Report of AR Memory Space

SD5511 Artificial Intelligent UX Project

GROUP 08

WENDY LU HONG | YANG SANQI | YANG SIQI | ZHANG MINGYUAN

Content

Abstract

The pandemic has caused sedentary lifestyles and psychological issues like loneliness for most people, especially those who live alone in China. So this study focuses on people who live alone and aims to find a solution to encourage them to go outside and provide a private and safe way to socialise with acquaintances. Questionnaires were conducted with different age groups for user segmentation. A semi-structured interview was implemented to identify users' behaviours and experiences of travelling and recording to find their needs. Results reveal that young people aged 18–30 are willing to go out, record, and share their travel memories. However, they are most concerned about their privacy issues and find it inconvenient to record and share their travel memories with current APPs. So this study proposed an intelligent system of travel experience visualisation and memory reappearance, which provides users access to quickly recording impressive, exciting things in travel or even little surprises encountered in daily life, and also creates a unique AR memory space based on their experience, which could be a channel for them to recall memory.

Keywords

Solo living, prolonged loneliness, travel, life recording, memory reconstruction, augmented reality, artificial intelligence.

Introduction

92 million people are living alone in China and nearly ¼ households have only one family member. Between 1999 and 2020, the one-person households jumped from 6% to 25%. At the same time, the pandemic also causes physical as well as psychological problems for these people who live alone. The problem of loneliness has been exacerbated and evolved into a social recession. People prefer staying at home alone rather than going out. They are feeling emotionally exhausted, anxious, and depressed during the long-term lockdown and lose the ability to actively socialise. Even before the pandemic, health officials expressed concern over how much time people spend sitting – a number that increased by four hours per day in 2020. As consumers adjust to a more sedentary lifestyle, three in five report new aches and pains about working from home. 2021 WHO statistics reveal that one in four adults, and four out of five adolescents, do not get enough physical activity.

In addition, with the rapid change in the information environment, the attraction to social media has gradually declined. Social media like Wechat was originally a social field for acquaintances. However, the addition of strangers gradually blurs the social boundary and weakens the "acquaintance" attribute of many social media. Users face information overload, social overload, privacy security, and other problems. From enthusiastic participation of users to quiet exit, mental burnout on social media is increasingly apparent.

Although privacy permission settings of most social media have protected user information to a certain extent, there is still an inevitable risk of data leakage. On the one hand, social media users worry about information security on their social platforms, but on the other hand, they actively disclose their personal information. Barnes calls this contradiction the "privacy paradox". (Barnes, 2006) In addition, users do not have a good grasp of the blurry boundaries of privacy security. They often unconsciously disclose personal information to the outside world, such as setting a real avatar and displaying the location, which is beyond the scope of privacy security. Moreover, people cannot completely block the "snooping" from strangers, which reduces users' desire to show themselves. Sharing life on social media gradually lost its original charm.

Thus, in this report, we focus on people who live alone (especially in Tier 1 and Tier 2 cities) and try to find a way to encourage them to go outside. In particular, we consider their passive social attitude and want to provide a comfortable way for them to socialize and at the same time, protect their privacy. This report is organized as follows. In the next section, a review of the literature on technologies we may utilize is provided. In section three, we conduct questionnaires and interviews to explore their pain points and needs related to travelling and socializing. Finally, in the last two sections, we report what we have designed and the limitations of the design and provide insights for future study.

Literature Review

As we are going to design an intelligent system for users to easily record and revisit their memory, we will make a literature review in this part to confirm the technical feasibility of our design idea.

Facial Emotion Recognition

1. Reasons for choosing Facial Emotion Recognition

Emotions are intentional. They "imply and involve relationships with a particular object" (Frijda, 1994). We are satisfied with someone, amazed at something, and enthusiastic about a few activities throughout our tour experiences, whether or not on an organized excursion or day-by-day life. We comprehend and recollect a selected object, scenery, or occasion with exceptional emotions. We record emotions as a vital input item in our system.

2. What is Facial Emotion Recognition and how it works

Facial Emotion Recognition is one of the facial recognition technologies that use a specific programme to examine and process the expressions on a human’s face. Artificial intelligence could study and detect different facial expressions like a human brain.

Emotion recognition is essentially the mapping relationship between characteristic data of emotions and intrinsic emotional states(何志鹏， 李自娜，梁艳， 邱丽娜, 2020). It has a general workflow, which can be illustrated in Fig.1(Canal *et al.*, 2022):

1. Input step: Input images or video sequences.
2. Image preparation step: improve image quality, reduce noise, reduce dimensionality in spatial resolution, or simplify the colour information into grayscale.
3. The feature extraction part: extract relevant features from the input image.
4. The classification part involves putting previously obtained features into the classification model, which could return a similarity or dissimilarity score for different facial emotions. The highest score expression is the recognition result for an input image.

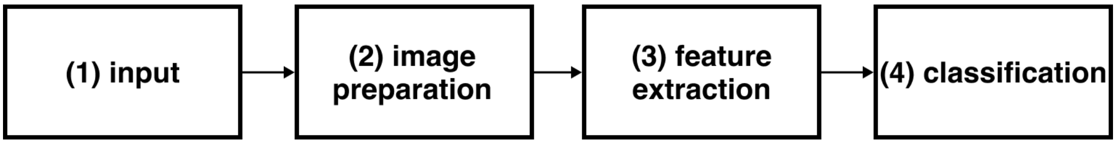

Figure 1. Facial emotion recognition workflow (own collection)

3. Technology in facial emotion recognition

Classification is the most critical stage in the recognition process. Some classical approaches have done an excellent job and have been widely used until now. These classical algorithms are Support Vector Machine (SVM) (Wang, 2005), Dynamic Bayesian Networks (DBNs) (Silander & Myllymaki, 2012), and Fuzzy Logic (Zadeh, 1965). Besides those classical approaches, neural network-based algorithms have been successfully used in various areas, including convolutional neural networks, multi-layer perceptrons, and neural networks. Convolutional Neural Networks (CNN) (Albawi, S. et al., 2018) are one of the main algorithms for solving problems like image recognition, classification, object detection, and face localization. This method sees the input image as a pixel matrix and then puts this matrix into a series of convolution filters, pooling and functions. It is possible to classify images with considerable accuracy using a trained network. CNN could be used in step 2 (image preparation) for face detection and step 4 (classification) for emotion classification.

Feature extraction is another crucial stage. It aims to select only the most essential and descriptive pieces of information, removing what is irrelevant to the given problem, as lots of data in each sample leads to the high computational complexity of training the classification models. Some algorithms work for feature extraction, such as the Active Shape Model (ASM) (Cootes et al., 1995), Local Binary Patterns (LBP) (Ojala, Pietikäinen and Harwood, 1996), Histogram of Oriented Gradients (HOG) (Bouwmans et al., 2018), Facial Action Coding System (FACS) (Ekman & Friesen, 2019), etc.

4. Challenges and solutions in our scenario

During the COVID-19 pandemic, every pedestrian in the street wears a face mask, which may become an obstacle to recognising emotions while travelling. It is a challenging task in the research area of computer vision, and many researchers work on it. As illustrated in Fig.2 (Yang, Jianming and Hattori, 2021), a two-stage attention model in the literature review has been proposed to improve facial emotion recognition accuracy with face masks. In stage 1, a masked/unmasked binary deep classifier has been trained, generating attention heatmaps to distinguish the masked facial parts from the unobscured region. Furthermore, in stage 2, a Face-Emotion-Recognition classifier has been introduced, which could capture the essential area that indicates facial expression.

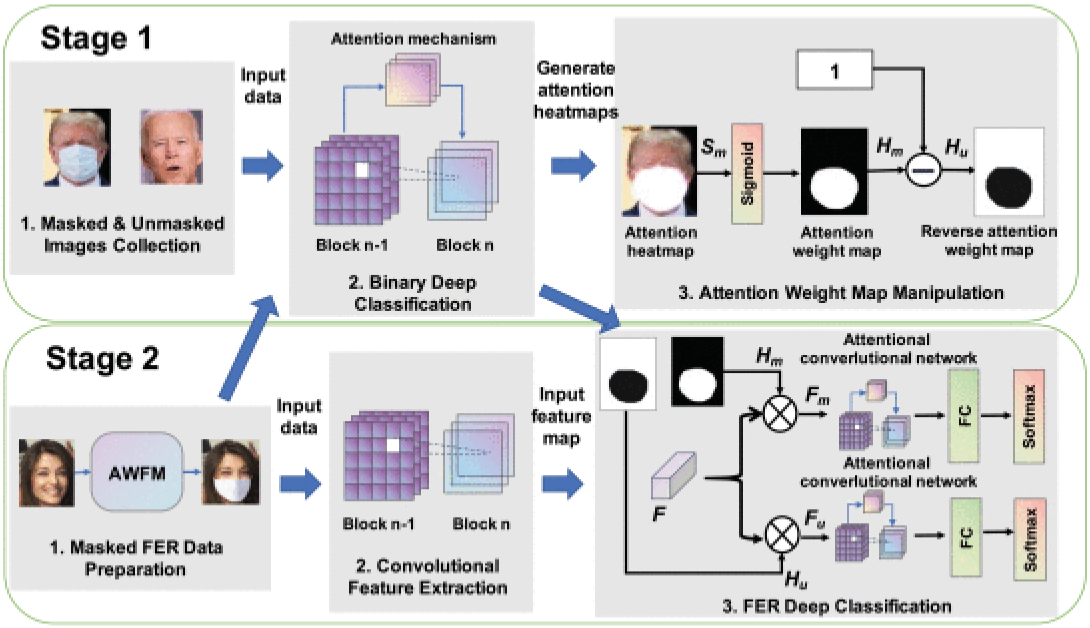

Figure 2. Two-stage attention framework (Yang, Jianming and Hattori, 2021)

Image-to-Text

1. Reasons for choosing Image-to-Text

As we want to deconstruct the memory of users' travel experiences and reconstruct those memory elements by using AI, we need to turn the image into a descriptive text and then use the text to generate a new idea. This section describes in detail the technical research of the image-to-text conversion first.

2. Technology in Image-to-Text

The technique for machines to automatically generate a textual description of images is called image captioning. It focuses on mapping the semantics of visuals to text, which involves different fields such as Natural Language Processing (NLP), Computer Vision (CV), and Artificial Intelligence. So it needs not only object recognition in the image but also an understanding of the semantic content of images, including the scene location, object properties, and their interactions.

There are some traditional methods for image captioning. In these methods, the computer imitates the human visual system to tell what it is watching; then, through research on human visual cognition abilities, the computer describes what they see in a simple natural language. Related methods are Retrieval-Based Image Captioning and Template-Based Image Captioning. Besides, deep learning methods have made significant progress nowadays. It can directly map images to text using a robust dataset and generate descriptive sentences.

Some image captioning samples are shown in Fig.3 (Ming et al., 2022).

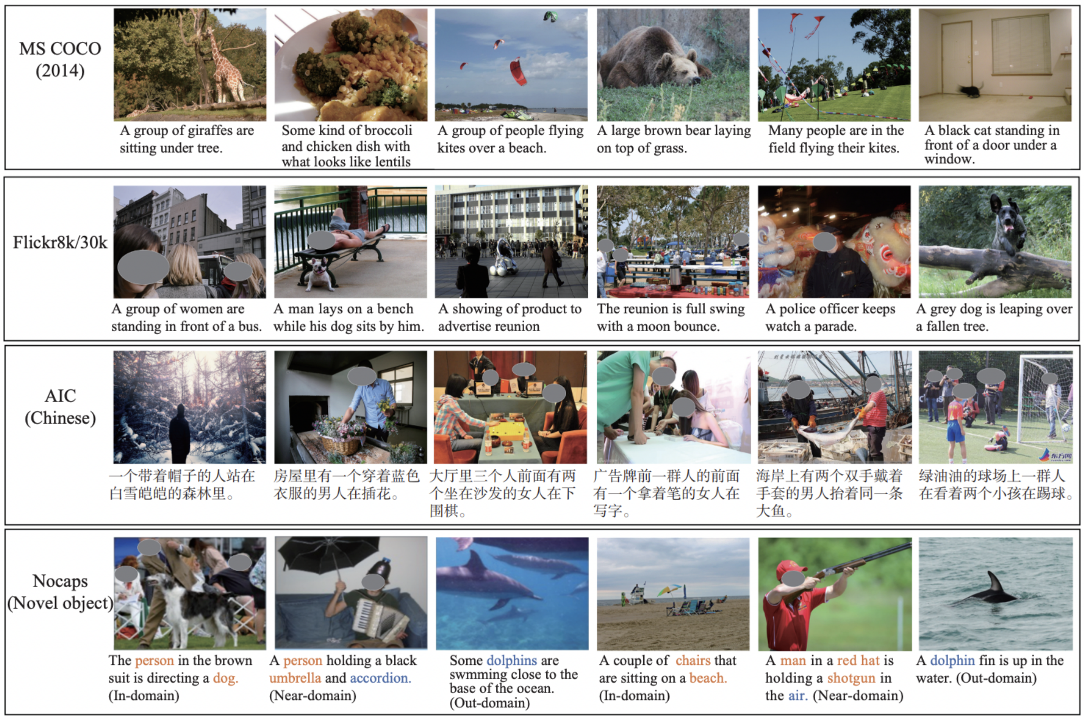

Figure 3. Examples from typical image captioning datasets (Ming *et al.*, 2022)

Text-to-Image

1. Reasons for choosing Text-to-Image

As for the stage of memory reconstruction, we consider using two kinds of techniques. One will map concrete objects as abstract objects, which can be achieved by creating a mapping database. Another is AI-generation art, which can automatically paint an excellent graph. This section will research some popular text-to-image diffusion models in AI painting.

2. Technology in Text-to-Image

As shown in Fig.4, DALLE 2 (OpenAI, 2022) is a prevalent AI painting system created by OpenAI. It can create original, realistic images and art from a text description that combines concepts, attributes, and styles.

Figure 4. DALL·E 2 (OpenAI, 2022)

Another AI painting model illustrated in Fig.5 is Imagen(*Imagen: Text-to-Image Diffusion Models*, no date a), published by Google, which claims to surpass DALLE 2. It possesses an unprecedented degree of photorealism and a deep level of language understanding.

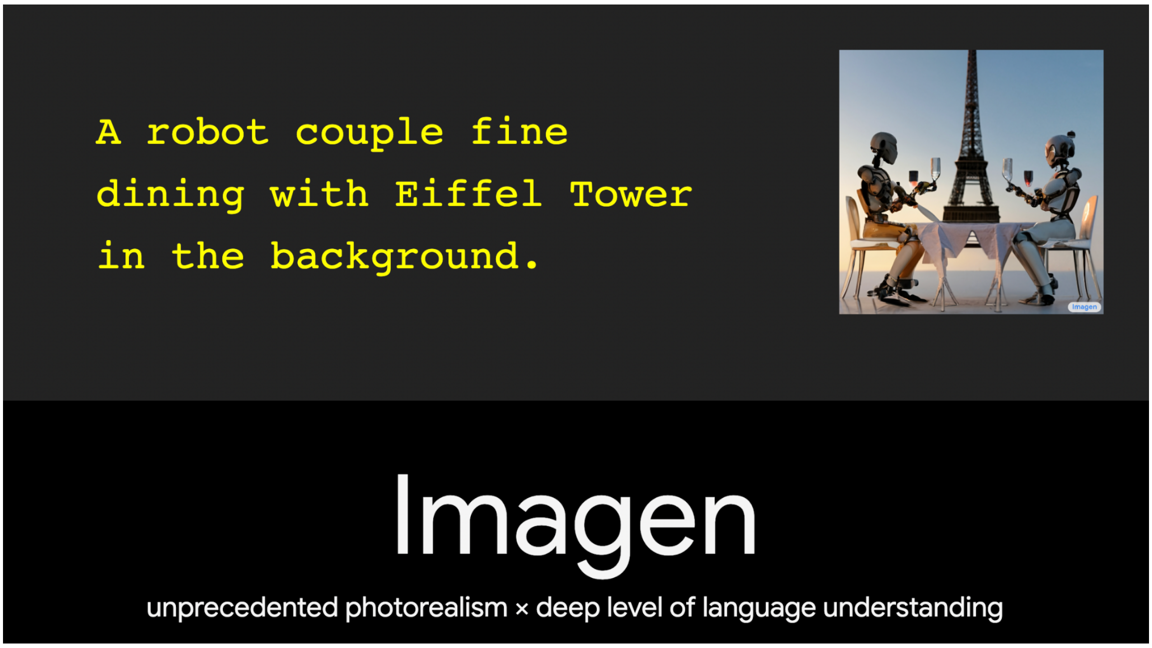

Figure 5. Imagen (*Imagen: Text-to-Image Diffusion Models*, no date b)

Method

The questionnaire

We conducted a questionnaire for quantitative research. A structured questionnaire was developed to collect data for this research. The questionnaire is composed of three sections. All the questions and their sources are reported in Appendix A.

The first section included 4 questions on the socio-demographic characteristics of the subjects(gender, age, education, profession). The second section included 7 questions. In this section, we tend to get to know their willingness to go out and record their travel as well as their previous experience of going out. In Particular, Q10 investigated what kinds of factors had caused their bad experience. The third section included 4 questions. A five-point Likert-type scale was utilized to evaluate their interest (from “not at all” to “extremely”) in augmented reality technology and digital collection.

1. Procedure and sample

The developed questionnaire was pre-tested with a sample of 20 people to check for the clarity of questions. As a result, small changes in the wording of some sentences were done to improve their clarity.

The questionnaire was developed by Wenjuanxing and administered through social media channels (e.g., Wechat). The survey was distributed to people living in Hongkong in November 2022. The sample consisted of N = 159 participants (36.48% Male, 57.86% Female)In our survey, all the questions were mandatory: without having provided an answer to one question, respondents were disallowed to continue to the next questions.

2. Result and discussion

From our analysis, we found that all ages suffer from psychological problems under COVID-19. People aged 23 to 30 feel more anxious, stressed, and lonely. Most of the respondents have the intention to go outside. However, the confusion about where to go, tiredness of making plans, and the complexity of transportation are the main reasons that restrict their willingness to go out. At the same time, they tend to record their travel through multimedia including photo, video, and social media, among which photo is the most frequently used recording method. And people aged 18 to 30 have the most interest in AR technology and the questionnaire reveals that they are more likely to accept AR technology.

Through the questionnaire, we did user segmentation and defined our target users as young people aged 18 to 30. And we have a basic understanding of their situation regarding travelling and recording, which guided us to deeper investigate their needs through qualitative research.

Interview

Based on the analysis of previous questionnaires, we employed a qualitative approach and semi-structured interviews were conducted to identify common user behaviours and experiences of traveling and recording, their needs and frustrations, and the process and emotions that they experience during their trips.

1. Interviewees and Recruitment

Our target subjects are young people aged between 18 and 30 years old. Understanding their perceptions and behaviour of them is highly relevant as it can guide us to better meet their true needs. The participants were recruited via Wechat and offline in various cities across China. Fourteen individuals were interviewed face-to-face offline or through Tencent Meeting online. The interviews were audio-recorded upon receiving consent from the participants. Data collection stopped when new interviews did not bring new insights to the study. All participants gave their informed consent for inclusion before they took part in the study.

2. Interview Guide

The interviews followed an interview guide (AppendixB) that was divided into 3 sections: (1)general behaviours and experiences of travelling and recording, (2)Needs and frustrations during travelling and recording, and (3)the feeling and perception of our probable solutions. Moreover, the interview guide was pretested to eliminate a potential lack of clarity.

3. Result

Most interviewees don’t want to share photos on public social media as they don’t want to be observed by other people they are not familiar with. Most interviewees are more willing to share their experiences to close friends and gain feedback. Some interviewees want to make their experience public, but don't want to reveal too much information.

General result and discussion

1. Persona

By conducting questionnaires and interviews, we have a relatively clear perception of our target user and draw a persona accordingly. (See Figure 6)

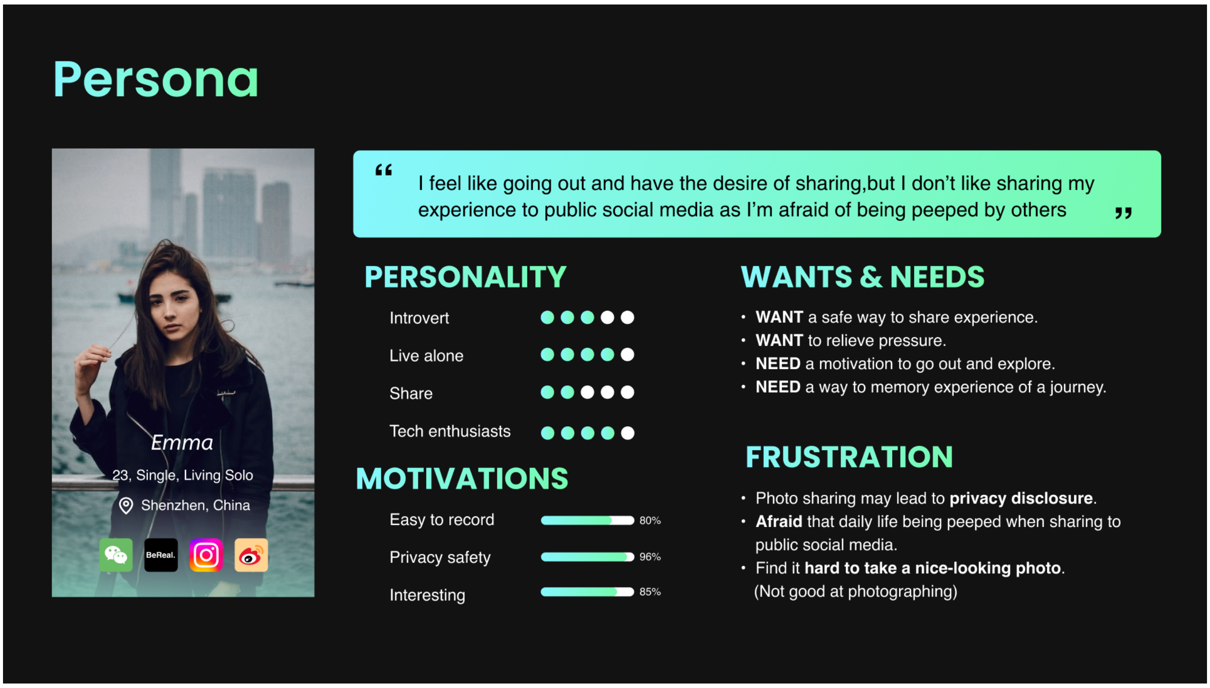

Figure 6. Persona (own collection)

2. User journey map

In order to gain a clear understanding of the user behaviour, their thinking and emotion during travelling and recording, a user journey map was conducted for better visualization and to help us figure out users’ pain points and our design opportunity. (See Figure 7)

The user experience is divided into two stages: during travel and after travel(which includes sharing and memory). In the first stage, users leave home and find their friends to have recreation, after which the user goes back home. During this stage, the user may find it hard to shoot a good video due to poor shooting skills. And sometimes, they may forget to take photos or find it inconvenient. Also, it’s hard to choose good-looking photos without editing and they may feel frustrated when trying to post them. So the opportunities are.

In the second stage, users tend to share photos on WeChat moments and it will waste a long time to pick up good photos that could be posted in moments. So the probable solution is to share an AI-generated key to moments, anyone who is close to users could open their experience memory. Users also want to record their memory in albums but don’t have enough time to record in detail sometimes. So it’s an opportunity to generate memory automatically and users could review their memory in AR space.

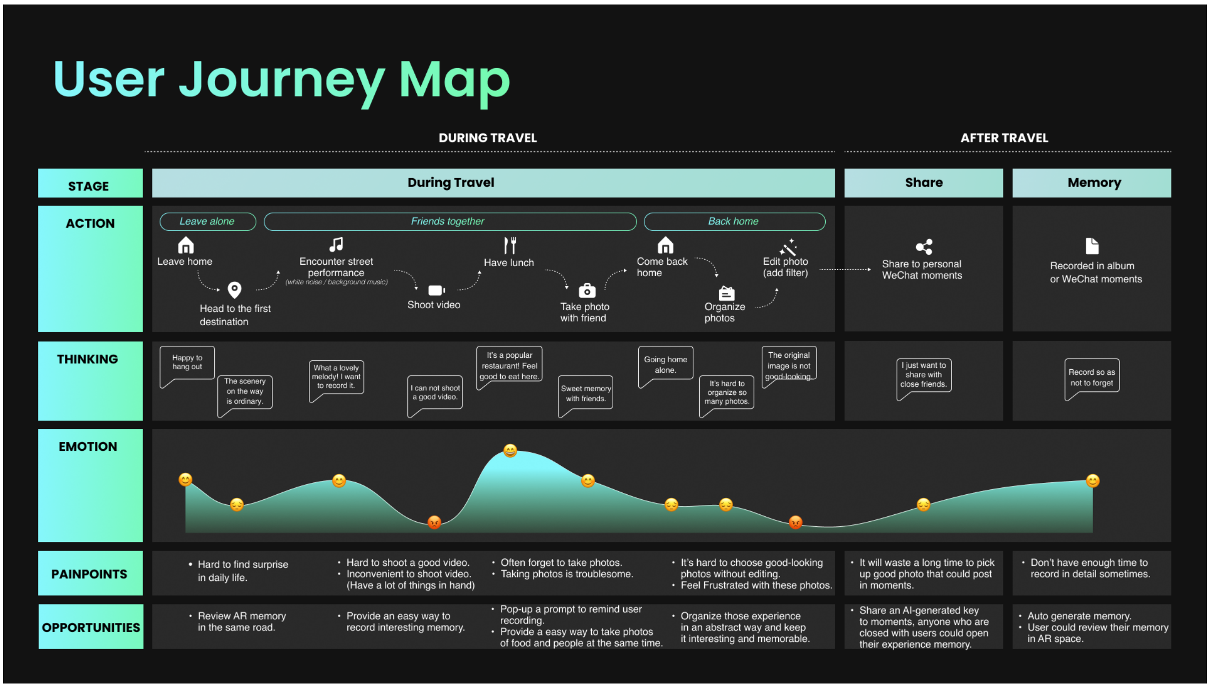

Figure 7. User journey map (Own collection)

Design Implementation

Intelligent System Design-information input and output

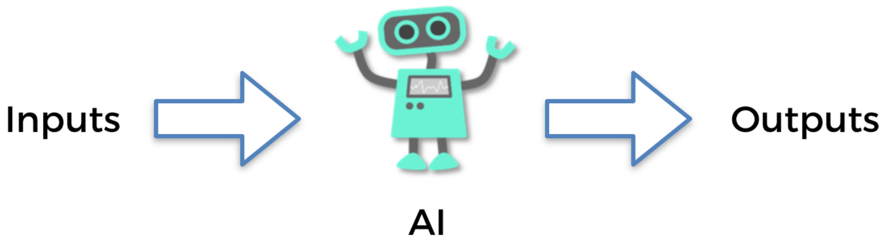

Figure 8. The input and output system (*5 key reinforcement learning principles explained by AI expert, Hadelin de Ponteves*, 2019)

Fig. 8 shows reinforcement learning input and output. The input is named the state or the input state. The result's the action performed by the AI. Fig 2 presents the connection between input data and output data.

Fig 9 shows that the full system will generate a travel map, ar memory space, and a memory box (the memory box is the entrance for users to enter memory space).

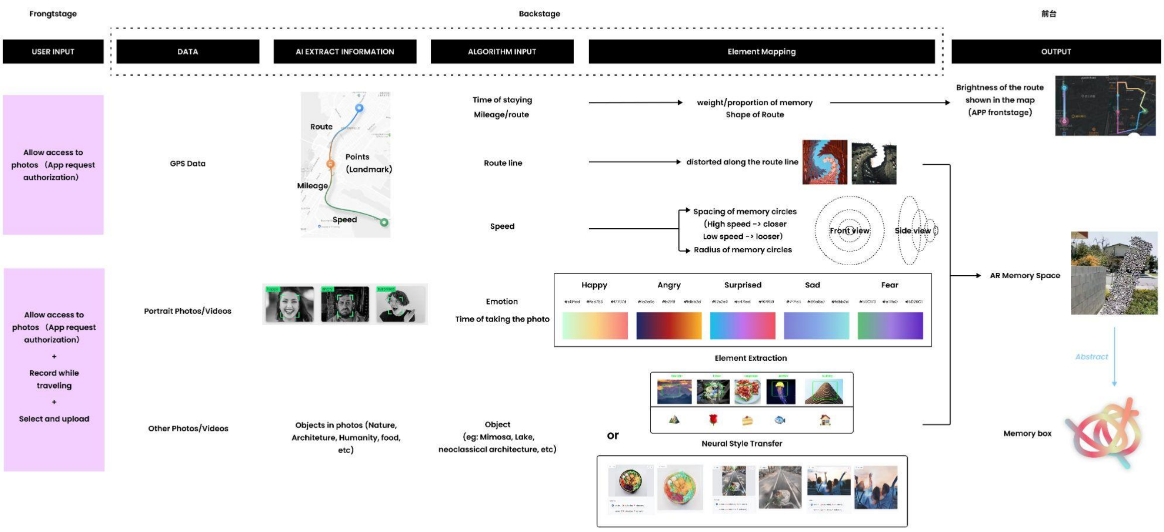

Figure 9. Information put and output (Own collection)

1. Data-Access Layer (DAL) (Fig 3)

A Data-Access Layer (Fig 10) has simple and simplified access to data kept in persistent storage. it's a layer between the Business Logic Layer (BLL) and therefore the storage layer.

Whenever an application (BLL) needs a particular data question or update, the request is shipped to the data access layer, which takes the appropriate action to confirm that the requested data or update is executed. This ends up in a level of abstraction wherever the business layer can solely understand which perform or methodology to call; it does not need to know the existing architecture or details of the entire database. (Data-Access Layer, 2022) (Fig 11)

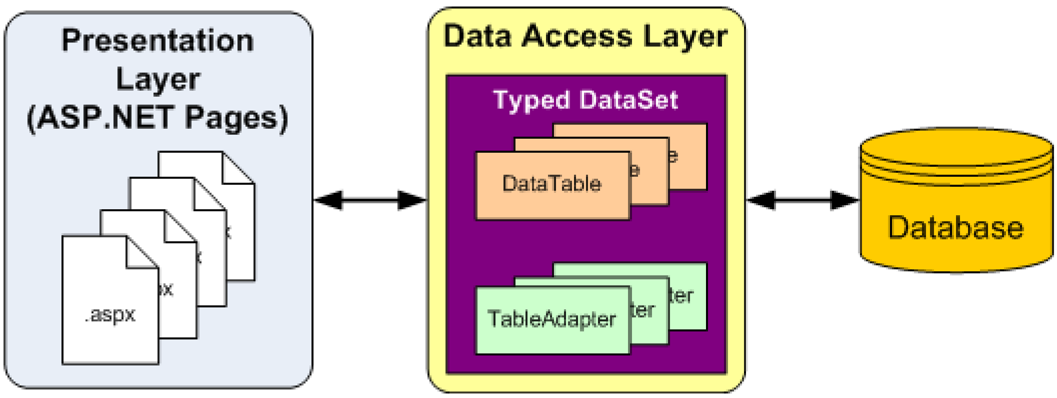

Figure 10. All Data Access Code is Relegated to the DAL (Rick-Anderson, no date)

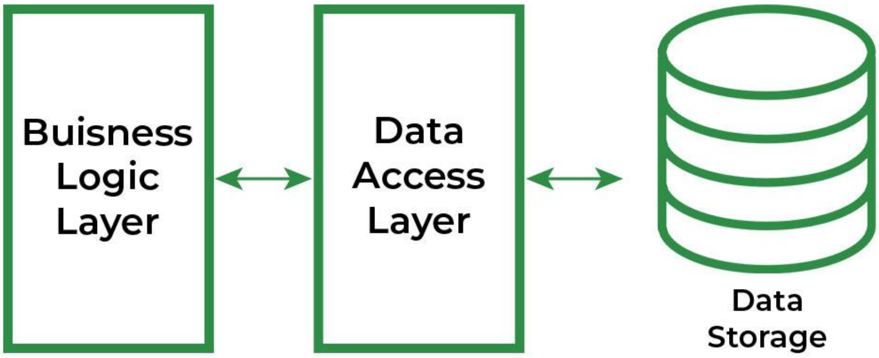

Figure 11. Data Access Layer (*Data-Access Layer*, 2022)

2. Generation of GPS Data - track points/route line/speed

By analyzing GPS data, the system can learn the user's information, like track points, route line, speed of this journey, etc.

2.1. Map View-Route line & Track Points

The route line and track points determine part of the travel map. In Figure 12, it is obvious to see that, depending on the mood of the person walking the route, the map will show colour changes for different moods. Track points mainly show the duration of the user's stay and mood changes. When the system determines that the user has stayed in place for more than 45 minutes, the system will prompt for a manual recall. Also, when the user's mood changes, track points record the points at the junction of the two moods and record them in the colour of the previous mood to show the change in mood.

Figure 12. Route line & Track points (Own collection)

2.2. AR View - Path visualization

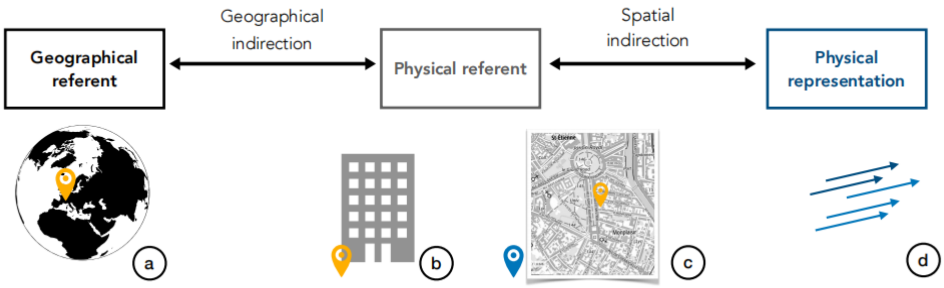

Figure 13. Schema representing the geographical referent (a) corresponding to a geographical position, the physical referent (b-c), and the physical representation (d). The geographical indirection represents the distance between the geographical referent and the physical referent. They can be the same (a) or a scaled referent such as a map can be used (c). The spatial indirection represents the distance between the physical referent and the physical representation. (Lobo and Christophe, 2020)

As in Fig 13, the route information is linked to the physical space and the series of references using geo-relations in the AR view. So we communicate the route information through a directionally guided path. The directionally guided route is superimposed on the realistic scene and displayed in the real world on the mobile device, enhancing the recall of the route at the time and the path of exploration of the generated elements. (Fig 14).

Figure 14. AR route in AR space (Own collection)

2.3. Generating the size of the AR space by the speed of the journey

AR is one part of the general area of mixed reality. Both virtual environments (or virtual reality) and augmented virtuality, in which natural objects are added to virtual ones, replace the surrounding environment with a virtual one. (2010) (Van Krevelen and Poelman).In order to differentiate the AR space according to the different trips, we determine the AR space's size by the trip's speed. Fig 15 shows the effect of speed on space, specifically, viewed from both the front and side views. Application to the subsequent production of the AR space.

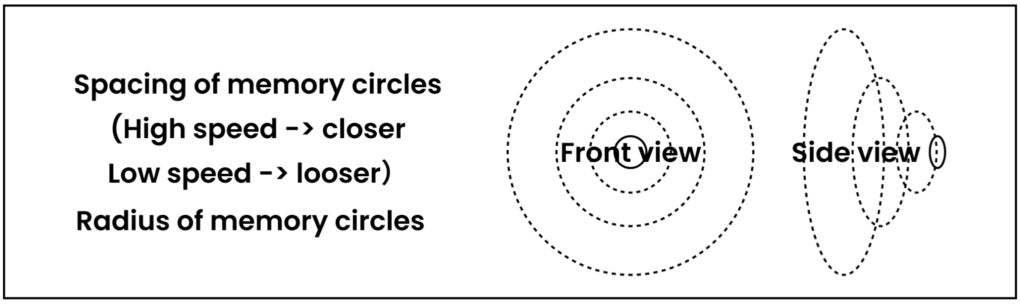

Figure 15. Memory space radius of AR space (Own collection)

3. Front and back camera user input

3.1. Front camera-Expressing emotional changes in facial expression Recognition in visual form

In Fig. 16, emotion recognition allows different emotional states to be analysed by facial expressions. AI performs emotion recognition on portrait photos. From the literature review, multimodal face emotion recognition proves that we can extract emotions from photos and videos efficiently.

Emotion	Facial Expression
Anger	Lowered and burrowed eyebrows Intense gaze Raised chin
Joy	Raised corners of mouth into a smile
Surprise	Dropped jaw Raised brows Wide eyes
Fear	Open mouth Wide eyes Furrowed brows
Sadness	Furrowed brows Lip corner depressor
Anxiety	Biting of the lips

Figure 16. Emotions along with their common corresponding facial expressions (*Emotion Recognition: Introduction to Emotion Reading Technology*, 2021)

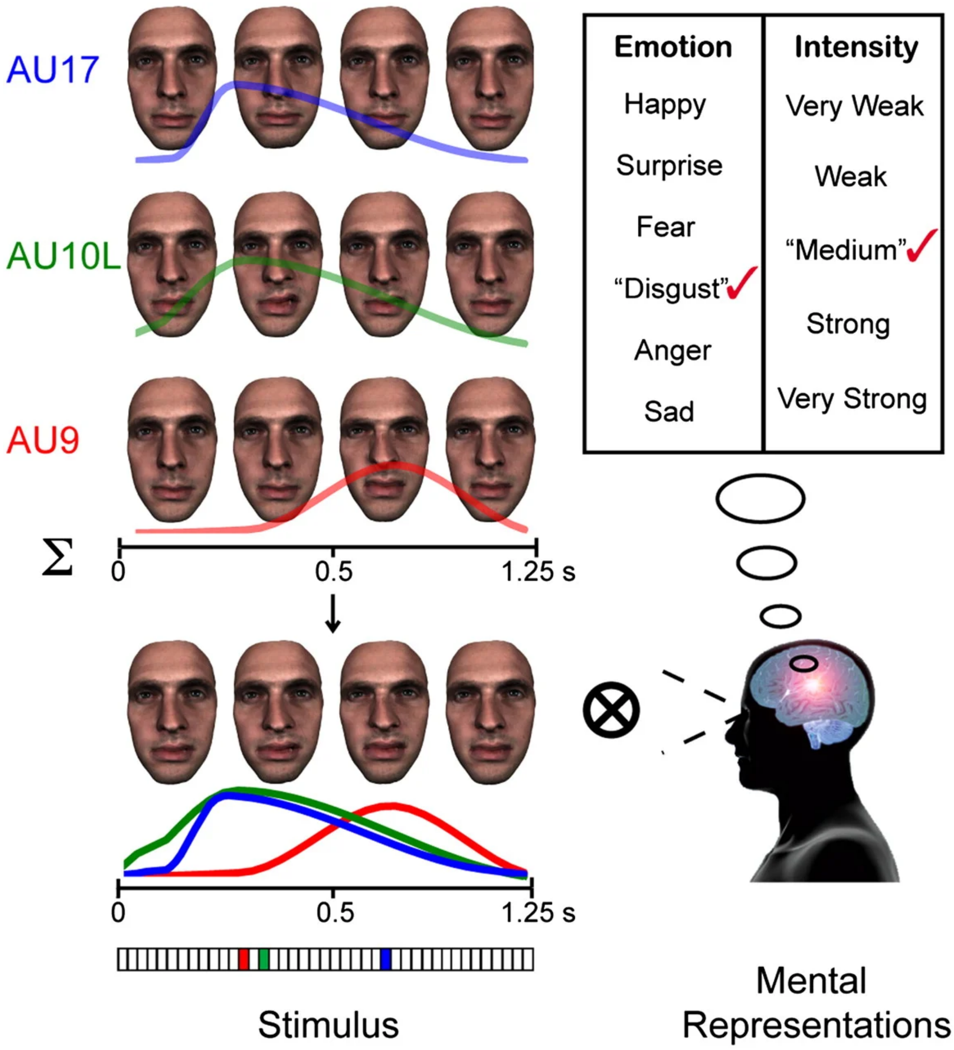

Figure 17. Random generative grammar of facial movements and the perceptual categorization of emotions. (*Website*, no date)(*Website*, no date; Jack *et al.*, 2012)(*Website*, no date)

Through AI, facial expressions are generated to classify the perception of Emotion and intensity (fig 17). We studied people's emotional associations with colour. The study of emotional colour (fig.18) identifies our norms for colour expression (fig. 19).

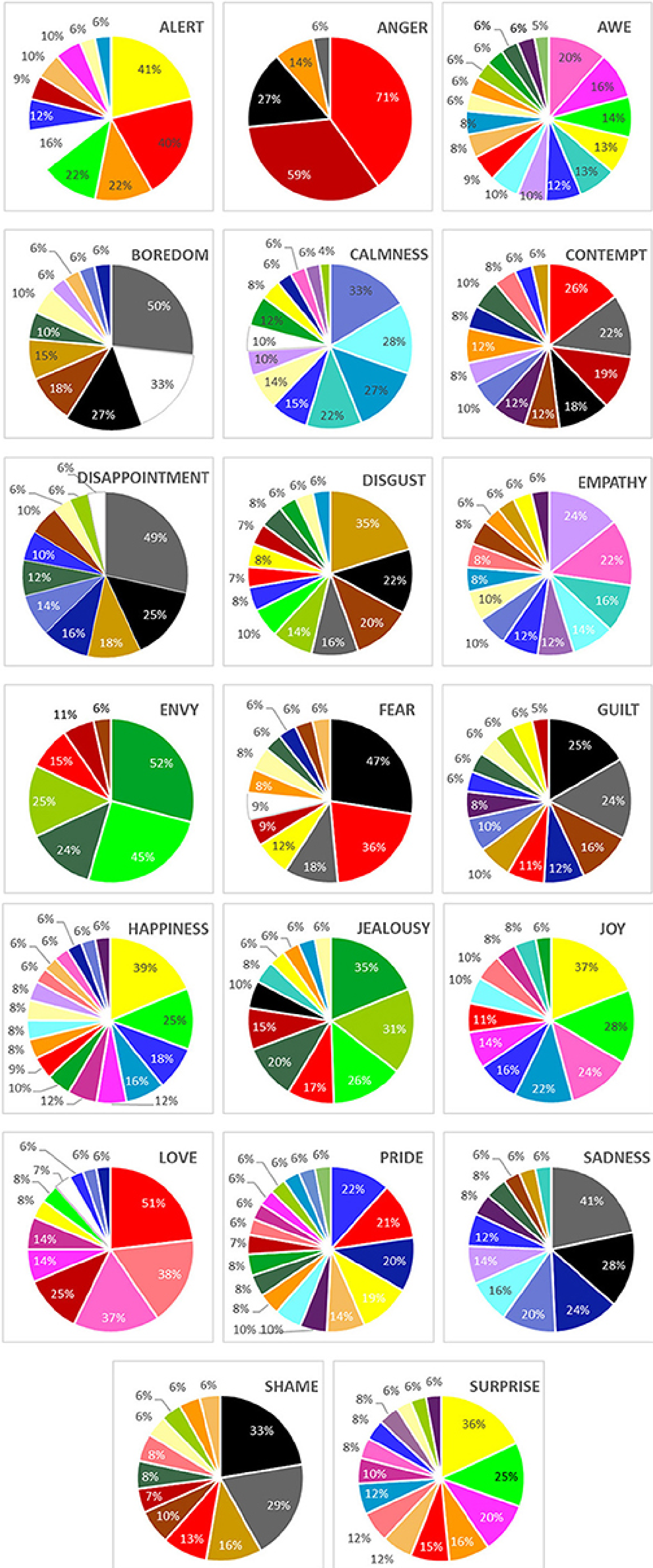
ALERT
ANGER
AWE
BOREDOM
CALMNESS
CONTEMPT
DISAPPOINTMENT
DISGUST
EMPATHY
ENVY
FEAR
GUILT
HAPPINESS
JEALOUSY
JOY
LOVE
PRIDE
SADNESS
SHAME
SURPRISE

Figure 18. Frequency of selection of emotions for each colour ((Fugate and Franco, 2019))

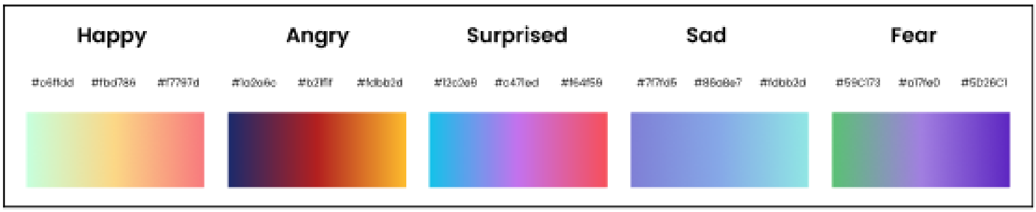

Figure 19. Mood colour specification (Own collection)

3.2. Back camera: AI recognises user-inputted photos and videos.

The AI performs emotion recognition on portrait photos, and its output corresponds to the colour change of the travel map route and the colour change of the memory space element. For other photos, the system will first convert them into descriptive text, such as "strawberry crispbread on the blue plate at home." The system will recognise the primary object "crispbread" as the main keyword. The main keyword is used to map an abstract element, and the entire descriptive text is the input of the AI painting artwork.

4. Generation of “AR memory space”&“Memory box”

4.1. Element generation in AR space

The whole generation process includes two main parts, automatic system generation and user upload section. In fig 20, "Element Generation", with GPS, we can get the route of the road and show the track points and speed in the app and the radius and route in the AR space. Photos uploaded by users from the front and back camera extract the user's mood, corresponding to the relevant mood colors. At the same time, AI recognition of text messages then generates elements from the database and finally generates AI painting.

4.2. AR space compressed into a “Memory box”

In fig 21, The AR memory box is a thumbnail sketch of the "AR memory space", using the AR drawing and mood (optional), the road map as a background, and extracting the more weighted elements from the AR space. Fig 14 also demonstrates some of the combinations that can occur.

Figure 20. Generation of AR space and memory box (Own collection)

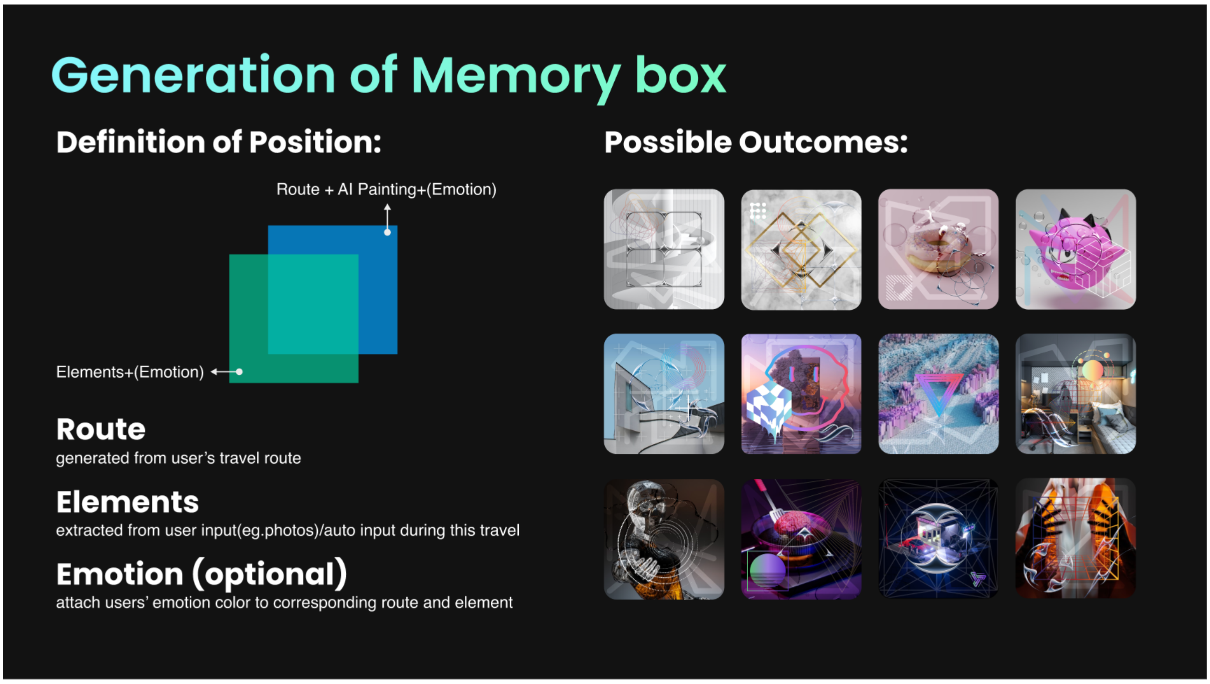

Figure 21. Memory box (Own collection)

5. Data input and output

Fig.22 is our data flow graph, which aims to describe in detail how our data is used in our intelligent system.

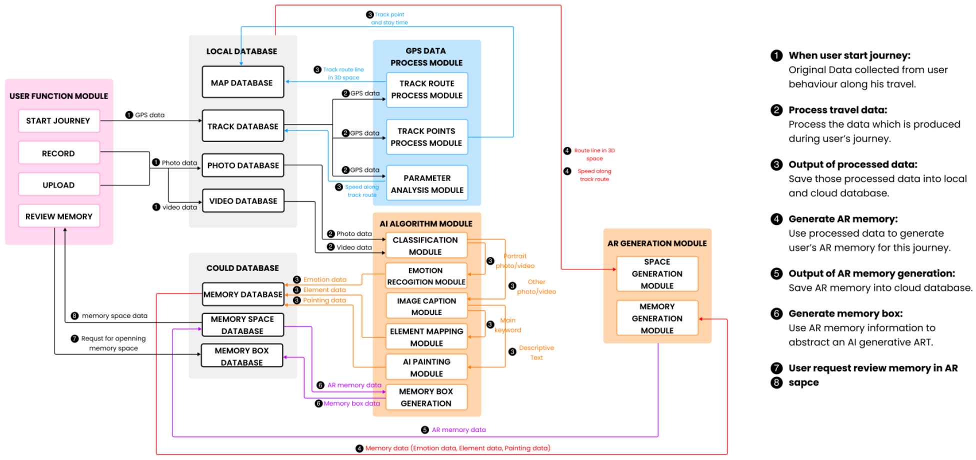

Figure 22. Data flow graph (Own collection)

5.1. Record travel data

Data input starts from the user's function; it could be roughly divided into two kinds of data. The GPS data is automatically recorded, and the photo or video data is manually recorded.

The GPS data is recorded in the Track Database (local) when the user starts travelling. In contrast, photos and videos that are recorded and uploaded by the user are stored in the Photo Database (local) and Video Database (local), respectively.

5.2. Process travel data

The AI Algorithm Module is classified into GPS Data Process Module and AI Algorithm Module by different functions.

The GPS Data Processor Module analyses GPS data and extracts critical information, including route information, track points, and speed. Then the processed data is stored in the Map Database (local) and Track Database (local).

Photos and video data are processed mainly through the AI Algorithm Module. Videos will be processed as a sequence of photos, so photo and video data will be called images later. First of all, the Classification Module will classify images. The output of portrait photos will be input into the Emotion Recognition Module, and other photos will be input into the Image Caption Module. The Image Caption Module outputs descriptive text for images and inputs this text information to the Element Mapping Module and the AI Painting Module. At last, the emotion data, element data, and painting data output by these AI Algorithm Modules are stored in the Memory Database (cloud).

5.3. Create travel memory space

The Track and Memory Database data are used to generate AR memory space. Data from the Track Database first generates the space container, and then data from the

Memory Database generates memory items at this pace. The whole memory space data will be stored in the Memory Space Database (cloud).

The Memory Box Generation Module will take the data in the Memory Space Database and fold it into a memory box (in digital space, like extracting primary information from lots of data). Afterwards, it will store the memory box in the Memory Box Database (cloud).

5.4. Review travel memory

The user's Review Memory function can read their corresponding memory space data in the cloud database. The Review Memory function requests open memory space by using its memory box, and then the Memory Space Database returns the space data to the user end.

User Experience Design

1. User flows

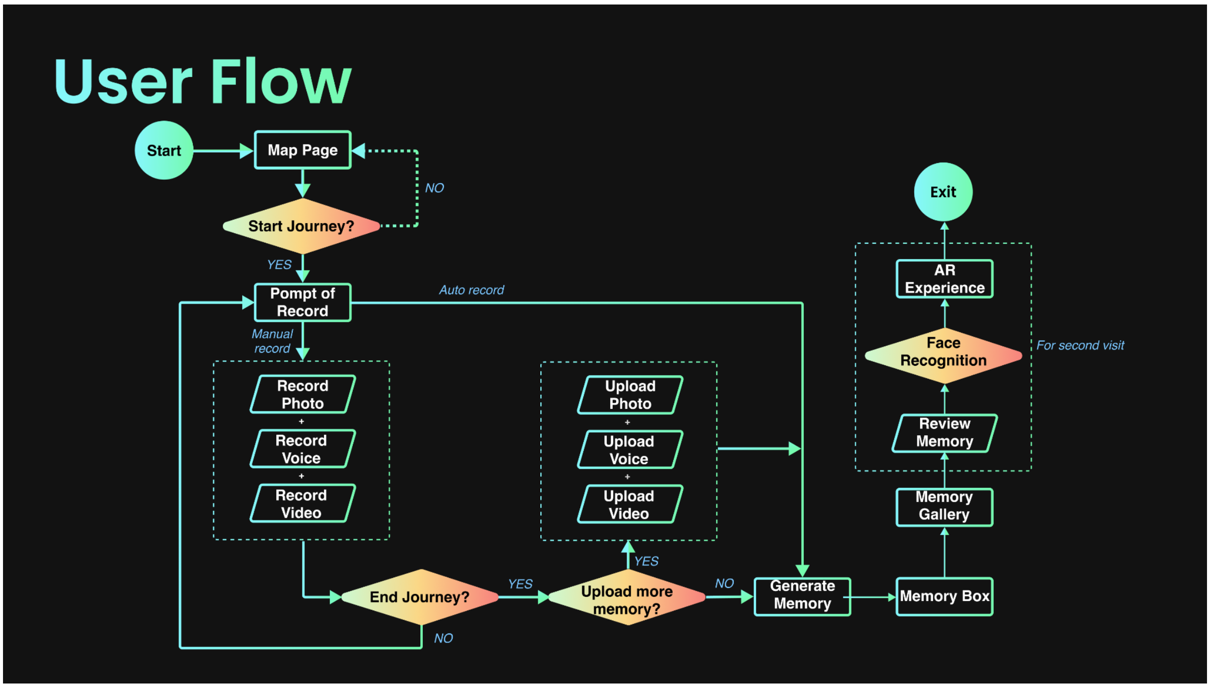

Figure 23. User flow chart (Own collection)

The goal of this user flow is to record a journey and generate memory space for a second visit. Starting from the map page, a user needs to make the first decision: to start the trip or not. A user agrees to give the system access to their location information by creating a trip. And this is used to auto-record a user's journey. A user can manually record a trip by recording photos, voice, and video. After a while, he can end the trip, and there is an additional option to upload more information. The auto-record and manual record data will come across at this point to generate a memory space. This new memory space will be placed and stored in the memory gallery as a memory box. The user can revisit his memory space anytime by using face recognition as a key to enter the memory box.

2. Storyboard

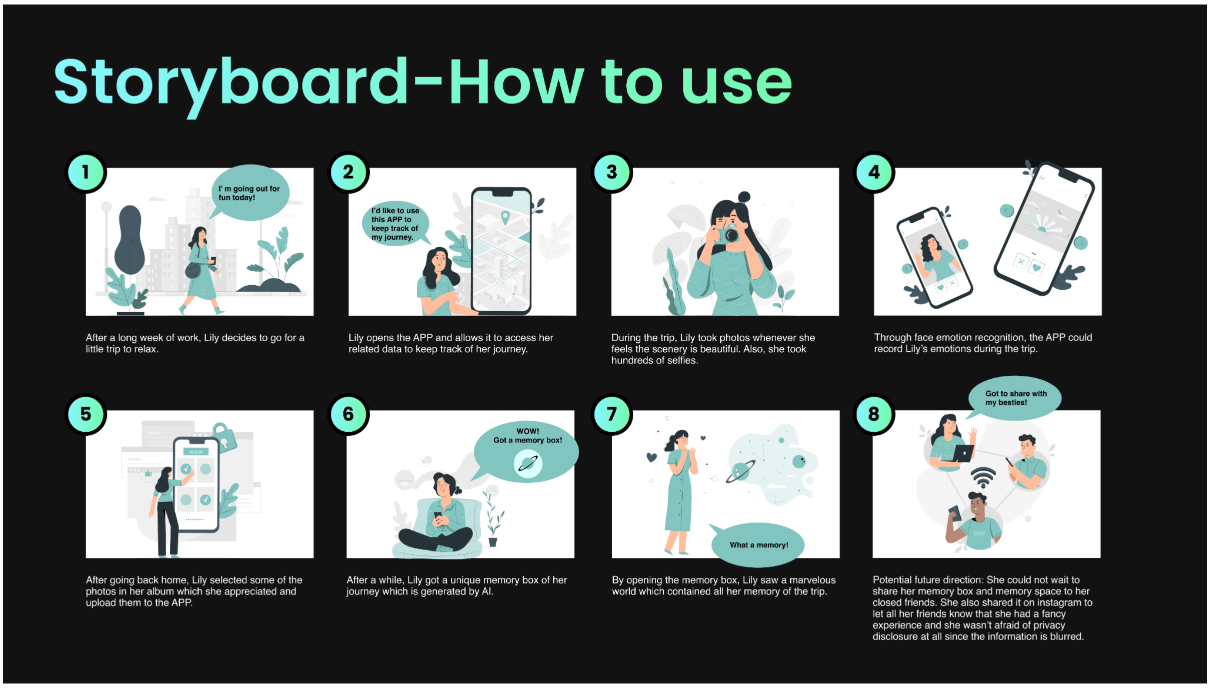

Figure 24. Storyboard (Own collection)

The purpose of this storyboard is to use Lily's experience as an example of how to use our product, thus better-aiding team communication and visualising the memory box generation process.

Final Outcome

AR space explanation

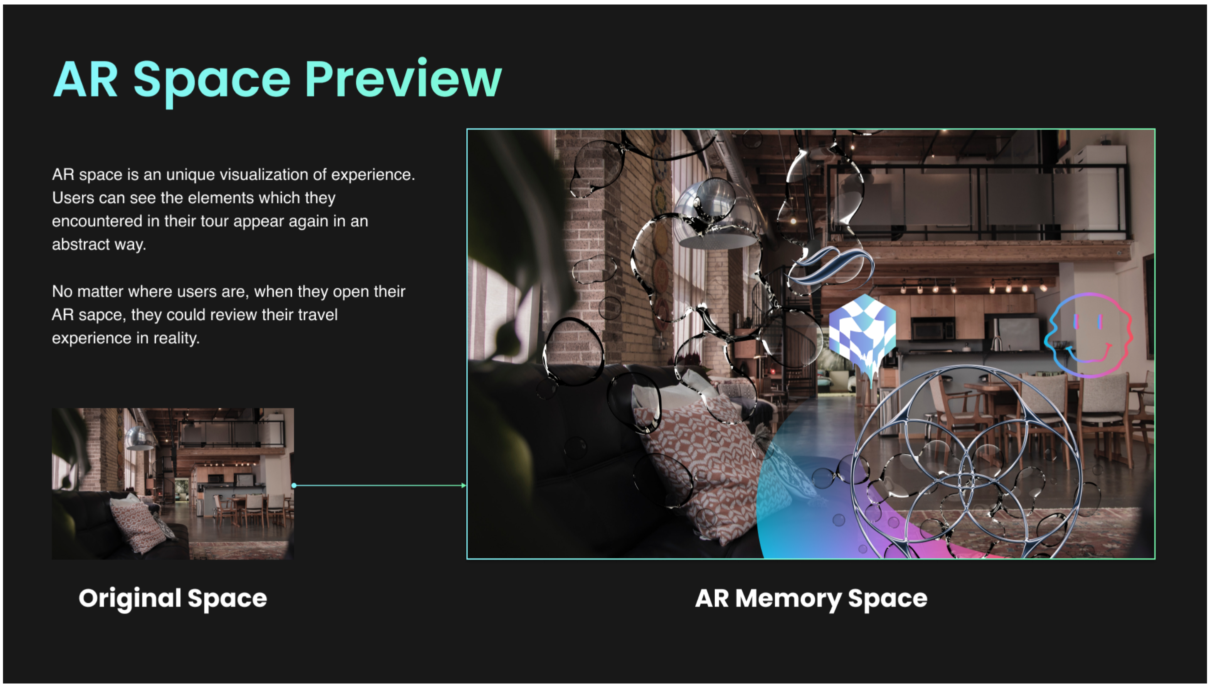

Figure 25. AR Space Preview (Own collection)

Through the input of the information introduced above, we will output an AR space.

In Fig 25. AR Space Preview (Own collection), the previously collected abstract elements are gathered together and could be viewed by turning on the AR function at any location.

This AR space is not limited to geographical location and size and can recall its own memory at any location.

Generation of memory box

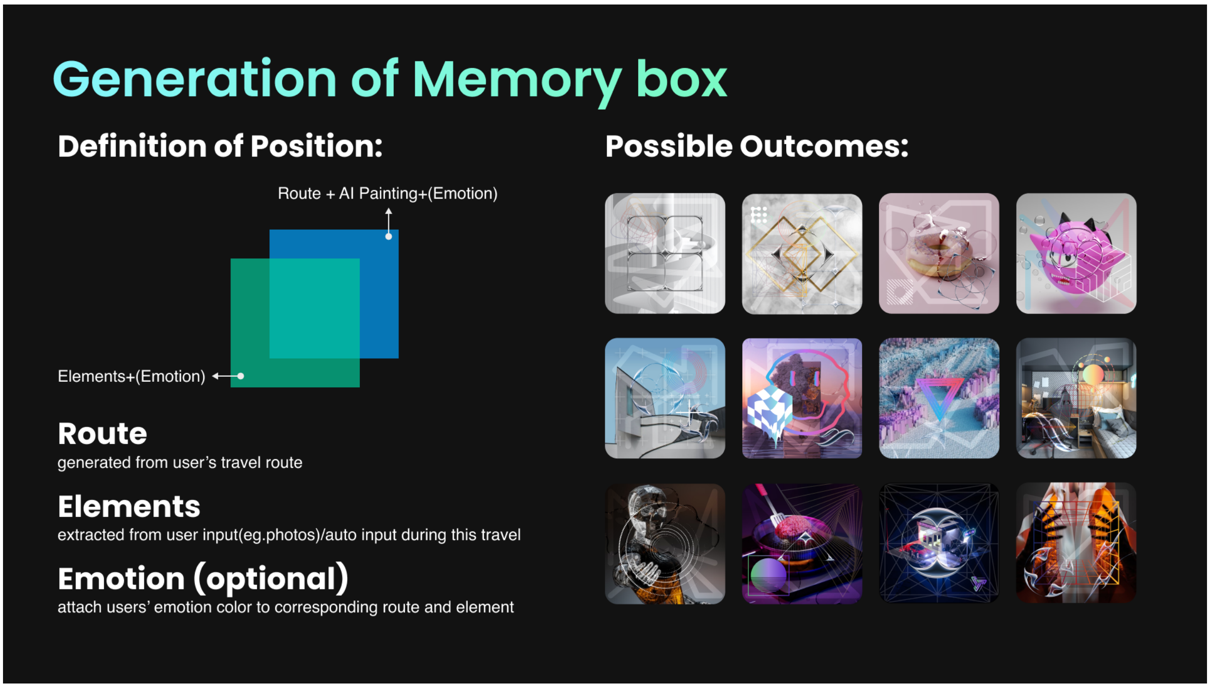

Figure 26. Generation of memory box (Own collection)

Each AR memory space can generate a unique memory box as a container to store the AR memory space. Three types of information input can be extracted from memory space and form the look of a memory box. Route: generated from the user's travel route; Elements: extracted from user input (e.g., photos) or auto-input during this trip; Emotion (optional): emotions are used to attach the user's vibrant colour to the corresponding route and element.

The front layer of the memory box is the elements and optional emotion; the back layer is the route and optional emotion (displayed as colour information).

In other words, the elements collected from the information input form an abstract illustration, like the cover of the space, which is helpful for users to identify the content of the experience quickly.

Mood Route

Figure 27. Mood Route (Own collection)

The following slides are the introduction to the main functions of our APP. The Mood Route function reflects the road line, via point and speed in the APP. During the process of action, the background will always input the information mentioned above. While users start their travel, APP will keep track of users' footprint and mood changes. The information recorded during the travel would be extracted and visualized in real time.

Realtime Record

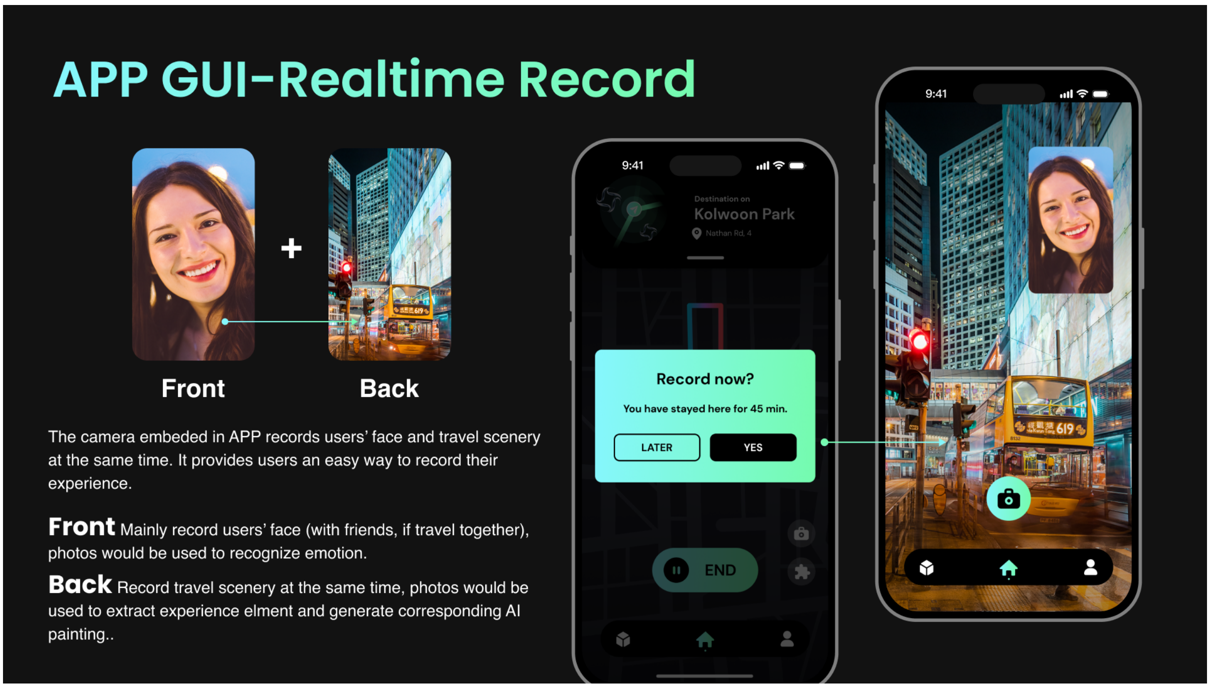

Figure 28. Realtime Record (Own collection)

"Realtime record" is a function that the system will automatically trigger after we stay in a particular

location for more than 45 minutes. The front and back cameras will be turned on at the same time. The front one recognizes the user's mood as input through a face recognition system. The back one is used for the user's own digital information input to generate elements and AI Painting.

Memory Space

Figure 29. Memory Space (Own collection)

The following is a simulated case reference in AR space to help us better understand the whole process. Through GPS, we can get the Road line, and display the Via point and speed in the APP. From photos uploaded by users, it extracts the mood, and determines the elements and AI painting. As shown in the right button, based on the space, a memory box will be automatically generated.

AI Painting & Pat pat

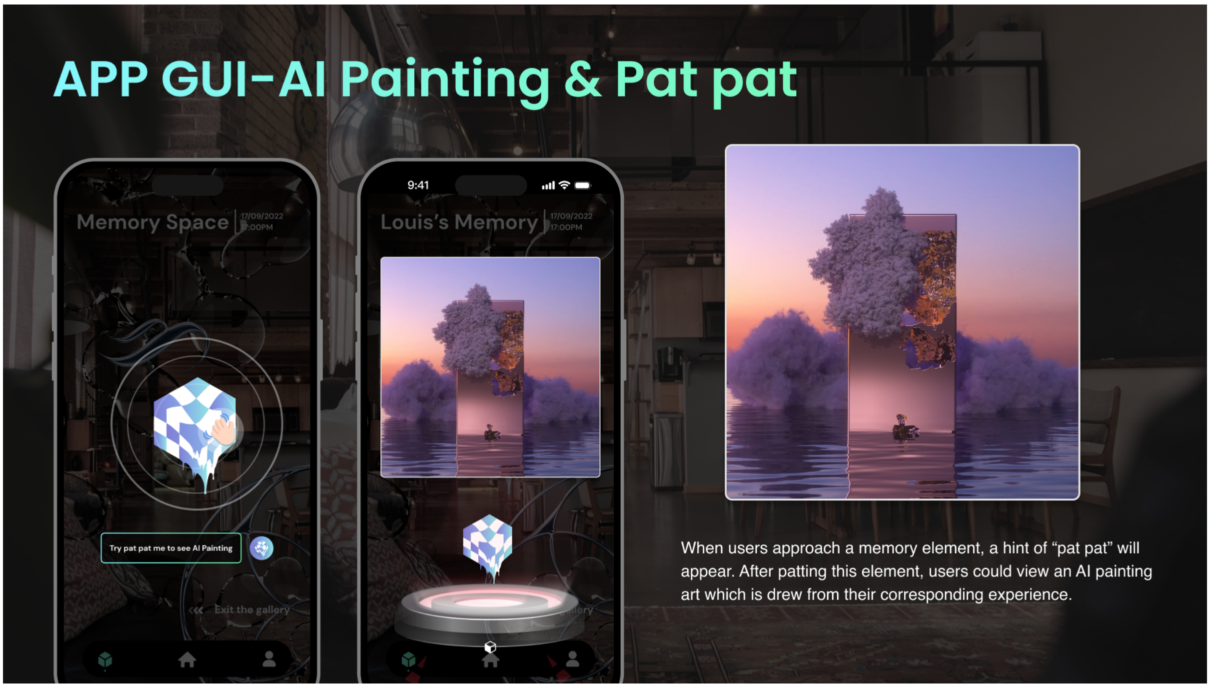

Figure 30. AI painting & Pat pat (Own collection)

In AR space, each element corresponds to a different AI painting.

When the user approaches the element in the AR space, there will be a [pat pat] reminder, and the content of the AI painting will be displayed after the operation.

Conclusion

Current product conclusion, pros

This study aimed to understand the behaviour and perception of young people aged from 18 to 30 who live alone regarding the intention to go out, record, and share the memory of travelling. Findings through questionnaires and interviews documented that they are willing to go out, but current recording and sharing methods have restricted them due to their concern about privacy. This source of information was utilised to design an intelligent system of travel experience visualisation and memory reappearance. It enables users to record their travels conveniently. And by using AI, a unique AR memory space is generated after the travel for them to memorise and look back at their trips immersively and vividly.

Current limitation

By offering an encrypted sharing medium to the target population, the initiative is now focused on tackling the resistance of young people who live alone and their loss of motivation to share. However, the generated output from the process of encrypting data into AI painting and AR spaces is too creative in its manner of art style, and there is a possibility to provide consumers with the option of customising their output.

The collection of AR settings and AI paintings in this project is not now a compelling reason for consumers to engage with the immersive experience over the long term. Is there a mechanism for rewarding users or converting values to make users stickier?

Additionally, it's critical to think about users' interest in AR and whether a manner helps them to switch between different modes.

Future direction

Our design solution mainly focuses on memory collection and reconstruction by using intelligent technologies. As mentioned in the Limitation, the functions we currently design in the app do not fully meet the needs of users.

From the perspective of APP features, we are going to explore more specific ways to provide users with the ability to generate images of their imagined memories, rather than just having the AI guess at what images to generate for them.

For user experience, we want to find out more touch points that could encourage users' continuous use. For example, adding a close friend circle in our APP may be a good idea to retain current users and attract new users, which is mentioned in our interview.

Considering the means of implementation, in addition to AR, we can also consider the current emerging and popular technology trends, such as metaverse, blockchain, etc. Metaverse can create a more immersive environment for users, while blockchain is the core technology solving the problem of data safety.

Appendix

A. Questionnaire

1. Opening

Hello everyone! Thank you for taking the time to participate in this questionnaire. We are master students of Hong Kong Polytechnic University. This survey is for the in-depth understanding of a user experience project. We hope to provide you with a better exploration experience by understanding your experience in exploring a city. All information collected by the questionnaire will be kept confidential and used only for academic research.

2. Background Information

1) Gender:

a. Female b.Male c.Could not give specifics

2) Age:

a. Under 15 b. 15-20 c.20-25 d.25-30 e. Over 30

3) Education background:

a. High school and below b. High school c.Bachelor degree d.Master degree and above

4) Occupation:

a. Student b. Administration c. Government-affiliated institutions d. Enterprise e.others

3. Behaviour/attitude/cognition Information

1) How many years have you lived in your current city？

a.Within half a year b.0.5-1 year c.1-2 years d.2-4 years d.Over 4 years

2) What aspects of the city do you tend to know

a.Food b.Shopping c.Scenery d.Entertainment d.Fitness e.others

3) Are you willing to explore new places or activities in a city?

a.Extremely b.Moderately c.Somewhat d.slightly d.Not at all

4) What are the reasons that prevent you from exploring new places or activities

a. Don't know where to go b.Unaccompanied c.Difficult to schedule d. Routes are too complicated e.others

5) How do you explore new places or activities in a city?

a. Map APP (such as Baidu Maps) b. Food APP (such as Meituan) c.Social APP (such as Little Red Book) d.others

6) What causes affect your satisfaction with using these APPs

a. Difficult to retrieve the desired information b. The operation is too complicated c . Need to switch between different platforms e. Can't share information with friends f.Others

7) How do you record your travels?

a. photo   b.video c.diary e. social media(eg. wechat moments) f. none

8) Have you ever been exposed to augmented reality technology (AR)?

a. yes b. no

9) Are you interested in augmented reality technology (AR)?

a.Extremely b.Moderately c.Somewhat d.slightly d.Not at all

Appendix B. Interview Guide

1. **Opening**

Hello, I'm the interviewer of city Explorer, I'm XXX, and this is my classmate XXX. Thank you for joining our interview. This time, we specially invited a group of users to interview, hoping to learn about your travel and social media use in the post-epidemic era, so we would like you to briefly introduce your situation.

We hope we can get some information and opinions from you in the form of a chat. It takes about an hour. There is no right or wrong question in this process, you just need to share your real thoughts and experiences. If you have any questions or opinions, you can explain them directly without any concern.

The information you provide can help us better understand users and further develop their experience. My classmate XX will record the whole conversation. The content of the conversation may be large. We hope you can authorize us to record. The recorded content is only used for academic research and will not be disclosed to ensure your privacy. Do you mind? (No recording if you mind).

During the chat process, we may ask questions about some details, which may be more thorough. Please don't mind, thank you!

2. **Basic Information**
    1) Age
    2) Gender
    3) Residence
3. **More Information**

Section1/ general behaviours and experiences of travelling and recording

1) Does the epidemic affect the frequency of going out?
2) What kind of places do you like to go to
3) What Reasons for going to this destination and the feelings
4) Do you have the habit of recording your experience of going out? Do you like recording?

Section2/ Needs and frustrations during traveling and recording

5) How do you record your experience of going out?
6) Why do you choose this way to record?
7) Are you willing to share your experience of going out?
    a) (willing)
        i) What methods (platforms, people) do you prefer to share?
        ii) What platform to share
        iii) Why do you choose this way of sharing? What's good is not good
        iv) How do you feel when sharing
    b) (unwilling)
        i) Why don't you want to share? What makes you unwilling to share? What concerns do you have?

Section3/ the feeling and perception of our probable solutions

8) If there is an application that can record your footprints, take photos, and generate NFT artworks specifically for you by AI based on your experience, how do you feel

References

5 key reinforcement learning principles explained by AI expert, Hadelin de Ponteves (2019) *Packt Hub*. Packt. Available at: https://hub.packtpub.com/5-key-reinforcement-learning-principles-explained-by-ai-expert/ (Accessed: 15 October 2022).

Albawi, S. et al. (2018) 'Understanding of a convolutional neural network', in Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017. [Online]. 2018 pp. 1–6.

Barnes, S.B. (2006) 'A privacy paradox: Social networking in the United States', *First Monday*, 11(9). Available at: https://doi.org/10.5210/fm.v11i9.1394.

Canal, F.Z. *et al.* (2022) 'A survey on facial emotion recognition techniques: A state-of-the-art literature review', *Information sciences*, 582, pp. 593–617.

Data-Access Layer (2022) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/data-access-layer/ (Accessed: 16 October 2022).

Emotion Recognition: Introduction to Emotion Reading Technology (2021) *RecFaces*. Available at: https://recfaces.com/articles/emotion-recognition (Accessed: 18 October 2022).

Fugate, J.M.B. and Franco, C.L. (2019) 'What Color Is Your Anger? Assessing Color-Emotion Pairings in English Speakers', *Frontiers in psychology*, 10, p. 206.

Imagen: Text-to-Image Diffusion Models (no date a). Available at: https://imagen.research.google/ (Accessed: 13 October 2022).

Imagen: Text-to-Image Diffusion Models (no date b). Available at: https://imagen.research.google/ (Accessed: 19 October 2022).

Jack, R.E. *et al.* (2012) 'Facial expressions of emotion are not culturally universal', *Proceedings of the National Academy of Sciences of the United States of America*, 109(19), pp. 7241–7244.

Lobo, M.J. and Christophe, S. (2020) 'Opportunities and challenges for augmented reality situated geographical visualization', *ISPRS Annals of Photogrammetry Remote Sensing and Spatial Information Sciences*, V-4-2020, pp. 163–170.

Ming, Y. *et al.* (2022) 'Visuals to Text: A Comprehensive Review on Automatic Image Captioning', *IEEE/CAA Journal of Automatica Sinica*, pp. 1339–1365. Available at: https://doi.org/10.1109/jas.2022.105734.

OpenAI (2022) *DALL·E 2*, *OpenAI*. Available at: https://openai.com/dall-e-2/ (Accessed: 13 October 2022).

Rick-Anderson (no date) *Creating a Data Access Layer (C#)*. Available at: https://learn.microsoft.com/en-us/aspnet/web-forms/overview/data-access/introduction/creating-a-data-access-layer-cs (Accessed: 16 October 2022).

Website (no date). Available at: https://www.pnas.org/doi/10.1073/pnas.1200155109.

Yang, B., Jianming, W. and Hattori, G. (2021) 'Face Mask Aware Robust Facial Expression Recognition During The Covid-19 Pandemic', *2021 IEEE International Conference on Image Processing (ICIP)* [Preprint]. Available at: https://doi.org/10.1109/icip42928.2021.9506047.

何志鹏， 李自娜， 梁艳， 邱丽娜 (2020) ‘多模态情绪识别研究综述’, 智能系统学报 [Preprint]. Available at: http://html.rhhz.net/tis/html/202001032.htm.