.2.2 Visualizing the Lifelogs. The photos of a day are displayed as sequential thumbnails in a grid format. All photos are displayed in the original temporal order without filtration nor editing. One page of a single screen (i.e. day) shows a maximum number of 703 photos (19 columns * 37 rows). This is approximately 3.3 cm horizontally X 2.47 cm vertically and 0.1 cm separators. We chose the size of the thumbnails heuristically through pilot experiments on four colleagues to reduce visual overload. The photos by design are small enough to mask details to avoid mental fatigue and show the overall structure of events in the day through visual similarity. However, they are large enough for the viewer to perceive the salient features such as number of people and prominent colors to facilitate the search process. The grid rows are numbered to facilitate limiting the search space if the participant knows the approximate temporal order of the searched event. Additionally, the numbers give an intrinsic sense of the time consumed doing a particular activity. The visualization supports pagination to show the rest of the day. A day on average comprises 2.3 pages. This calculation assumes logging for 14 hours[2]. A 24-hour day requires 4 pages.

3.2.3 Interacting with the Lifelogs. Participants could enlarge a photo for closer inspection using a wireless keyboard and mouse, chosen to facilitate navigation across the large display area. A "find my mouse" feature was added after pilot studies showed participants often lost the cursor. To select a photo, participants double-clicked and confirmed their choice in a dialogue box to prevent false selections.

[2]Assuming continuous logging of an average time = 14 hours (8 am to 10 pm)

4 Methodology

In this work, we conducted an in-the-wild experiment using lifelogging as a memory prosthetic. We envisioned that if individuals could review their daily activities in under a minute to enhance natural memory, applications could range from external memory aids to supporting patients with memory impairments, adherence to medical routines, aiding physicians in differential diagnoses, and assisting psychiatrists in co-analyzing behavioural patterns with patients. However, we questioned whether traditional lifelogging solutions, given today's data overload, would still yield the positive memory augmentation effects reported in prior studies (e.g. [10, 27, 28, 53]). A key challenge in our study was evaluating the impact on memory in a privacy-preserving manner, without the researchers' getting access to the data for validation, which required careful consideration of the experimental design.

4.1 Hypotheses

Our goal is to conduct a privacy-preserving evaluation of how reviewing timelapse summaries of daily activities affects: 1) the formation and retention of autobiographical memories, and 2) the quality of the generated summaries. We know from psychology literature in lab settings that regularly reviewing information helps people remember it better and healthy adults should benefit from it. Thus, we specifically had *five* hypotheses:

- **H1:** Daily summary reviewing increases the quantity of recalled events, i.e. events' number.
- **H2:** Daily summary reviewing increases the quality of recalled events, i.e. accuracy of recalled details.
- **H3:** Daily summary reviewing increases the participants' confidence about the correctness of their recalled events.
- **H4:** Daily summary reviewing increases "events crystallization", i.e. recalling more events related to the summary photos, while forgetting more about other events.
- **H5:** The selection algorithm of the reviewed photos centralizes around important events only and can predict them.

4.2 Experimental Design

We conducted a within-subject design experiment with a primary independent variable SUMMARY-DAY representing if participants see the summarized timelapse of the day (Condition 1: review) or they do not get a summary as a baseline (Condition 2: no-review). Figure 2 summarizes the study flow.

4.2.1 Experimental Block. The experiment ran for four weeks. The participants lifelogged their day through a wearable camera and provided our system with the photos to generate a daily summary. The experimental block is a week, where participants are exposed to one condition (e.g. viewing the summaries), followed by another three days of the other conditions and a day of break then an evaluation lab session. We counterbalanced the conditions' order between participants and between the blocks.

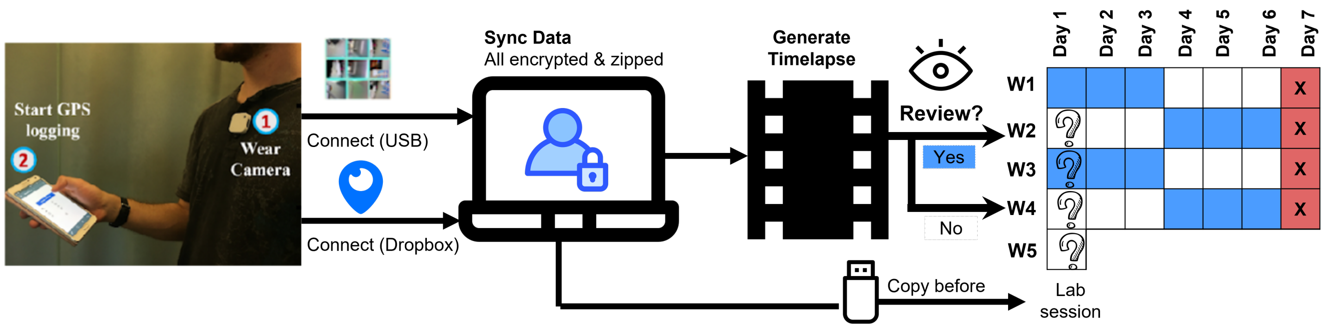

Figure 2: Summary of experimental flow from the participant's perspective

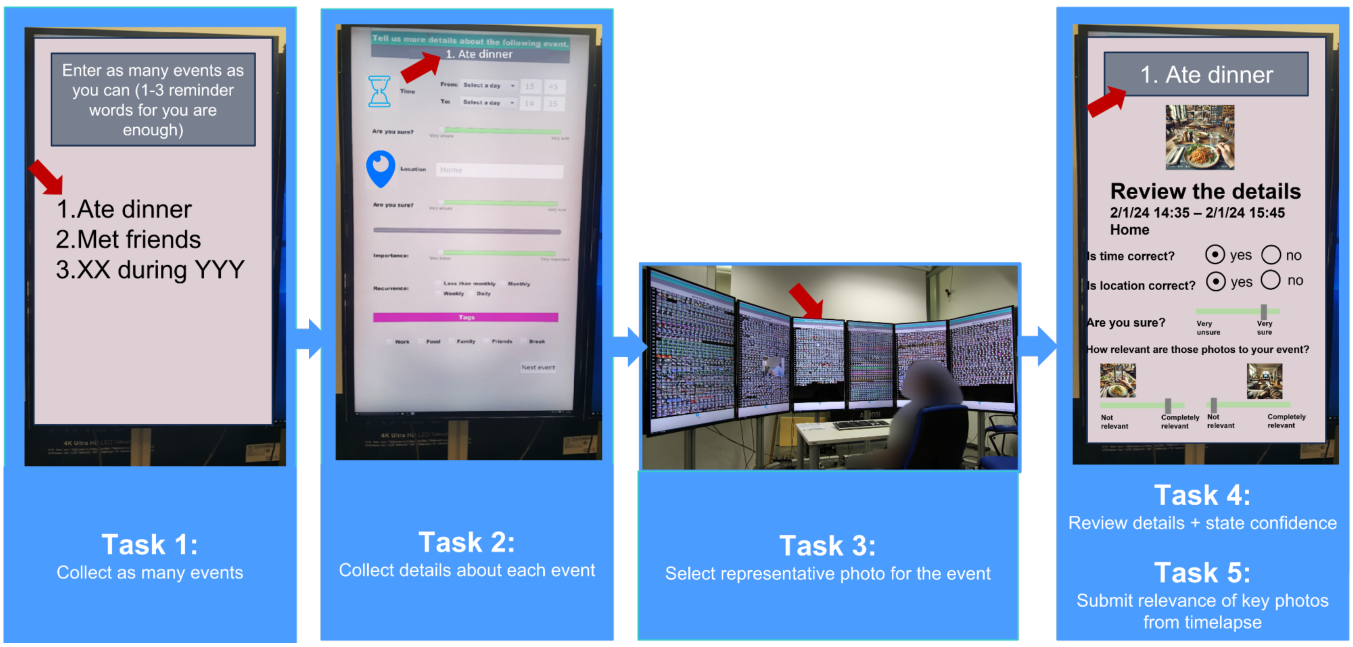

Figure 3: Flow of the lab session. The data is retrieved from the USB the participant provides before the lab session. Task 2 and Task 3 are real screenshots while the rest are mockups for how the system looked like. The tasks are summarized in Table 1.

4.2.2 Lab Session Design. Figure 3 summarizes the session flow. The lab session goal was to examine the impact of the system on the recall quantity and quality by collecting free-recall events while preserving the participant's privacy. The main idea was: 1) collect information from the participant, 2) collect relevant sensor data to the information, 3) collect the subjective relevance of the technology to the memory changes, 4) automatically validate the information to evaluate memory qualities without the researcher seeing it nor keeping full logs of both data types. It is important to note that this approach is *novel*. The session has four views without time constraints (except view 1).

First (see Figure 3: Task 1), participants recall as many events as they can in 8 rounds * 30 seconds (corresponds to approximately 20 events [45]) and we log only the character count. *A recalled event* refers to an incident freely recalled by the participant within a single day. This approach is inspired by the well-known memory task of time-limited free recall of autobiographical events [4, 45, 49]. The selected intervals are because the spontaneous recall rate drops to about one item every 30 seconds after four minutes [45]. The original method does not mandate time subdivision (e.g., 30-second rounds), but we adopted it to reduce overthinking about event granularity, sustain participant engagement, and minimize unnecessary detail in descriptions, as these are not reviewed by researchers.

Afterwards (see Figure 3: Task 2), the participant provides additional details about the start and end timing of the event (date and time) as well as the location of each recalled event (example metrics [32, 36, 52]). Additionally, they report on the importance of the event (numerical scale 1-100) and how recurrent it is. We log the recalled time and the number of characters in a location –like the event content– to conceal it. This approach is inspired by the well-known UbiComp and psychology evaluation method of cued recall of details [32, 52].

Next (see Figure 3: Task 3), the participant is shown the lifelogs from the previous experimental week (6 days) using the "Pixel Memories" prototype (see Section 3.2) to select one representative photo per recalled event (we refer to it as *recalled photo*). Its metadata such as timestamp and location is used to evaluate the corresponding recalled memory elements. The metadata is hidden from the participants to avoid cueing them. We log only abstract data such as the difference between the recalled time and the photo's timestamp. Alternatively, participants can tick a box indicating that they did not find a representative photo. This approach does not directly map to an existing literature method.

In the final view (see Figure 3: Tasks 4 and 5), the participant uses the recalled photo as a cue and reports any incorrect time and location details in the second view. They indicate the event's relevance (numerical scale 1-100) to two key photos from the summary, i.e. the closest prior and following photo to the recalled photo. It is worth noting that we show the key photos from the summaries in both review and no-review days, i.e. not all key photos have been seen before by the participant. The mix gauges the selection algorithm's quality. We collected the relevance as a percentage continuum for better distribution of collected data [43]. The reviewing approach is based on [18], the psychology standard for measuring memory changes post-intervention, and the "remember, know, and guess" methods [23]. However, the relevance approach does not directly map to an existing literature method. Similarly, we collected the participants' confidence about the recalled time and location separately (numerical scale 1-100).

After the lab session, the researchers score the collected details (events and their features) against the evidence (metadata of the recalled photos) *programatically without seeing the data* and log abstract correctness scores. We selected the accuracy of recalled times as an example analysis metric. However, other researchers can use alternative metrics. This approach is inspired by the coding and scoring phase of the free recall contextual interviews (e.g. [36]). However, we do not code the qualitative data manually and we validate the correctness rather than considering only if the detail is present or not like the original method. We present later in Section 4.6.1 the validation process.

4.2.3 Interviews Design. We also conducted two semi-structured interviews, in the first and fourth lab session to understand user expectations of the technology, the attitudes towards privacy and usage of data, and the challenges of lifelogging. We used almost the same questions based on the context. Appendix A shows a list of the questions.

4.3 Apparatus

Figure 2 provides an overview of our system. Each participant is given a chest-mounted wearable camera (Narrative Clip brand) set to automatically capture a photo every 30 seconds, a commercial GPS tracker software, a USB stick, a laptop to generate and view the timelapse summaries, and the respective chargers. Participants synchronize their camera with the given laptop every night and our software generates the timelapse (see selection details in Section 3.1). The next day, the system prompts the user to watch the created timelapse and we log interaction metrics like when the timelapse was opened and if the participant finished it. The timelapse of a day was always shown the following day to encourage maximum logging time and to overcome technological limitations in required processing time[3]. The photos were not accessible during the weekdays to eliminate learning effects. All key photos were shown for an equal amount of time (3 seconds per photo [36]). The timelapse and the raw photos are stored in encrypted formats.

The participant copies the encrypted data via the given USB stick to the lab session. The data is decrypted on the lab computer to run the lab session described above in Section 4.2.2 without the researcher seeing it. The described views run on the same hardware used for the "Pixel Memories" (see Section 3.2: six 50-inch displays and a PC with high graphics card) within the same integrated software. All the abstracted and anonymized data from the lab session is collected to a master database hosted on the institute's server then the researcher deletes all the data from the lab's computer and the USB under the participant's supervision.

4.4 Procedure

We provided an individual orientation session for each participant about their legal and ethical rights as well as the co-participants' and strangers' rights during photo captures and gave them the hardware. We explicitly encouraged participants to practice mindful behaviour while wearing the camera. Participants signed consent inform in the session. It lasted for approximately an hour. We instructed the participants to wear the camera and activate the GPS logger at all times except in private moments and during sleep. We left the definition of private moments to their discretion. Every day, participants logged, synchronized, and generated timelapses that were only seen in specific experimental days. The timelapse was viewed only once. Every eighth day, the participant came to an individual lab session. The researcher would copy the weekly data and operate the lab software that starts with an empty screen then the researcher goes behind the screens to avoid seeing any content. The participant starts the experiment and is encouraged to make vocal comments during the session and we recorded the notes. The experiment also ends with a blank page to enable the researcher to do the data cleanup.

At the end of the last lab session, the researcher would thank the participant, take back the hardware, and provide a monetary compensation for the participation in the study. Afterwards, the researcher would ask the participant if they would like to keep a copy of their one-month lifelogs. If they wished to, the researcher would decrypt the dataset on the given laptop and transfer the data to the participant's chosen storage device. The researcher would also ask the participant if they were willing to donate their collected photos to the research team for further analysis, explicitly explaining that this is completely voluntary and would not impact the compensation nor entail extra compensation. If they agreed to this, they were encouraged to review their dataset and remove

[3]The required time to process and generate the timelapse was around four hours on a core i5 laptop. Thus, showing the timelapse on the same day required the participant to plan their sleeping schedule to account for synchronization and generation of the timelapse.

sensitive information and were asked to sign another release consent form highlighting the acceptable usage level of data[4]. The semi-structured interviews were done at the end of the first and fourth lab sessions. An ethics approval was obtained to run the study.

4.5 Participants and Recruitment

The experiment lasted for four weeks with a total of four lab sessions and twelve reviewing timelapse sessions per participant. Three participants withdrew from the study within the first week due to concerns over potential work privacy breaches, leaving a total of 11 participants for the one-month experiment. We recruited the 11 participants (5 females, 6 males) through Facebook groups of university students and generic expat groups. The mean age of participants was 24.63 years ($min = 19, max = 28, std = 3.38$). The participants' pool included seven nationalities from Europe, Asia and the Middle East. Although participant background diversity was not a primary focus, it provides additional value by reducing potential cultural bias. We excluded participants with full-time jobs to eliminate professional confidentiality breaches. We reimbursed the participants 200€ for their participation. We provided no additional compensation for participants consenting to donating their captured datasets (photos, GPS data and timelapses). On average, the participants spent 20-30 minutes daily to synchronize the data, charge the devices, and view the timelapses. A lab session lasted for two hours on average and the interviews lasted for approximately extra 30 minutes each.

4.6 Analysis

We had two types of analysis: quantitative analysis for validation of the recalled information and evaluation of the changes in recall and qualitative analysis for understanding the users' impressions about the experience.

4.6.1 Quantitative Analysis. We applied here the core idea of the paper, i.e. automatically validating the recalled information through the collected sensor data (metadata of the recalled photos) without the researchers' intervention. We later explain in detail the compared parameters before each hypothesis evaluation in Section 5.

We used Bayesian Factor Analysis[5] to understand if the reviews have no impact on the memories of our participants or not. On the contrary, frequentist inferential statistics would have only allowed us to detect if their usage makes a difference, as they only detect differences between groups by rejecting the null hypothesis. In the model, we employed a standard non-informative prior where the probabilities are distributed equally. The analysis was conducted using Jasp software [31] where the corresponding Bayesian version of all appropriate regular frequentist tests was used [31, 46, 51].

Data Processing We first filtered the data by dropping the first session from all participants to avoid skews related to novelty effect / lack of training. Thus, we analyzed the data from three experimental sessions covering a duration of three weeks. We also removed all events that participants reported were spanning several days (38 events) or where their start time was older than their end time (15 events). We analyzed a total of 737 events.

Variable Names and Abbreviations We use *IV* to denote the independent variable(s), and DV to denote the dependent variable(s) in the test analysis. The first IV is SUMMARY-DAY and it has 2 conditions: review, indicating a day where the participant was presented with a summary video, and no-review, indicating a day where the participant did not get a summary video. The second IV is the LAB-SESSION and it has 3 conditions: Lab sessions (2 to 4) included in the data analysis.

4.6.2 Qualitative Analysis. We only analysed parts of the interviews that were relevant to the paper's narrative. The goal from looking at the data was: 1) identify the privacy concerns from the user's perspective and the co-participants' perspective, and 2) to identify use cases for "Pixel Memories" and understanding the preferred visualization for promoting reflection from the experiment. For the first part, we did semi-closed coding for the data from the following codes: users-perspective concerns and co-participants concerns in addition to a theme. Afterwards, we grouped the themes and chose themes that were represented at least once. Two interviews were coded independently by two researchers and discussed then the rest of the interviews were coded by one of them. The coding is relatively straightforward so we did not opt for more coders. For the second part, we had a question asking participants directly about their preferred method to review their weekly activities: timelapse summaries, "Pixel Memories", or the free recall of events in the lab session. Thus, we directly counted the frequencies. One researcher did semi-closed coding, looking for each method's pros and cons. We reported on themes that appeared at least once. The purpose of the coding is not to show the frequency of the themes due to the small sample size and the variation of the reports but rather to give an intriguing account of potential issues.

5 Analysis and Results: Impact Of Summaries on Memories

We summarize here our quantitative results from examining the hypotheses proposed in Section 4.1.

5.1 Did the Summaries Impact the Quantity of Recalled Events?

Analysis: For each participant, we counted the total number of events recalled in review and no-review days. Afterwards, we normalized them by calculating the corresponding percentage compared to the total events recalled by the participant. We also calculated the percentage of recalled events within each lab session.

Results: The Bayesian Paired Samples T-test (IV= SUMMARY-DAY, DV= percentage from overall events) indicates anecdotal evidence that showing the summaries did not affect the amount of recalled events ($BF_{10} = 0.398$). Another Bayesian Repeated Measures ANOVA (IV= LAB-SESSION, DV= percentage from overall events) indicates substantial evidence that there is no difference between the quantity of recalled events across the lab sessions of the experiment ($BF_{10} = 0.242$).

[4] Anonymous analysis and/or usage in publications.
[5] The ascending ordinal magnitude of evidence as used in this paper: no evidence < anecdotal < substantial < strong < very strong < decisive. Some resources group them to weak, moderate and strong.

■■ *Takeaway message 1:* There are no measurable memory augmentation effects resulting from using summaries as memory prostheses in terms of remembering more items. However, the results are not decisive. Nevertheless, we can rule out temporary augmentation effects due to the novelty effects of using the system. Further analysis in the upcoming sections would reveal if this could be attributed to quality of the algorithm (e.g. showing irrelevant events to the participants) or not. Thus, **H1** is *rejected*, i.e. reviewing summaries did not increase the quantity of recalled events.

5.2 Did the Summaries Impact the Quality of the Recalled Events?

Analysis: We use the evaluation of recalled time as an example for how to apply AMV. For each event, we compared the recalled start and end times to the timestamp of the recalled photo (*RP*) that represents the event from the comprehensive lifelogs in the session. Afterwards, we marked it with a OBJECTIVE-CORRECTNESS: 1) correct is an event where the participant accurately recalled the time ($RP >= start\ AND\ RP <= end$), 2) semi-correct is an event where the participant accurately recalled the day of the event but not the exact time ($day(RP) == day(start|end)$), and 3) wrong is an event where the participant missed the correct day and the time of the event. Afterwards, we calculated the Recalled Correctness Score (RCS) that is weighted summation of the OBJECTIVE-CORRECTNESS for each participant. Events marked as correct got 2 points, semi-correct got 1 point and wrong got 0 points. We calculated the RCS for review and no-review days separately for each participant. Higher RCS indicates better quality of recall. The labels categories and the RCS calculation are inspired by [19] but the labels' definition is adapted to the context here.

Results: The Bayesian Paired Samples T-test using (IV= SUMMARY-DAY, DV= RCS) indicates anecdotal evidence that showing the summaries did not affect the quality of recalled events ($BF_{10} = 0.577$). Another Bayesian Repeated Measures ANOVA (IV= SUMMARY-DAY, DV= LAB-SESSION) indicates substantial evidence that there is no difference in the quality of recalled events across the lab sessions of the experiment ($BF_{10} = 0.239$).

■■ *Takeaway message 2:* Similar to section 5.1, there are no measurable memory augmentation effects from using the summaries in terms of remembering further details about an event (namely in this case, an event's time). However, the results are not decisive. Nevertheless, we can also rule out temporary augmentation effects due to novelty effects of using the system. Thus, **H2** is *rejected*, i.e. reviewing summaries did not increase the quality of recalled events.

5.3 Did the Summaries Impact the Confidence of the Participants regarding Recalled Events?

Analysis: We evaluated the median confidence rating about **recalled time** for each participant in review and no-review days. We also calculated the median confidence score for each of the OBJECTIVE-CORRECTNESS to investigate if for example participants are more confident about correct answers. As another application for AMV method, we compared the participants' perceived correctness of the reported times and their objective correctness from the ubiquitous data we collected. Each participant reported two Boolean flags at the end of each event indicating if they think they recalled the event start time and end time correctly after they were exposed to the comprehensive set of lifelogs. We calculated a metric called SUBJECTIVE-CORRECTNESS that also has three conditions similar to the OBJECTIVE-CORRECTNESS: 1) correct when the participant thinks both times are right, 2) semi-correct when the participant thinks only one is right, and wrong when they think both are wrong. We calculated the difference between OBJECTIVE-CORRECTNESS and SUBJECTIVE-CORRECTNESS (delta) to generate one of three labels: 1) overestimate indicating the participants think they remembered correctly while they did not, 2) underestimate indicating the participants think they remembered wrong while they remembered correctly, and 3) same-as-system indicating the participants' evaluation and the objective evaluation match. Afterwards, we calculated the percentage of events in each of the three categories compared to all recalled events of the participant. We also wanted to see if the confidence patterns changed during the experiment. Thus, we calculated for each of the lab sessions: the median confidence per participant and the median delta. We also evaluated the median confidence rating about **recalled location** for each participant in review and no-review days. We did not further analyze the remaining metrics like the temporal data, as the GPS data was not utilized to assess location accuracy.

Results: The Bayesian Paired Samples T-test (IV= SUMMARY-DAY, DV= median confidence in recalled time) indicates anecdotal evidence that showing the summaries did not affect the participants' perceived high confidence about the correctness of the recalled time for each event ($BF_{10} = 0.395$) ($Mean_{review} = 67.72, Mean_{no-review} = 65.13$ | $Median_{review} = 77$, $Median_{no-review} = 66$ out of 100). Another Bayesian Repeated Measures ANOVA (IV= OBJECTIVE-CORRECTNESS, DV= median confidence) indicates substantial evidence that participants' confidence about the recalled times differs according to the correctness of the answers ($BF_{10} = 4.895$). We used the default T-test with Cauchy Prior for post-hoc comparisons, which the default in JASP software. The evidence regarding a difference in the participants' confidence about semi-correct and either correct or wrong answers in anecdotal. The post-hoc test results suggests that there is no difference in confidence between correct and semi-correct answers ($BF_{10} = 0.855$)) but a tendency of increased confidence in semi-correct answers compared to wrong ones ($BF_{10} = 1.006$). Answers here refer to recalled timings of the events. However, there is substantial evidence that participants are more confident about correct answers than wrong answers ($BF_{10} = 3.758$)($Mean_{correct} = 75.95$, $Mean_{wrong} = 56.64$).

The Bayesian Repeated Measures ANOVA (IV= SUBJECTIVE-CORRECTNESS, DV= percentage of events) indicate a decisive evidence that there is difference in the tendencies of participants to subjectively judge the accuracy of their recalled memories ($BF_{10} = 1.286e+14$) across the three conditions. We followed by default T-tests with Cauchy Prior for post-hoc comparisons. Participants have a very clear tendency to overestimate the correctness of their answers (75% of the cases), followed by judging them correctly similar to the objective evaluation (21.7%), and lastly doubting themselves by dismissing an objectively (semi-)correct answer (2.6%).

We also followed by a Bayesian Repeated Measures ANOVA (IV= LAB-SESSION, DV= median confidence) that indicates anecdotal evidence that there is no difference in the confidence levels of the participants about the correctness of their recalled events across the lab sessions ($BF_{10} = 0.402$). Using another Bayesian Repeated Measures ANOVA (IV= LAB-SESSION, DV= median `delta`) indicates substantial evidence that there is no difference in the participants' subjective perception of their correctness across the sessions.

The Bayesian Paired Samples T-test (IV= SUMMARY-DAY, DV= median confidence in recalled location) indicates anecdotal evidence that showing the summaries did not affect the participants' perceived high confidence about the correctness of the recalled location for each event ($BF_{10} = 0.455$) ($Mean_{review} = 94.04$, $Mean_{no-review} = 93.4$ | $Median_{review} = 98$, $Median_{no-review} = 98$ out of 100).

■■ *Takeaway message 3:* Similar to the previous sections, there are no measurable changes in the participants' confidence about their memories from consuming the summaries. However, naturally without intervention, participants are more confident about correct than wrong ones and can barely differentiate semi-correct memories from wrong ones. Participants also tend to massively overestimate the correctness of their recalled memories. This trend did not get better across the experiment which indicates that the summaries did not particularly cause participants to be more reflective about the potential accuracy of their memories. Thus, **H3** is *rejected*, i.e. reviewing the summaries did not increase the participants' confidence about recalled events. A limitation in contextualizing the results that invites future work is our focus on temporal and geographical confidence alone, excluding other potential memory-related metrics such as associations with people, emotions, and natural elements.

5.4 Did the Recalled Events Crystallize around the Photos in the Summaries?

Analysis: We want to understand here if participants tended to remember only events related to the summaries they saw at the expense of occluding other events. We calculated a Boolean flag (RELEVANCE) for each event to indicate if it is relevant to one of the key photos in the summaries if the participant reported that the event is relevant to one of the closest key photos in time (either the before or after one). It is important to note that we created summaries for every day of the experiment. However, participants saw only half of it to create the SUMMARY-DAY independent variable. Thus, we evaluate here if there is a difference in the amount of relevant and irrelevant events only within the *review* days. Afterwards, we calculated separately the number of events relevant and irrelevant to key photos in the summaries for each participant. We normalized the data by dividing it across the number of events in review days for each participant and not the total number of recalled events.

Results: The Bayesian Paired Samples T-test (IV= relevance in review days, DV= percentage of events) indicates substantial evidence that there is no difference between the amount of relevant and irrelevant events to the key photos in the summary during review days ($BF_{10} = 0.298$).

■■ *Takeaway message 4:* Participants did not forget about other events on the expense of the reviewed ones through the summaries indicating there was no detectable retrieval-induced forgetting in our case [3]. Thus, **H4** is *rejected*, i.e. reviewing the summaries did not lead to event crystallization.

5.5 What is the Quality of the Selection Algorithm of the Summaries?

Analysis: Similar to Section 5.4, we used the calculated RELEVANCE flag but to evaluate the relevance of recalled events to key photos selected by the system in no-review days. Our hypothesis was that if the majority of the events are relevant to the key photos, then the algorithm is good and selecting important events for people. We also calculated the normalized percentage of relevant and irrelevant events for each day type by dividing by the total number of recalled events per participant. Next, we calculated the median importance for the events in review and no-review days.

Results: The Bayesian Paired Samples T-test (IV= RELEVANCE in no-review days, DV= percentage of events within no-review days) indicates substantial evidence that there is no difference between the amount of relevant and irrelevant events to the key photos in the summary during no-review days ($BF_{10} = 0.323$) ($Mean_{relevant} = 52\%$, $Mean_{irrelevant} = 47\%$). To further clarify the results, we followed by a a 2X2 Bayesian Repeated Measures ANOVA (IV= RELEVANCE and SUMMARY-DAY, DV= percentage from total events) which also suggested anecdotal evidence that that there is no difference between the amount of relevant and irrelevant events to the key photos in the summary ($BF_{10} = 0.306$) regardless of the presence of the summaries ($BF_{10} = 0.380$). There was also substantial evidence of the absence of interaction effect ($BF_{10} = 0.116$).

The Bayesian Paired Samples T-test (IV= SUMMARY-DAY, DV= median importance) indicates substantial evidence that there is no difference in the importance of recalled events between review and no-review days during no-review days ($BF_{10} = 0.324$). Participants generally tended to recall important events ($Mean_{review} = 62.545/100\ points$, $Mean_{noreview} = 61.364/100\ points$)

■■ *Takeaway message 5:* The summaries showed relevant and irrelevant events (almost equally on average). Participants generally tended to recall important events. However, reviewing the summaries did not cause them to recall more important or mundane events. Thus, **H5** is *rejected*, i.e. the selection algorithm did not centralize around important events only nor predicts them.

6 Results: Privacy Considerations

We reflect on common privacy themes regarding participants (local storage of data, relaxed accidental sensitive captures) and themes regarding co-participants (agitation from the camera in the beginning, treating the camera as a social entity towards the end). This data should help us better contextualize the participants' experience with using the system.

6.1 Participants' Perspective

Participants were at first concerned about the storage location of their lifelogs and their accessibility to the researchers and/or other parties. Only 27% (3/11 participants) stated they would have participated in the experiment regardless of the storage location, while the rest praised our design decision for storing the data locally on a device only accessible by the participant. They also commented that they would not accept the photos being stored in the cloud even if it was on the institute's secure server. Despite their reservations, participants were generally relaxed about accidental captures of sensitive situations like going to the bathroom. For example, **P11** explains *"A couple of times I forgot to take it off while going to the toilet and I remembered only when I was back. Then I thought, okay if no one else is seeing it then it's okay"*.

Interestingly, the conservative attitude changed through the course of the experiment. At the end of the experiment, five out of eleven participants granted us full access to the recordings of the 30 days (including GPS location, timelapses and comprehensive lifelogs). They refused to review the dataset to remove potentially sensitive information before handing it despite the researchers urging them to do it.

6.2 Co-participants' Perspective

Co-participants are special type of bystanders that are regularly captured such as room mates or partners. Unlike participants, co-participants were initially agitated around the camera. However, the participants could not easily empathize with their concerns. For example, **P11** told us about a dinner in a friend's house *"... because they are not wearing proper clothes and not combing their hair. So, somehow the pictures should not be recorded and then, I explained to them that no one is going to see this pictures except me who is already seeing you in this dress. So it's okay"*. We postulate that the co-participants' rejection is aggravated because of the camera disguise as they thought it is an MP3 or medical device. A potential solution is wearing brightly colored camera. However, this seemed to cue unwelcome interaction from the co-participants and bystanders with the participants. For example, **P9** said *"My friends in class wanted to ask me out of curiosity because okay the color is orange and I didn't have any orange so it was very colorful"*. Similarly, **P8** commented *"sometimes the people in the train looked at me with a strange face, and they looked directly to the camera"*.

Interestingly, co-participants' acceptance of the camera increased over the course of the experiment. They started to positively interact with the camera and treat it as an active entity. For example, **P1**'s friends tried to *"pose for the camera"*. Similarly, **P9** said *"They [his co-participants] say hi to the camera sometimes. They also touch it saying take my pictures please"*. However, new social power dynamics emerged within groups where co-participants remained conservative about the camera usage. For example, **P11** said regarding the same group who refused the recording in the beginning *"I took it [the camera] off for like 30 minutes, because I was trying to gauge their [co-participant's] response. If my friend sounds serious, I take it off but put it on again later"*.

7 Results: Visualizations Promoting Reflection

Initial evaluation shows that 7 out of 11 participants preferred using the large screens setup "Pixel Memories" to review their weekly activities, while 3 preferred the free recall task in the lab session, and 1 preferred both. None chose the daily summaries as means to reflect on their weekly activities. The problem was not being able to recognize the content sometimes and it offered a visualization for single days separately. However, the "Pixel Memories" was chosen as it provides as a holistic view on the data, clearly shows the day activities separately and the photo size was convenient to see everything clearly. On the other hand, free recall provided unusual pressure on the participants to remember their week which they liked. **P10** explains *"here I have to think, because I was told by [researcher X] that I can't look up stuff so I had to use my mind"*. One participant also commented that she did not notice how much time she was wasting on social media until she saw the number of lines wasted on the screen. This shows the potential of using the prototype as a standalone memory prosthetic beyond the original purpose for the research protocol.

8 Discussion and Lessons Learnt

In this section, we reflect on takeaways that can benefit the broader community beyond the scope of this experiment.

8.1 Use the Research Methodology to Evaluate UbiComp Interventions & Advance Memory Research

One *core and novel* contribution of this work is distilling the research methodology, that we call *Automated Memory Validation (AMV)* (summarized in Table 1). It can be used beyond the scope of our study by other researchers to 1) conduct and evaluate other lifelogging experiments, 2) use it in psychology to further conduct memory research, 3) use it in the healthcare domain to help physicians automatically validate patients' narratives, particularly in the field of psychiatry, monitoring of chronic diseases, and monitoring of dietary habits, and (most relevant to this community) 4) study the impact of any technological intervention on the memory to evaluate HCI and UbiComp systems. The idea, as outlined in the study design, is to first collect ground truth data from the participant (phase 1) and the sensors (phase 2) after removing sensitive information. Next, the subjective relevance of an intervention on memory changes is evaluated (phase 3), followed by programmatic validation of the collected data (phase 4).

The key constraint is that the researcher does not have access to the raw data either during the lab sessions (e.g. via monitoring the computer) or after them. Although the tasks build on existing methods, the combination and novel alterations present a powerful, innovative approach to *privacy-aware* memory research. This privacy-aware methodology reduces the omission of important memories by easing concerns about exposing sensitive information. It speeds up evaluation by eliminating manual coding, though at the cost of potentially deeper insights. It also lessens researchers' legal responsibilities under regulations like GDPR by avoiding sensitive data storage. However, it places significant responsibility on participants to manage co-participants' privacy, which complicates informed consent and could lead to ethical concerns such as post-removal of sensitive data of co-participants on demand, proxy reliance of co-participants on the presence of the technology to capture "their lives", accidentally or deliberately recording co-participants' sensitive data. Similar to other privacy-sensitive technologies like social media, societal acceptance evolves over

Phase	Experimental Task	Inspired By	Our Alterations (Novelty)
1. Collect participant's ground truth	1. Participant recalls as many events as possible in 8 rounds * 30 secs (ca. 20 events)	Psychology task: time-bound free recall of autobiographical events [4, 45, 49]	*Abstract logging*: Log only number of events not events' content
	2. Participant fills details questionnaire about each event (e.g. time, location). Choose details that can be validated from sensors.	Psychology task in Ubicomp evaluations: cued recall of details [32, 52]. Details' examples: [32, 36, 52]	*Abstract logging*: Log only pointers to details like number of characters for location or correctness score
2. Collect sensors ground truth	3. For each recalled event, the participant selects one lifelog photo to best describe it without seeing any metadata (recalled photo).	X	Novel idea
	4. Participants are shown the the recalled photo (Task 3) as a cue to review the details (Task 2) + report their confidence about the correctness of the answers.	Psychology & HCI standard: pretest-posttest design + the remember, know, and guess memory method [23]	We use a recalled item (the recalled photo) to be a new cue (stimuli) rather than reviewing without it.
3. Subjective relevance of intervention on memories	5. Participants rate the relevance of technology cues (e.g. parts of the lifelog summaries) to the events they recalled.	HCI literature: Use continuum scale (1-100) [43]	Novel idea
4. Validate the data programmatically	The researcher codes formulas to compare the metrics from Task 2 & 3 automatically + labels the output as: correct, semi-correct, or wrong. Data from Task 4 & 5 help the researcher understand the subjective impact of the intervention	Psychology & HCI literature: Coding and scoring phase of the free recall contextual interviews [36]	The metadata and/or photo content (Phase 2) validate the information from Phase 1. No manual open-coding + Details are predefined (Task 2) + Details are validated for correctness, not just presence.

Table 1: The proposed Automated Memory Validation (AMV) research method. The mix of methods and the alterations is novel.

time, shaping new norms for data management. This progression was evident during our one-month deployment, as participants began perceiving the camera as a "social entity". It is worth noting that AMV cannot validate subjective experiences directly, but advancements in Human Activity Recognition (HAR) broaden the range of validatable parameters. The integration of tools like the "Pixel Memories" setup may also enhance participant engagement with existing standardized approaches. The method is also sensitive to data loss from inconsistent user compliance to tracking or external factors such as hardware malfunction or poor synchronization. A potential solution is to use environmental lifelogging [19] as a backup in critical use cases. However, even with data loss, it provides an improvement over classical methods outlined in Table 1, where ground truth data is absent.

To sum up, AMV method is superior when the main goal is to protect the participants' privacy and/or have objective metrics to validate memories. However, classical methods such as time-bound free recall and cued recall of details in interviews are superior for speedy short evaluations without ground truth or complex use cases where detail depth is needed and highly trained staff is available or when the ground truth is missing for any reason such as hardware failures or the user stopping captures.

8.2 Explore "Pixel Memories" as a Tool for Self-Reflection

Although initially developed only as a part of the experimental apparatus, the "Pixel Memories" setup (see Section 3.2) turned out to be one of the most exciting outcomes of this work worth picking up by other researchers. It effectively addressed the challenging task of selecting a representative photo from a set of approximately 7,000 images without using metadata filters. Participants were able to complete this task up to 20 times in under 20 minutes, with no reported difficulties . Additionally, participants used the setup for spontaneous self-reflection, reviewing how they had spent their time, despite this not being part of the study design. Some even expressed excitement about attending lab sessions just to see how their week went. Informal setup demos consistently impressed attendees with how much information could be seen at a glance, even from datasets of strangers. We encourage further exploration of use cases for this setup and variations, such as using a projector for home settings or virtual reality implementations to maintain the immersive dome setup and investigating its impact on the user experience. Our setup is *novel*. The closest setup to the "Memory Pixels" is the "color of life wall"[33] (see website and demo [25]). However, the "color of life" shows the dominant color(s) of each photo while we show the full photo, which enables our users to decode other cues like faces and time of the day and use it for searching and activity recognition during reflection.

8.3 One-Minute Timelapses are Not the Magical Memory Pill

Our results were surprising as all proposed hypotheses (see Section 4.1) were rejected. The analysis suggest that the designed timelapses were not successful in imparting knowledge on the participant and acting as a memory prosthesis triggering memory augmentation neither in the quantity nor the quality of the recalled memories (**H1 and H2**). Luckily though, they did not cause confusions nor contribute in memory degradation nor alteration. We also expected to observe a novelty effect during the experiment where participants tended to remember more in the beginning because of being excited about the system and because of knowing their memories are being tested. However, the data showed otherwise and there were no changes in their recall patterns across the experiment.

Participants are also naturally overly-confident about the quality of their recalled memories. While prior work (e.g. [1]) show that lifelogs could alter participant's perception of what happened in specific use cases, we did not observe that effect with natural data here from consuming the summaries. We expected participants to become more reflective about their memory errors as a side effect during the course of the experiment as they saw more content about their lives. This also did not happen and the confidence patterns remained unchanged, particularly about the time and location of events (**H3**). It is worth noting though that the focus of the experiment was not to highlight "overestimation of correctness" problem. Thus, we did not attempt to explicitly visualize it for them.

While one might think that this could be attributed to the quality of the selected events in the summary, our results suggest otherwise. The summaries already contained relevant and important events. Thus, we cannot claim that they were focused for example on the very important events that participants would have remembered anyways nor on the mundane events that they did not care about (**H5**). However, we can see that there were no differences in recall quantity and quality from consuming them anyways. A limitation of our study is that we did not experiment with various frequencies for repeating the the reviews of the summaries. Thus, we cannot make claims at the moment if reviewing them more would still lead to the same findings or not. An interesting design aspect for future exploration is the presentation of summaries on the following day, which may negatively influence the quantity and quality of recall, as sleep plays a crucial role in consolidating newly acquired memories [34]. However, our approach increases the relevance of our results to the community as it aligns with the current commercial practices, where personal photo-based reminiscence tools present content on subsequent and often random later days.

We also suspected that participants might tend to remember mostly content from summaries because it is being reviewed rather than having a holistic overview of events that happened to them. However, the data proved otherwise that there were no event clustering detected successfully in our experiment (**H4**).

These findings show that while next-day summaries were not particularly useful in augmenting the memories, they did not also threaten the natural recall of the participants. Thus, one might use them over platforms if users feel engaged and interested in them. However, they are not the right tool to combat natural forgetting.

8.4 Informed Consent and Diversity of Physical Features are Challenging

We had two challenges that easily apply to other lifelogging experiments. The first is *informed consent*. Participants underestimate what is captured despite detailed explanations. In our study, several participants dropped out after unintentionally recording sensitive work material. To prevent this, researchers should discuss participants' job tasks, recommend obtaining manager consent, and provide a testing period for participants to adjust to the camera before starting the study. The second challenge is *non-inclusive camera design*. Participants with longer hair and/or fuller upper bodies reported significant discomfort wearing the camera and poor photo quality, with many images being obstructed or misaligned, often capturing the ceiling instead of their surroundings.

8.5 Limitations and Future Work

One limitation of our study is the participant sample, which did not include individuals with full-time jobs. This tradeoff was made to minimize privacy risks after early participants unintentionally captured sensitive work content. However, this restriction limits the generalizability of the findings to broader professional settings. Additionally, all participants were expatriates, as local individuals, who were more privacy-aware regarding lifelogging, declined to participate despite recruitment efforts. This may slightly influence the participants' perception, as cultural norms and societal integration could affect their views on lifelogging. Another limitation is that the AMV method used in the study is more privacy-preserving but also crude compared to manual qualitative coding by skilled researchers. While this study serves as a proof-of-concept for the method, claims about memory effects should be interpreted cautiously, especially when relying on time-based metrics that require high accuracy and relatively smaller sample sizes. For example, we did not evaluate whether the recalled events were true or not as the current standard methods (e.g., classical psychology approaches referenced in Table 1) do not require such validation, and the concept of "true" is inherently relative. Nonetheless, our designed protocol includes a proxy for this. Participants can indicate instances where they "did not find a representative photo" within the "Pixel Memories" and provide a reason. Analyzing the frequency of such instances could offer an educated estimate of "questionable events". However, this approach has limitations as it might also reflect capture failure. We also did not explore potential confusion with other memories. Despite these limitations, our protocol still improves upon classical methods, where ground truth is often entirely absent. Future research could explore new metrics and algorithms for memory evaluation, incorporating additional sensor data to improve precision. Another limitation is that the evaluation of the "Pixel

Memories" prototype remains preliminary, as it was not a primary focus of the experiment. However, the final interviews suggested its potential, and we encourage further exploration of this tool in future studies. Lastly, leveraging generative AI for summarizing lifelogs or refining automatic validation metrics presents a valuable direction for future work. Techniques like OpenAI's CLIP or ResNet models could simplify tasks such as image clustering with minimal fine-tuning, offering promising opportunities for improving the efficiency and depth of lifelog analysis.

9 Conclusion

In this paper, we present a lifelogging deployment (n=11, tested for a month) that produces daily automated timelapses shorter than a minute to review daily activities. We also present a novel research method (*Automated Memory Validation (AMV)*) to evaluate memory prostheses that focus on protecting the users' privacy and automatically validating the research data. We also present a browsing system of large screens to efficiently review a week-worth of pictorial lifelogs. We learnt that in contrast to prior work showing that reviewing helps us remember better (e.g. [28, 52]), the intrinsic memory enhancements in our case were not substantial in real scenarios from basic short reviews despite the relevance of the presented content. Thus, further research is required to identify the required number of review repetitions leading to quantifiable memory enhancements while keeping the system interesting and usable. We postulate that lifelogs are harder in imparting information because participants are not familiar with the photo angles and perspectives. This was reflected in several of our participants not being able to recognize the context of their lifelogs or if the photos belonged to them or not. Our proposed research methodology (AMV) successfully helped us evaluate a technological intervention affecting autobiographical memories while being positively perceived by participants in protecting their privacy. Our work will help future researchers evaluate their memory prostheses, explore new psychological memory constructs, and highlight the still ongoing-challenge in the wide adoption of lifelogging technologies.

Acknowledgments

We want to acknowledge that ChatGPT-3.5 and ChatGPT4o have been used to restructure some paragraphs following current ACM policies. This work is partially supported and funded by the following entities: the Future and Emerging Technologies (FET) Programme with the 7th Framework Programme for Research of the European Commission, under FET grant number:612933 (RECALL), the HumanE AI Network under the European Union's Horizon 2020 ICT program (grant agreement no. 952026), and the AI4EUROPE Project (grant agreement no. 101070000).

References

[1] Delene Adams, H. Paterson, and H. MacDougall. 2020. Law and (rec)order: Updating memory for criminal events with body-worn cameras. *PLoS ONE* 15 (2020).

[2] Paarijaat Aditya, Rijurekha Sen, Peter Druschel, Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele, Bobby Bhattacharjee, and Tong Tong Wu. 2016. I-Pic: A Platform for Privacy-Compliant Image Capture. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (Singapore, Singapore) *(MobiSys '16)*. ACM, New York, NY, USA, 235–248. doi:10.1145/2906388.2906412

[3] Michael C Anderson, Elizabeth L Bjork, and Robert A Bjork. 2000. Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic bulletin & review* 7 (2000), 522–530.

[4] Alan Baddeley, Michael W. Eysenck, and Michael C. Anderson. 2009. *Memory*. Psychology Press Taylor and Francis Group.

[5] Deborah Barreau, Abe Crystal, Jane Greenberg, Anuj Sharma, Michael Conway, John Oberlin, Michael Shoffner, and Stephen Seiberling. 2006. Augmenting memory for student learning: Designing a context-aware capture system for biology education. *Proceedings of the American Society for Information Science and Technology* 43, 1 (2006), 1–6.

[6] Emma Berry, Narinder Kapur, Lyndsay Williams, Steve Hodges, Peter Watson, Gavin Smyth, James Srinivasan, Reg Smith, Barbara Wilson, and Ken Wood. 2007. The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychological Rehabilitation* 17, 4-5 (2007), 582–601. doi:10.1080/09602010601029780 arXiv:https://doi.org/10.1080/09602010601029780 PMID: 17676536.

[7] Agon Bexheti, Anton Fedosov, Jesper Findahl, Marc Langheinrich, and Evangelos Niforatos. 2015. Re-Live the Moment: Visualizing Run Experiences to Motivate Future Exercises. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 986–993. doi:10.1145/2786567.2794316

[8] Georgina Browne, Emma Berry, Narinder Kapur, Steve Hodges, Gavin Smyth, Peter Watson, and Ken Wood. 2011. SenseCam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory* 19, 7 (2011), 713–722.

[9] Yi Chen and Gareth J. F. Jones. 2010. Augmenting Human Memory Using Personal Lifelogs. In *Proceedings of the 1st Augmented Human International Conference* (Megève, France) *(AH '10)*. ACM, New York, NY, USA, Article 24, 9 pages. doi:10.1145/1785455.1785479

[10] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1143–1152. doi:10.1145/2556288.2557372

[11] Martin A Conway. 2009. Episodic memories. *Neuropsychologia* 47, 11 (2009), 2305–2313.

[12] Nigel Davies, Adrian Friday, Sarah Clinch, Corina Sas, Marc Langheinrich, Geoff Ward, and Albrecht Schmidt. 2015. Security and Privacy Implications of Pervasive Memory Augmentation. *IEEE Pervasive Computing* 14, 1 (2015), 44–53. doi:10.1109/MPRV.2015.13

[13] Tilman Dingler, Passant El Agroudy, Gerd Matheis, and Albrecht Schmidt. 2016. Reading-Based Screenshot Summaries for Supporting Awareness of Desktop Activities. In *Proceedings of the 7th Augmented Human International Conference 2016* (Geneva, Switzerland) *(AH '16)*. Association for Computing Machinery, New York, NY, USA, Article 27, 5 pages. doi:10.1145/2875194.2875224

[14] Tilman Dingler and Niels Henze. 2014. That's the Dog from my Wedding – Algorithms for Memory Shaping. Workshop Paper. Adjunct Proceedings. CHI Workshop on Designing Technology for Major Life Events. Retrieved June 24, 2016 from https://sites.google.com/site/techmajorlifeevents/Dinger.pdf?attredirects=0.

[15] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language learning on-the-go: opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services*. 1–12.

[16] Aiden R. Doherty, Katalin Pauly-Takacs, Niamh Caprani, Cathal Gurrin, Chris J. A. Moulin, Noel E. O'Connor, and Alan F. Smeaton. 2012. Experiences of Aiding Autobiographical Memory Using the SenseCam. *Human–Computer Interaction* 27, 1-2 (2012), 151–174. doi:10.1080/07370024.2012.656050 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07370024.2012.656050

[17] Aiden R Doherty and Alan F Smeaton. 2008. Automatically segmenting lifelog data into events. In *2008 ninth international workshop on image analysis for multimedia interactive services*. IEEE, 20–23.

[18] Passant ElAgroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Andreas Bulling, and Albrecht Schmidt. 2019. Can Privacy-Aware Lifelogs Alter Our Memories?. In *Proceedings of the 37th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Glasgow, Scotland UK) *(CHI EA '19)*. ACM, New York, NY, USA, 6 pages. doi:10.1145/3290607.3313052

[19] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Ekta Sood, Andreas Bulling, and Albrecht Schmidt. 2023. Impact of Privacy Protection Methods of Lifelogs on Remembered Memories. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–10.

[20] David Elsweiler, Ian Ruthven, and Christopher Jones. 2007. Towards memory supporting personal information management tools. *Journal of the American Society for Information Science and Technology* 58, 7 (2007), 924–946.

[21] Jason R Finley, William F Brewer, and Aaron S Benjamin. 2011. The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam. *Memory* 19, 7 (2011), 796–807.

[22] I. Flammer. 2016. Genteel Wearables: Bystander-Centered Design. *IEEE Security & Privacy* 14, 5 (Sept.-Oct. 2016), 73–79. doi:10.1109/MSP.2016.91

[23] John M Gardiner, Cristina Ramponi, and Alan Richardson-Klavehn. 2002. Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory* 10, 2 (2002), 83–98.

[24] Rúben Gouveia and Evangelos Karapanos. 2013. Footprint Tracker: Supporting Diary Studies with Lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 2921–2930. doi:10.1145/2470654.2481405

[25] Cathal Gurrin and DCU. 2015. Colour of Life Wall. Science Gallery Dublin. https://dublin.sciencegallery.com/lifelogging-exhibits/cathal-gurrin-amp-dcu-ie Accessed: 2024-09-12.

[26] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Foundations and Trends® in Information Retrieval* 8, 1 (2014), 1–125. doi:10.1561/1500000033

[27] Cathal Gurrin, Alan F. Smeaton, Zhengwei Qiu, and Aiden Doherty. 2013. Exploring the Technical Challenges of Large-Scale Lifelogging. In *Proceedings of the 4th International SenseCam and Pervasive Imaging Conference* (San Diego, California, USA) *(SenseCam '13)*. Association for Computing Machinery, New York, NY, USA, 68–75. doi:10.1145/2526667.2526678

[28] Morgan Harvey, Marc Langheinrich, and Geoff Ward. 2016. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive and Mobile Computing* 27 (2016), 14–26. doi:10.1016/j.pmcj.2015.12.002

[29] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*. Springer, 177–193.

[30] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 2014. Privacy Behaviors of Lifeloggers Using Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 571–582. doi:10.1145/2632048.2632079

[31] JASP Team. 2025. JASP (Version 0.8.2) [Computer software]. https://jasp-stats.org/

[32] Vaiva Kalnikaite, Abigail Sellen, Steve Whittaker, and David Kirk. 2010. Now let me see where i was: understanding how lifelogs mediate memory. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2045–2054.

[33] Philip Kelly, Aiden R Doherty, Alan F Smeaton, Cathal Gurrin, and Noel E O'Connor. 2010. The colour of life: novel visualisations of population lifestyles. In *Proceedings of the 18th ACM international conference on Multimedia*. 1063–1066.

[34] Jens G Klinzing, Niels Niethard, and Jan Born. 2019. Mechanisms of systems memory consolidation during sleep. *Nature neuroscience* 22, 10 (2019), 1598–1610.

[35] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. 2016. Enhancing Lifelogging Privacy by Detecting Screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 4309–4314. doi:10.1145/2858036.2858417

[36] Huy Viet Le, Sarah Clinch, Corina Sas, Tilman Dingler, Niels Henze, and Nigel Davies. 2016. Impact of Video Summary Viewing on Episodic Memory Recall: Design Guidelines for Video Summarizations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (Santa Clara, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 4793–4805. doi:10.1145/2858036.2858413

[37] Matthew L. Lee and Anind K. Dey. 2008. Lifelogging Memory Appliance for People with Episodic Memory Impairment. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea) *(UbiComp '08)*. ACM, New York, NY, USA, 44–53. doi:10.1145/1409635.1409643

[38] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Blur vs. Block: Investigating the Effectiveness of Privacy-Enhancing Obfuscation for Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1343–1351. doi:10.1109/CVPRW.2017.176

[39] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and Users' Experience of Obfuscation As a Privacy-Enhancing Technology for Sharing Photos. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 67 (Dec. 2017), 24 pages. doi:10.1145/3134702

[40] Siân E Lindley, Maxine Glancy, Richard Harper, Dave Randall, and Nicola Smyth. 2011. "Oh and how things just don't change, the more things stay the same": Reflections on SenseCam images 18 months after capture. *International Journal of Human-Computer Studies* 69, 5 (2011), 311–323.

[41] Takuya Maekawa. 2013. A sensor device for automatic food lifelogging that is embedded in home ceiling light: A preliminary investigation. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 405–407.

[42] Ali Mair, Marie Poirier, and Martin A Conway. 2019. Memory for staged events: Supporting older and younger adults' memory with SenseCam. *Quarterly Journal of Experimental Psychology* 72, 4 (2019), 717–728.

[43] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The effect of visual appearance on the performance of continuous sliders and visual analogue scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5421–5432.

[44] Georgios Meditskos, Pierre-Marie Plans, Thanos G Stavropoulos, Jenny Benois-Pineau, Vincent Buso, and Ioannis Kompatsiaris. 2018. Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia. *Journal of Visual Communication and Image Representation* 51 (2018), 169–190.

[45] Bryan J Moreton and Geoff Ward. 2010. Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review* 17, 4 (2010), 510–515.

[46] RD Morey and JN Rouder. 2015. BayesFactor (Version 0.9. 11-3)[Computer software].

[47] Paulina Piasek, Kate Irving, and Alan F Smeaton. 2012. Case study in SenseCam use as an intervention technology for early-stage dementia. *International Journal of Computers in Healthcare 1* 1, 4 (2012), 304–319.

[48] Paulina Piasek, Kate Irving, and Alan F Smeaton. 2014. Using lifelogging to help construct the identity of people with dementia. (2014).

[49] Henry Roediger and Jeremy K Yamashiro. 2019. History of Cognitive Psychological Memory Research. In *The Cambridge Handbook of the Intellectual History of Psychology*, Robert J. Sternberg and Wade E. Pickren (Eds.). Cambridge University Press.

[50] Franziska Roesner, David Molnar, Alexander Moshchuk, Tadayoshi Kohno, and Helen J. Wang. 2014. World-Driven Access Control for Continuous Sensing. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. ACM, New York, NY, USA, 1169–1181. doi:10.1145/2660267.2660319

[51] Jeffrey N Rouder, Paul L Speckman, Dongchu Sun, Richard D Morey, and Geoffrey Iverson. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review* 16, 2 (2009), 225–237.

[52] Abigail J Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do life-logging technologies support memory for the past?: an experimental study using sensecam. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 81–90.

[53] Abigail J Sellen and Steve Whittaker. 2010. Beyond total capture: a constructive critique of lifelogging. *Commun. ACM* 53, 5 (2010), 70–77.

[54] Betsy Sparrow, Jenny Liu, and Daniel M Wegner. 2011. Google effects on memory: Cognitive consequences of having information at our fingertips. *science* 333, 6043 (2011), 776–778.

[55] Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. 2018. PrivacEye: Privacy-Preserving First-Person Vision Using Image Features and Eye Movement Analysis. *CoRR* abs/1801.04457 (2018). arXiv:1801.04457 http://arxiv.org/abs/1801.04457

[56] Diana I Tamir, Emma M Templeton, Adrian F Ward, and Jamil Zaki. 2018. Media usage diminishes memory for experiences. *Journal of Experimental Social Psychology* 76 (2018), 161–168.

[57] Robert Templeman, Roberto Hoyle, David Crandall, and Apu Kapadia. 2014. Reactive Security: Responding to Visual Stimuli from Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1297–1306. doi:10.1145/2638728.2641708

[58] Emma Woodberry, Georgina Browne, Steve Hodges, Peter Watson, Narinder Kapur, and Ken Woodberry. 2015. The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory* 23, 3 (2015), 340–349.

A Pre and Post Interview Questions

(1) Privacy and sensitive information
  (a) How often did you find sensitive pictures in your logs?
  (b) Were you aware of them when they were taken?
  (c) What were they about? (Broad categories)
  (d) Remembering events from last week (phase 1)
  (e) Did you start preparing for the remembering the events after the first week?
  (f) How did you feel about it? Was it boring / fun?
  (g) Was it stressful? When did you start feeling that? (Try to find an estimate for the maximum time you are happy doing the task)
  (h) What type of events did you usually remember? (Broad categories)
  (i) Do you imagine doing it for 3 minutes every day at the end of the day? Why?
  (j) Which scenario would be useful for presenting this data if you do it on daily basis? (Would it increase your sense of achievement?)

(2) Remembering details about events (phase 2)
  (a) How easy was it to remember the time of an event?
  (b) Do you usually remember it from the date or day or as an event? (how do they construct time)
  (c) How easy was it to remember the location of an event?
  (d) Do you usually remember it as a concept or address
  (e) How often did you remember events from past weeks or yesterday? When do you realize it: phase 1 or 2?

(3) Data usage patterns
  (a) Would you like to keep your lifelogging data? Why? What would make you want to keep it?
  (b) Would you prefer to keep the whole data set, the videos only or a another subset? Why?
  (c) Were there instances where you were happy that something got captured? When?
  (d) Were there instance where you wished you had access to the photos and could search it? Give examples
  (e) Is there data from work you would be useful to look up later?

(4) Video quality
  (a) Did the algorithm remove the blurry and dark pictures successfully?

(5) Selection of a photo (phase 3)
  (a) How overwhelming/hard was the task of searching for a photo to represent the event?
  (b) How did you search? (Filter by time or pure visual search .. etc.)
  (c) Which features would have made your life easier?
  (d) Did seeing the photos show you something you didn't know about your days? Examples?
  (e) Did seeing the whole week made you notice something you didn't know about your activities? Examples?
  (f) Would you prefer seeing the whole week or day by day on a single screen?
  (g) Did seeing the photos on a large screen give you a sense of reviewing about your events or you only used it for searching?

(6) Expectations about the technologies
  (a) Did the quality of photos match your initial expectations before the experiment? (Quality as in frequency of capturing, angle of the content ..)
  (b) Did you have any misconceptions about lifelogging before the experiment that changed after the experiment?
  (c) If there were cheap alternatives for the camera, would you imagine doing lifelogging on daily basis?
  (d) Did your attitude towards the camera change from the beginning till the end of the experiment? (E.g: you enjoyed it in the beginning but were bored in the end or vs. versa)
  (e) What did you wish to have in the system but was not there?

(7) Effect on self reflection, achievement and memory
  (a) Which method gave you the best sense of your weekly activities: videos, photos on large screen or recalling events? Why? Did each of them provide you with something different?
  (b) Do you feel better / worse towards how you spend your time during the week after participating in the experiment? Why?
  (c) Did the recall phase give you a sense of achievement?
  (d) Did watching the video affect your mood: postively or negatively?
  (e) Were you always able to know which event does the photo in the video belong to in the day?
  (f) Did you see events in the photos that you have already forgotten but recognized them in the photos?
  (g) Did you see photos where you were not sure which event do they belong to?
  (h) How often did you realize in phase 4 that you have entered incorrect data?
  (i) Was it usually the time or the location?
  (j) How did you realize it? (You remembered on your own, seeing the pictures .. etc)

B Architecture of Prototypes

Fig. 4 gives an overview of both prototypes developed for the experiment, that is the home routine for summarizing the photos through a timelapse and the lab experimental setup including the "Pixel Memories" and the data logging mechanisms for the proposed research method AMV. We highlight in the figure the key blocks in the system.

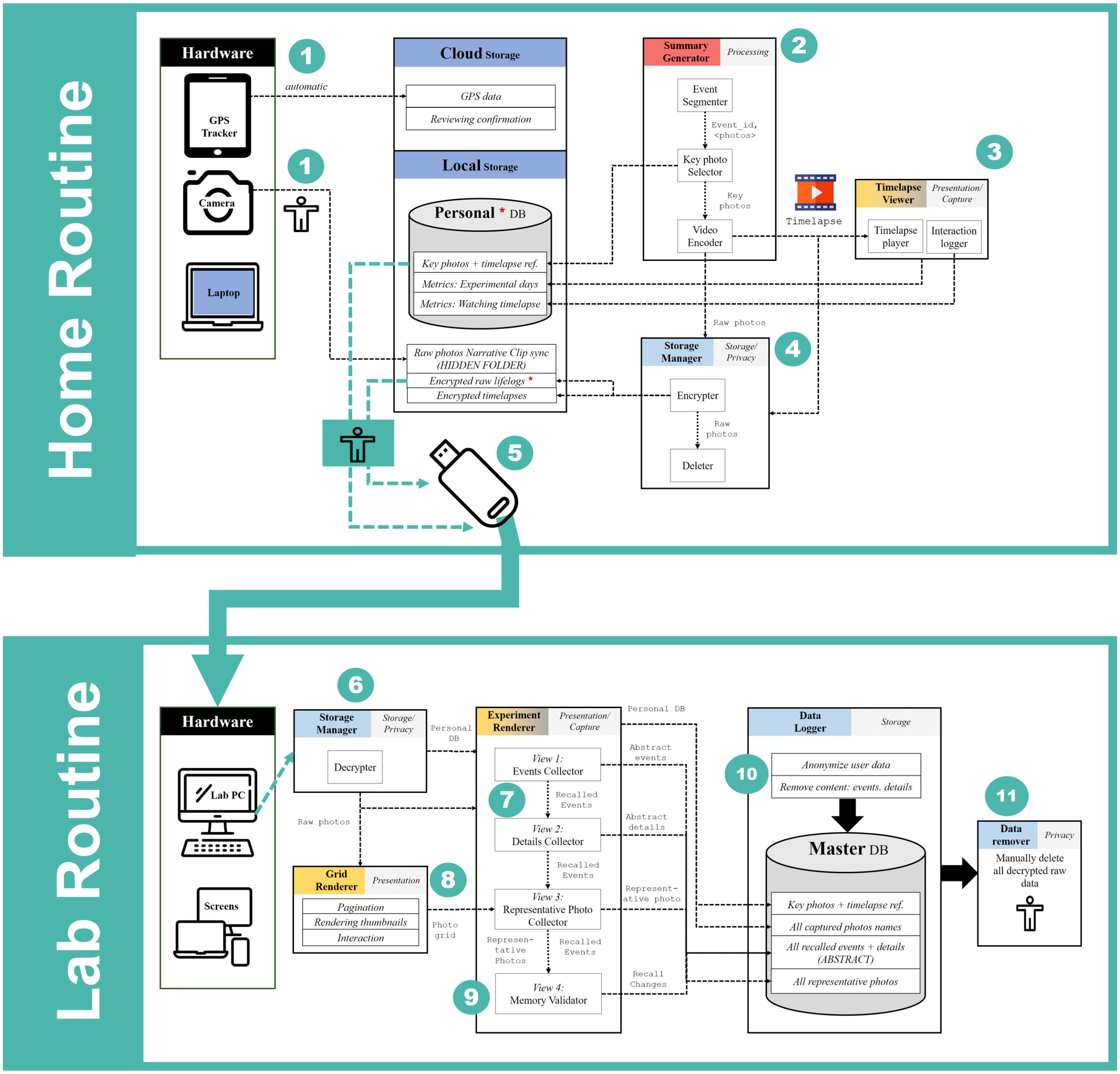

Figure 4: Overview on the of the developed prototypes.