MASTER THESIS

Enhancing Active Reading: A Human-Machine Co-Creation Journey for Visualized Narratives Reading

Meng CAI
July 2024

Enhancing Active Reading: A Human-Machine Co-Creation Journey for Visualized Narratives Reading

Delft University of Technology (TU Delft), The Netherlands

Faculty of Industrial Design Engineering

MSc. Strategic Product Design

Master Thesis

July 2024

Supervisory Team

Chair - Dr. R.S.K. Chandrasegaran
Mentor - Dr. J.S. Love

Author
Meng Cai

Executive summary

This thesis explores the enhancement of active reading experiences through a human-machine co-creation process, focusing on visualized narratives. The project addresses the limitations of traditional reading materials that often neglect the individuality of readers in visualizing narratives. To bridge this gap, the project adopts a modified ZEN design approach, emphasizing user rituals and the integration of AI tools for visualizing mental models. By integrating artificial intelligence (AI) and generative technologies, the project combines human creativity with machine efficiency, aiming to create personalized digital reading experiences that cater to individual cognitive abilities and preferences.

The project begins with a comprehensive literature review that establishes the theoretical foundation for understanding narrative books and active reading. Realizing each one's active reading process and mental images regarding the same contents are different, the core of this project is to create a personalized active reading journey. Then, design vision is proposed to identify design goals and value delivery. To understand the actual user ritual, qualitative user research is conducted to study the reading activities and mental models of active readers. Then, through competitive products research, mental image visualization is found to be the active reading function currently missing in the reading products on the market. Thanks to the emerging AIGC technology, it becomes possible to generate desired visual images for people with different mental imagery abilities with high efficiency and quality. By strategically synthesizing the user requirements with image generation workflow, design round 1 focuses on content personalization. 2 solution flow was designed and tested by 10 participants with different mental imagery abilities (results from VVIQ test). After iterating the solution flow of mental image visualization function, design round 2 is conducted to realize toolkit personalization. Scenebites as an integrated active reading platform is introduced, empowering diverse user rituals of active processes. The value proposition of design vision and key user needs are validated through focus group research.

The final design output of this project is a human-machine collaborative system flow for personalized digital reading experience on knowledge construction, together with business model propositions for future libraries. The results demonstrate that personalized visualizations significantly enhance active reading experiences by reducing cognitive load and increasing engagement. And the assistance of visual options and editable mood boards in the user journey can positively match the requirements from users with diverse visual imagery abilities.

The project also suggests improvements in AI models to enhance precision and control in generating visuals. It also recommends strategies for libraries and educational institutions to integrate personalized reading services, thereby transforming the digital reading landscape into a more interactive and user-centered journey.

Table of contents

1. Introduction

Narrative books offer readers the freedom to envision multiple scenarios and characters, but visualization of text based on individual visual imagery ability and cognitive experiences perform an important role in creating an active reading experience —- think beyond default text to discover, construct knowledge and build mental model (Victor, 2011; Masson, 2023). When reading narratives, many people visualize the characters and plot, and mentally construct storylines and scenes based on the information provided (Williams, 2023). The project emphasizes the significance of visualizing mental models to actively construct knowledge in reading experience. Personalization — customized features offering and adaptive visual interpretations based on a reader's individual input (Adar et al, 2017; Hohman et al, 2020) becomes crucial as each individual's cognitive abilities and preferences are different. However, current reading material often overlooks the individuality of active readers in visualizing narratives. The project seizes the opportunity presented by narrative books as an ideal domain to visualize storylines and scenarios, and the design aims to deliver an active reading experience for knowledge construction.

However, building reading materials with domain-specific visuals is difficult, as each book must be independently developed, incorporating the subjective ideas of the team (Hohman et al,2020). Amidst the ongoing digital transformation in reading, the project acknowledges the possibilities of implementing new AIGC tools in supporting active reading. To ensure human domination on creativity, human-machine co-creation is at the heart of this design, combining human creativity with machine efficiency to create adaptive reading services tailored to individual preferences. This collaborative approach offers a user-controlled service for digital visual reading experiences.

Future-oriented libraries will pay more attention in creating personalized reading experiences and embracing emerging technology to be more attractive to active readers (Jayawardana et al., 2001), and promoting AIGC technologies can bring innovative solutions in achieving this (Jayavadivel et al., 2024). The project starts with researching the user activities and needs in active reading experience, then identifies the function gap of current digital narrative reader in supporting these activities. By embracing innovative AI image generation technologies, visualized mental models can be produced by active readers' input. Then, the ideal mental image visualization system will be integrated into the whole user journey in creating an optimized active reading process. Strategically implementing the personalized active reading service, libraries can upgrade the digital reading landscape into a more productive and user-controlled journey.

2. Project approach

This design project follows a modified ZEN design procedure as the user ritual is emphasized in this procedure. The basic principle of the ‘ZEN’ design method is: ‘Do not focus on the desired product for quite some time’. The primary focus are rituals and qualities, which may be the preferred method for designers who would like to achieve innovation in terms of functionality, social interaction and user experience (Persig, 1974; Van Boeijen et al., 2014). The following figures illustrated the overall design process.

This project approach provides a structured and user-centered roadmap for designing a personalized digital reading experience.

The first part of literature review establishes a solid theoretical foundation that informs subsequent design decisions.Then, identifying basic opportunities of the ritual helps understanding the technology landscape and pinpointing areas for innovation. Afterwards, user requirements and opportunities are strategically combined to inform the design. The design will firstly focus on the key function facilitation, then expand to the original design scope on the holistic product service design. With a clear concept of the product service and business model, the prototype will be validated and given implementation suggestions.

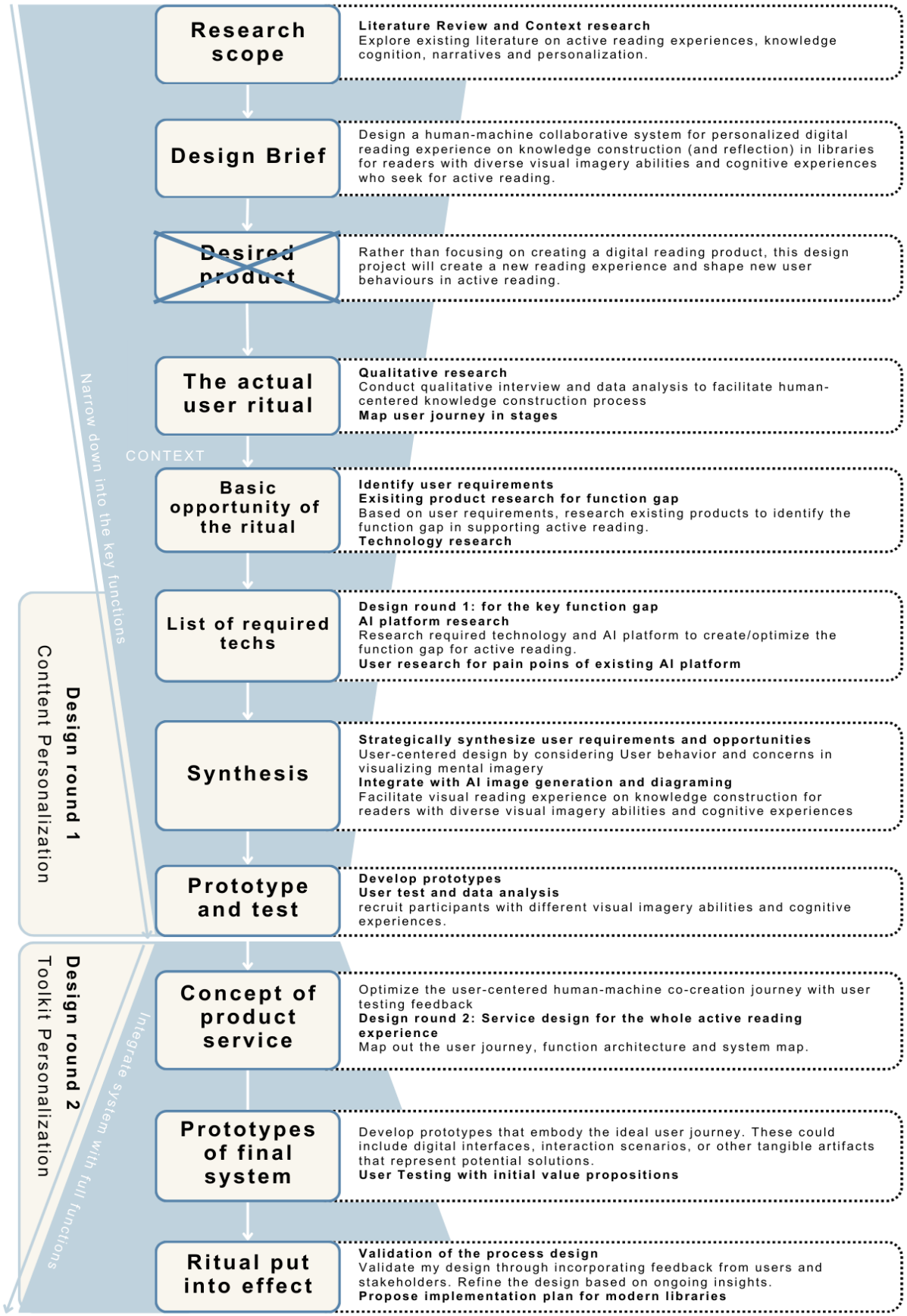
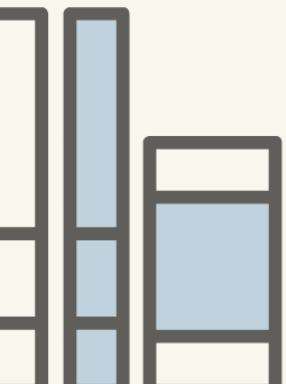


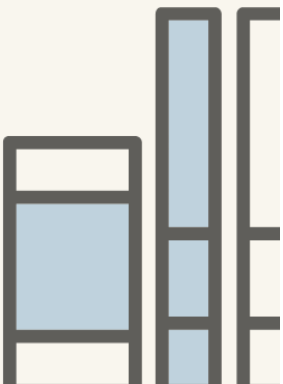
Figure 1: The project approach

Chapter 3

Literature Review

In this chapter:





3.1 Narrative books

Types of narrative books range from fiction and non-fiction. Fiction includes literary, fantasy, thriller/mysteries, and science fiction. In non-fiction, news, political writings, history books and biographies are commonly mentioned (Williams, 2023).

Different from other types of reading, narrative reading involves engaging with texts that convey a story structured around a plot and characters. They are presented to support information about the characters, conflict and resolution, and various other story-specific features (Boyd et al., 2020). It primarily focuses on understanding the sequence of events, character development, and the overarching themes that drive the narrative forward. This form of reading is distinct because it involves the construction of a mental model or a situation model that encompasses the setting, characters, and events described in the story. Readers engage in constructing inferences and integrating knowledge from the text with their existing knowledge to make sense of the narrative (Graesser et al., 1994). Appendix 1 shows several common formats of narrative structure analysis.

Narratives typically involve complex settings, characters, and events that readers want to visualize to fully understand and engage with the story. In this case, visualization helps in constructing a mental model of these settings and events, making it easier to follow the plot and empathize with the characters (Brosch, 2018). This mental imagery is less critical in other types of books like expository books, which focus more on facts and information than on creating a vivid scene or storyline.

3.2 Active reading

For narrative readers who actively read, they focus on interactive documents to explore, build intuition, make discoveries, and reinforce or challenge knowledge. “An active reader doesn’t passively sponge up information, but uses the author’s argument as a springboard for critical thought and deep understanding.” (Victor, 2011)

Pearson and Dole's (1991) research on reading determined that active readers, or "good readers”, employ several comprehension strategies before, during, and after reading. These comprehension skills include:

Before Reading

- *Survey the reading and assignment.*
- *Activate background knowledge.*
- *Understand the task.*

During Reading

- *Read actively: Highlight or underline important facts and information.*
- *Code or write important notes in the margins or use sticky notes to record ideas, questions, predictions and responses.*
- *Make connections with what I already know or have experienced.*
- *Monitor comprehension.*
- *Visualize to enhance comprehension.*

- *Make inferences, conclusions, evaluations, and interpretations.*

After Reading

- *Reflect: What did I learn? How is this important?*
- *Summarize.*
- *Synthesize and extend my thinking.*

Beyond passive reading, active reading process involve a rich interaction with the information in texts (Tashman & Edwards, 2011), demanding active reading activities including:

Annotation. Annotation involves embellishing the original text, as in highlighting, marginalia, etc. This process requires high efficiency, which can be compromised when switching between different annotation tools (O'Hara, 1996). It also requires support for unique markings and effective retrieval of previous annotations (O'hara & Sellen, 1997).

Content extraction. Content extraction involves enhancing the original text through methods such as highlighting and adding marginal notes. Users also must easily be able to organize, review, and trace their extracted content back to its source (O'Hara, 1996).

Coding. Contract or expand sentences to accurately express personal understanding. During this process, active readers may modify the information based on personal memory, information queried from other people or documents, and emotional responses.

Navigation. Navigation involves moving both within and between documents, as when searching for text, turning a page, or flipping between locations to compare parts of a text (Askwall, 1985; O'hara & Sellen, 1997).

Layout. Layout is the process of visually and spatially organizing one's understanding and materials, such as when cross-referencing, comparing items side by side, or obtaining an overview of a document (O'Hara, 1996).

Diagramming. Creating graphical representations by diagramming allows the reader to further integrate concepts into a cohesive mental model, which proves to be a more effective activity when compared to simple note-taking. (Subramonyam et al., 2020). Diagramming processes include: (1) selection, (2) organization, and (3) integration aids (Mayer, 1984).

Picturing. When engaging in reading, many people are compelled to conjure the plot, visualize the characters, and mentally construct storylines and scenes based on the information provided (Williams, 2023). Akin to active reading, the creation process is part of sensemaking. By enabling readers to select and decide on the best **visualization**, they actively participate in understanding the text and the data (Masson, 2023).

Questioning. Active readers like to raise questions about allusions, history, and plot settings in narrative reading, which is an important way to enhance comprehension and construct knowledge (Pearson & Dole, 1991).

Summarizing. Provide a summary, general description of the narrative. It often happens when active readers finish reading a narrative sector, which is helpful for mental reflection and knowledge organization.

Predicting: Make predictions about plot development and story endings based on personal reasoning or expectations.

Most readers engage in active reading to externalize their thoughts, optimize cognitive load, improve their understanding, emphasize key concepts, help memorization, aid in later retrieval and share their marks (Masson, 2023; Matthew et al., 2019). Therefore, knowledge construction shown as a mental model is a key delivery when users are doing active reading.

3.2.1 Knowledge construction

Reading comprehension is an active process where readers construct meaning as they decode the text. Typically, readers do not retain the exact wording of the text but instead develop more adaptable knowledge structures (Woolley, 2014). When learning from text, readers develop meaning by merging sentence-level information into a coherent mental model (Kintsch,1988). The research results of cognitive linguistics show that the process of readers reading and understanding text is the process of readers discovering schemas, activating schemas and updating schemas. Schema is a representation of the brain (Kant, 1787), which is the smallest unit that forms the human knowledge system. Besides, the formation of feelings, understanding and memory are all affected by people's original knowledge system (Bartlett, 1932).

Knowledge construction in reading activities refers to the process of **information gathering, information comprehension, knowledge integration and mental model construction** (Verhoeven & Arthur, 2008; Jayawardana et al., 2001).

Information gathering. When gathering information, the user may create new artifacts such as notes or highlighting, by integrating selected segments of reading materials with their own comments. These segments would be text, images, audio or video depending on the type of original source (Jayawardana et al., 2001).

Information comprehension. Information comprehension is the reader's gradual construction of a cognitive representation that captures the meaning of the text. The representation may be modeled as a network of concepts and semantic propositions, including mental images or the reader's emotional responses (Saux et al., 2021; Kintsch, 1998).

Knowledge integration. This is accomplished by prioritizing information that aligns best with the discourse context and the readers' prior knowledge, while deemphasizing less central or consistent information (Saux et al., 2021). To integrate information and construct the model, readers process both verbal and pictorial information, combining them with prior knowledge stored in memory (Verhoeven & Arthur, 2008; Mayer & Moreno, 2003).

Mental model construction. In the context of reading, a mental model refers to the cognitive framework or set of beliefs and understandings that a reader uses to interpret texts. It is essentially the internal representation of the external world that a reader relies on to process, comprehend, and integrate information from the text. (Garnham & Jane, 1996). Mental models are shaped by a reader's prior knowledge, experiences, cultural background, and personal beliefs,

influencing how they comprehend and engage with the material. To construct a coherent mental model, readers generate inference including induction and deduction that connect information across different texts and visuals (McCarthy et al., 2021). Theories generally agreed that a reader's mental model consists of multiple levels of representation (Kintsch, 1988).

Knowledge construction and mental model visualization happens in a self-directed, active reading environment —- It cannot rely solely on passive, memory-based understanding and reasoning (McCarthy, 2021). This process must be supported by facilitating the construction of information artifacts that represent the learner’s information gathering process (Jayawardana et al., 2001). Effective reading strategies of active reading, , such as highlighting, questioning and summarizing, can help readers construct and adapt their mental models.

3.2.2 Knowledge reflection

Reflecting on material that reader is learning and explain the knowledge architecture themselves — a learning technique called self-explanation — is known to have a positive impact on learning outcomes (Chi et al., 1989). In the process of knowledge construction, through generating explanations and refining them as new information is obtained, readers actively reflect on the information they absorbed in mind, which is beneficial for formulating deep understanding and long-term memory (Jayawardana et al., 2001). Different fields of research all indicate that mapping mental models will reflect the experiences, perceptions, assumptions, knowledge and subjective beliefs that a “model user” draws upon to reach his conclusion about some assumption (Kolkman et al., 2005).

Knowledge construction involves a large number of component skills, it chunks multiple elements of information into single elements, generalizations, or cognitive schemas that can subsequently be elaborated and reflected (Verhoeven & Arthur, 2008). However, the current reading environment in libraries does not encourage active reading experience. A typical reading tool, such as a book or digital application, displays the author's argument, and nothing else. The reader's line of thought remains internal and invisible, vague and speculative. We form questions, but can't answer them. We consider alternatives, but can't explore them (Victor, 2011; Schwartz, 2018). Initial efforts to bolster digital active reading focused on bringing the capabilities of pen and paper to screen-based environments and proceeded to extend them through powerful **annotation, content extraction, layout, navigation, diagramming, picturing, questioning and summarizing** (Masson, 2023; Matthew et al., 2019). Despite the hardware upgrading for digital transformation, the user experiences in constructing mental models still need to be optimized.

3.3 Visualization

Unlike images or films that present a predetermined story scene, narrative books empower readers to envision scenarios with unparalleled freedom. These literary works offer glimpses into diverse lives and cultures, occasionally presenting historical or exotic concepts that may seem incomprehensible to modern readers.

Recognizing that a decent percent of the population are visual learners (Imran et al., 2014), when we engage in reading, many people are compelled to conjure the plot, visualize the characters, and mentally construct storylines and scenes based on the information provided (Williams, 2023), thereby granting our minds big spaces for association. When readers actively visualize story events, they are able to make appropriate inferences because visualizing enables them to draw on their own prior knowledge and life experiences. As readers visualize while reading, they become more engaged with the text, enjoy what they are reading, and often imagine themselves in the story. Imagining story ideas during the reading process links information in working memory and makes the encoding and recall of information more efficient (Woolley, 2014).

Visual storytelling involves interacting with audiences using graphics, images, diagrams, and videos. Complex stories may be made simpler to grasp via the use of visual storytelling, which helps to convey a point more effectively. It makes tales as vivid, topical, and meaningful to viewers as humanly possible (Kowsalya, 2024).

3.3.1 Diagrams

Mind maps and diagrams help visualize relationships between different segments of information, facilitating easier comprehension and recall. This visual representation helps in understanding complex narratives by breaking them down into simpler, connected components, thus enhancing memory retention and recall (Eppler, 2006).

Relevant research proves that transforming texts into diagrams supports inference generation and reduces comprehension errors, with simplified diagrams being particularly effective for information integration (Butcher, 2006).

At the same time, the process of creating mind maps and diagrams can increase readers' motivation and cognitive engagement with the material. It encourages them to explore various narrative paths and possibilities, thus fostering creativity and a deeper engagement with the content (Winn, 1991).

3.3.2 Images

Images in books with text may increase cognitive efficiency and help comprehension. Specifically, the dual-coding theory suggests that people mentally represent information verbally and visually through two interconnected but distinct systems. Thus, presenting both text and images allows activating both systems and drawing connections between them to enhance understanding(Masson, 2023; Paivio, 1990).

Images in books alongside text can also enhance recall. According to the dual-coding theory, both text and images aid memorization by encoding the information through both verbal and visual systems (Paivio, 1990). Experiments testing the "multimedia principle" have shown that information is better learned when presented through a combination of text and related images (Mayer, 2014).

It has been acknowledged that imagery information usually helps knowledge comprehension and recall of text information. Cox et al. (1994)'s research further defined the effects of visual elaborations on the cognitive processes of adult learners in two forms. Internal imagery elaborations are generated internally by the learner. They are instructed in some way to "see" a specific image in their minds. For external imagery elaborations, the learners are asked to produce or complete an image on paper. In their test**, overall group means indicated that the group that generated external visual elaborations and drew sketches were higher than the means for both the group that used internal visual elaborations and the read-only control group on both the immediate and delayed tests of recall.** Therefore, visualizing mental pictures on screen has been proved to be effective in enhancing knowledge construction in terms of recalling.

3.4 Personalization

Continuing to assume that most people's reading experience is the same will obviously have potentially negative consequences (Williams, 2023). Each person's cognitive processes and experiences are unique, and factors including visual imagery abilities and personal preference also contribute to their imagination and understanding (Rahimi et al., 2021). In terms of mental imagery abilities, Schnotz & Rasch's research (2005) showed that manipulation pictures had an enabling function for individuals with strong visual abilities, whereas simulation pictures had a facilitating function for individuals with weak visual abilities. Williams's research (2023) studied the Impact of Aphantasia levels on reading experiences, showing that lack of mental imagery ability significantly affected their attention and comprehension. There are clearly differences in perceptions of the reading experience between Aphantasics and those who consider imagery to be an essential part of their reading experience.

For example, as depicted in the following figure, individual readers picture the same text Dragon into different images based on their cognitive experiences and imagery abilities. Eastern readers may think of an eastern dragon without wings, while western readers picture it with wings. Readers with weak mental imagery abilities find it hard to visualize the text, while extreme Aphantasia readers can picture it in a very vivid and colorful way. It's important to embrace these differences and empower personalized reading experience through design.

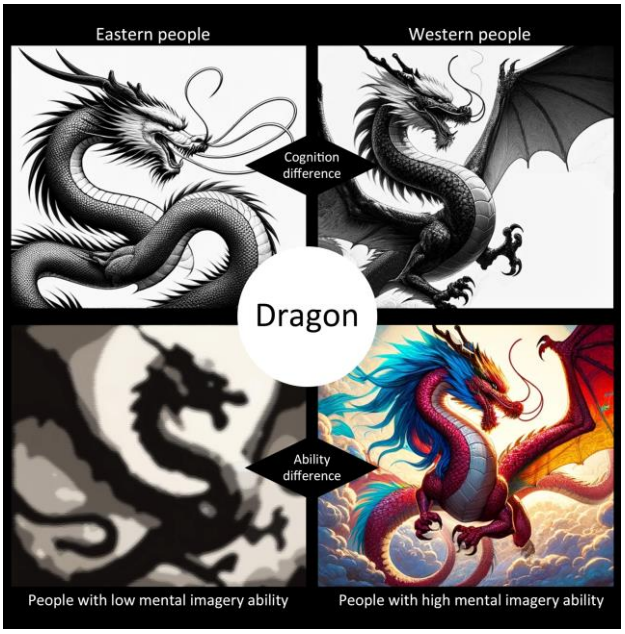


Figure 2: Mental imagery of people with different cognitive experiences and visual abilities for the same text

Cognitive procedure of reading narratives is powered by reader's past experiences and comprehension abilities, default pictures could promote or inhibit reader's understanding of text (Schnotz & Rasch, 2005; Verhoeven & Arthur, 2008). Existing books often overlook the individuality of how readers interpret and visualize narratives, and one-size-fits-all literary content leads to passive reading (Haiken et al., 2022). And existing images in narrative books are basically generated by humans, and cannot avoid subjectivity of the illustrator (Hohman et al., 2020). Readers with different cognitive abilities and preferences may disagree with the illustrator.

The personalized reading design has emerged in two dimensions, content and toolkit personalization. In Jayawardana's (2001) personalized information environment design, the two dimensions are phrased as Material personalization and collection personalization. Material personalization corresponds to facilities multimedia tools for learners to use library materials according to their individual requirements. Collection personalization, on the other hand, captures the learner's learning context and interest from the material personalization in order to provide a personalized knowledge view of reading materials.

PersaLog, designed by Adar et al. (2017), introduced content personalization, which automatically adjusts the facts presented in an article based on the reader's inferred characteristics. This approach has been shown to reduce information overload and enhance reader engagement. Personalized visualization offers a reading and learning experience tailored to the reader's individual visual capabilities, needs, and interests (Masson, 2023).

To comprehensively achieve personalized reading, the design scope has to touch upon not only **content personalization**, but also ***toolkit personalization***, which means different tools and inspirations to facilitate the visualized knowledge construction in order to read narratives actively. In Haiken's (2022) book *Personalized Reading: Digital Strategies and Tools to Support All Learners*, different strategies and technical tools are

recommended to foster active reading and comprehension for readers at different levels:

> For beginner readers, multimedia tools like images, storyboard and video can assist in synthesizing understandings of texts. Podcast and reader vibration can enhance engagement from auditory and tactile aspects, thereby enhancing knowledge construction.
>
> Advanced readers are introduced to Visual Curation Tools, enabling them to organize graphics and map concepts freely. Providing a choices menu with various tools such as Think-Tac-Toe, Bingo boards, and question cards empowers users to create their own reading tasks. Additionally, gamification, as an active reading strategy, proves valuable in sustaining enthusiasm and attention during reading activities.

The advancement of technologies has greatly contributed to making more personalized reading experiences such as personalized websites or mobile apps. In this way the system conforms to the readers abilities and preferences based on the user inputs. Research confirms that personalized reading platforms can create a uniquely joyful reading atmosphere, therefore promoting active learning and understanding (Rahimi et al., 2021).

However, Existing books often overlook the individuality of how readers interpret and visualize narratives to address the varied mental imagery abilities and cognitive experiences of readers, and one-size-fits-all literary content leads to passive reading (Haiken et al., 2022). Unfortunately, creating personalized reading materials today is difficult. Domain-specific visualizations must be individually designed and implemented, which unavoidably incorporates the subjective ideas of the team (Hohman et al., 2020; Adar et al, 2017). While personalized reading materials need to be intuitive and performant based on individual reader's input to achieve a nice reading experience.

3.5 Future library

Library development can be influenced by technological developments. The representative technological features are referred to as a symbiosis web, reading, writing, and executing simultaneously, web OS, middleware, and a massive web allowing intelligence interaction just like a human brain (Yıkılmaz, 2020). Supported by emerging technology, the future library will become an intelligent library where not only inference and research are available, but the system will analyze information by itself and discuss findings with users like a colleague (Noh, 2015). In the massive web, the users create and share knowledge as group intelligence and will expand open source contents. Users reading digital books do not simply receive searched data in the form of PDFs but begin to use all the available data through dynamic connection (Crane et al. 2006).

Artificial intelligence (AI) is one of the latest digital transformation (DT) technological trends the future library can use to upgrade digital reading experiences. However, the recent findings of Wheatley and Hervieux (2019) affirm that incorporating the usage of AI into library services strategic plans of the future is yet to be visible. Future-oriented libraries are encouraged to actively research and adopt AI technologies to deliver value-added innovative library services to achieve digital transformation (Okunlaya et al., 2022).

However, the current digital reading experience of libraries is the lack of customized service to support active readers in constructing and visualizing knowledge and mental models (Schwartz, 2018).

3.6 Human-machine co-creation

Human-machine co-creation harnesses the strengths of both human creativity and machine efficiency (Hermann, 2022; Rahimi et al., 2021). This synergy allows for the creation of a dynamic and adaptive reading platform that tailors visual elements to individual preferences with high efficiency. By involving users in the co-creation process, the design becomes more attuned to diverse needs, ensuring that the technology serves readers effectively and inclusively.

According to the depth of the machine's involvement, existing human-machine co-creations could be put into three categories: machine as an expression tool and machine as the medium of experience and machine as a creative service (Zhuo,2021). The cooperation between active readers and AI machines in this project exists as the creative services rather than as physical entities used by them. At this stage, machines provide users with personalized and flexible creative assistance. The machine will take its advantages to stimulate human's creativity.

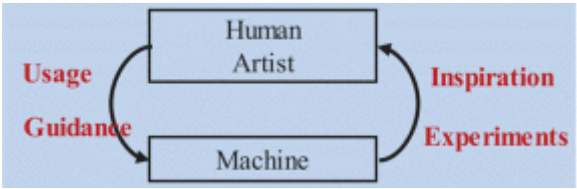


Figure 3: Creative services stage of human-machine interactions (Zhuo,2021)

To make people and machines live in harmony, we must first clarify their respective roles. Humans have unique advantages in creativity, emotional communication and decision-making. For example, artistic creation, psychological counseling, business management, etc. require inspiration, empathy and flexible decision-making ability, which are more suitable for humans to complete. Machines are good at highly repetitive, highly dangerous and precise calculation tasks (Leonhard, 2016). Through reasonable allocation, people and machines can do their best and maximize their efficiency.

3.7 Key take-aways

✧ Narrative reading:

- Reading narratives involves integrating the text's settings, characters, and events. Actively reading and mental model construction are crucial for understanding and engaging with the story, as it helps readers visualize complex settings and empathize with stories.

✧ Active Reading Strategies:

- Active reading is emphasized as an interactive engagement with texts that involves annotating, questioning, summarizing, and predicting etc. It encourages readers to not passively absorb information but to engage critically and constructively with the content.

✧ Visualization and Mental Model Construction:

- Envisioning story scenarios and elaborating story logics are natural mental activities in narrative reading.

- Mapping mental models can enhance knowledge comprehension and reflection. Diagramming and picturing texts can help readers integrate concepts and build cohesive mental models, making complex information more understandable and memorable.

✧ Personalized Reading Experience:

- Diverse Cognitive Abilities: It’s important to acknowledge the variation in visual imagery capabilities and cognitive experiences among readers, and suggests that personalized reading experiences should accommodate these differences.

- Content and toolkit personalization: This includes tailor the reading experience to individual preferences and cognitive styles not only in generating contents but also in toolkit offerings.

✧ Future of Libraries and Digital Reading:

- Future libraries will leverage emerging technologies like AI to transform reading experiences, suggesting a shift towards more intelligent, interactive, and personalized learning environments.

✧ Human-Machine Co-Creation:

- The synergy between human creativity and machine efficiency is harnessed to create personalized reading experiences.

Chapter 4

Design Vision

In this chapter:

The design vision to enhance active reading experience in future libraries is:

Design a human-machine collaborative system for personalized digital reading experience on knowledge construction (and reflection) in libraries for diverse readers who seek for active reading.

Figure 3: Possible future reading scenario

4.1 Functions

The most important functions for this design are:

- **Facilitating active reading:** The system should encourage active reading through understanding the characteristics of narrative books and the knowledge construction process of readers.
- **Visualizing narratives:** The system should be capable of generating visual interpretations of narrative content that are tailored to each user's unique visual imagery capabilities and cognitive experiences.
- **Integrating with future libraries:** The system should integrate smoothly with future library environments, in order to realize digital transformation.

4.2 Value propositions

Human-centered:

- Put human needs and rituals in the center of the design process.
- Encourage humans to construct and reflect narratives actively through visualization.
- Make sure humans dominate the human-machine collaboration process.

Personalization:

- Adaptive visual interpretations: automatically modifying multimedia with pre-sented texts based on a reader's individual input (Adar et al, 2017; Hohman et al, 2020). The visual

interpretations can change depending on the reader's knowledge and preference, such as using terms or values (Masson, 2023).

- Customized features offering: targetedly providing readers with different mental imagery abilities with different digital tools to help them freely and positively construct knowledge and reflect.

Future-oriented:

- Incorporates future technological advancements and user needs to deliver an innovative, adaptable, and sustainable reading experience that enhances learning, engagement, and personalization.
- Propose an AI-powered reading service to upgrade the user experience in future libraries
-

4.3 Interested Party

Stakeholders include **readers** seeking personalized and active reading experiences, and future-oriented **libraries** aspiring innovative technologies to ensure optimized reading experience and accessibility for readers with different imagery abilities.

4.4 Environment for use

This product is going to be placed into future libraries, which have rich narrative book resources, embrace the ever-changing AI technology, and provide a variety of digital devices to help readers actively read.

Chapter 5

User ritual research

In this chapter:

In order to understand the actual user ritual of active reading journey and mental model construction outcome, qualitative research was conducted in Rotterdam Public Library with modern narrative readers (N=10).

The core research question in investigating user ritual is: How do people approach active reading on narratives and visualize mental models? Through 30 mins user experiment and interview, the expected outcome includes but is not limited to:

- Active reading process to knowledge construction
- Components of expected mental model
- User needs for active reading

5.1 Research Method

Research on knowledge understanding and construction relies heavily on open-ended measures such as think-aloud protocols, interviews, and essays. In a **think-aloud protocol** (also called a talk-aloud or verbal protocol), readers are asked to verbalize their thoughts as they read. These utterances reflect the processes and strategies that readers engage in during reading. Although asking readers to engage in this type of reflective speech can be potentially damaging (Branch, 2000; McCarthy, 2021), research shows that the information gathered through thinking aloud reflects the natural reading process (Magliano & Millis, 2003).

Another way to investigate readers' mental models of literature is through **post-reading interviews**. Unlike thinking aloud, post-reading interviews do not require interruption of the reading process. Instead, readers are asked to reflect on what they have read, either by describing what they think they did during the reading process (Ericsson & Simon, 1980), or by answering structured or semi-structured questions.

In order to cater to reading styles and habits of different individuals, this research takes a combined method of think-aloud protocol and post-reading interviews. Since think-aloud can have instant reports and feedback of the participant's reading processes, I will recommend them to describe their thoughts while reading the provided materials. However, to minimize interruptions to the reader's in-depth reading of the narrative material, it's not a compulsory requirement, participants can perform silent reading as they wish. After reading the whole narrative paragraph, a post-reading interview will be conducted, when participants are asked to reflect on the reading materials and answer the semi structured questions in order to map their active reading process and outcome.

Besides, Phenomenological Research strategy (Moustakas, 1994) is applied: The essence of human experiences regarding a phenomenon, as described by participants, is identified by the researcher, and researchers act as outside observers that will set aside their own experiences. I will observe the behaviors and activities that participants take from an external perspective, then take notes as a validation of the following user interview.

The provided active reading journey mapping material and research equipment are listed in Appendix 2, as well as the detailed interview plan.

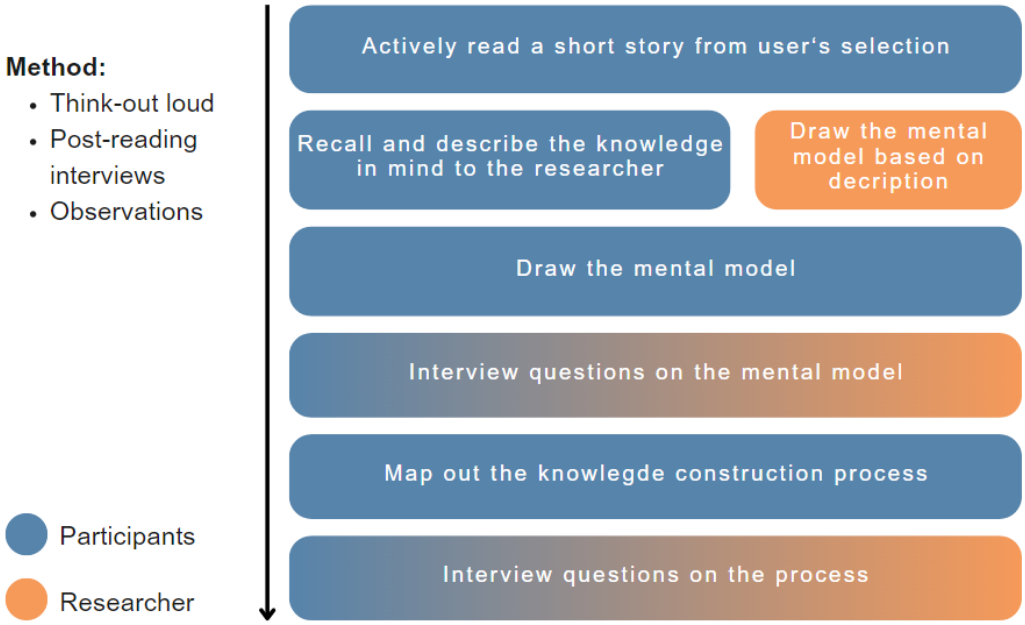


Figure 4: User ritual research facilitation

5.2 Data analysis method

To map out the active reading process to knowledge construction, the 10 participants' active reading activities across stages were integrated based on their essence. The integrated process map also includes the potential linear link that users thought out loud and the researcher observed.

Elemental disassembly analysis was adopted for the visualized mental models created by participants and collected online. From every partial component, the smallest units of activer reader's mental model are identified. Then, strategically combining the smallest units into categories, the components of the common mental model of knowledge construction of narrative materials are categorized. Also, I researched the relevant literature and bilaterally proved the components of the results of human's mental model construction of information.

Thematic analysis was adopted for the interview transcripts, coding and laddering techniques were applied to user replies, to identify the basic and latent user needs for smooth active reading experiences.

5.3 Results

5.3.1 Knowledge Construction Process

The 9 participants recalled their mental model construction process and mapped out the activities they did before reading, during reading and after reading by dragging the elements into the process map. Many participants found the elements are useful in offering directions when they look back into their mental model construction process, as they usually don't realize the exact mental activities when reading.

Generally looking into all the 9 process maps, people all have different mental and physical behaviors when doing active reading. This difference is shaped by individual preferences and abilities. It's unreasonable to offer the same functions to support different individuals' active reading process.

The following figure 5 depicts the integrated process of mental model construction, it includes all the activities the 9 participants have taken in their process map (the definitions of those activity terms are listed in 3.2.1). The arrows between steps indicate the linear connection, for example, when readers raise questions during reading, the next step they take is normally searching for explanations at once or making notes on that question so they can search for it later. This figure can serve as a foundation of the function architecture elaboration in the design phase.

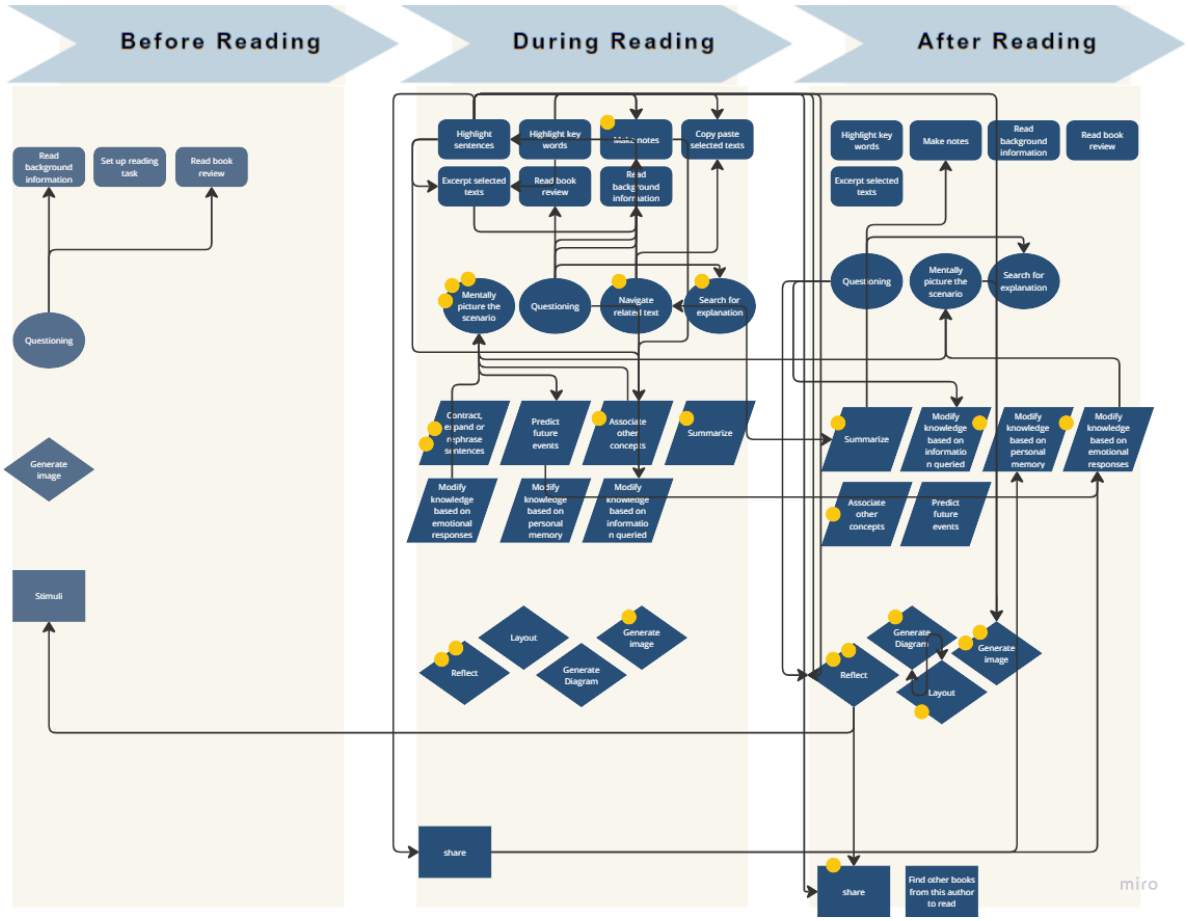


Figure 5: Integrated results of active reading process

Horizontally comparing the activities of the three categories of the process map (Figure 6), I saw the same elements appeared in the different categories. The following figure 1 demonstrates the explanations of the differences. In general, the activities done before reading are based on imaginative understanding of book contents, which is expected to be verified or solved in reading. Those activities done during reading are based on immediate understanding of specific contents, this understanding could be modified by new information gained from following contents. And those activities done after reading are based on holistic understanding of the whole book, sometimes combined with reflective understanding of personal experiences or predictions. Figure 6 can serve as an important consideration when designing the specific features offered in functions of different reading stages.

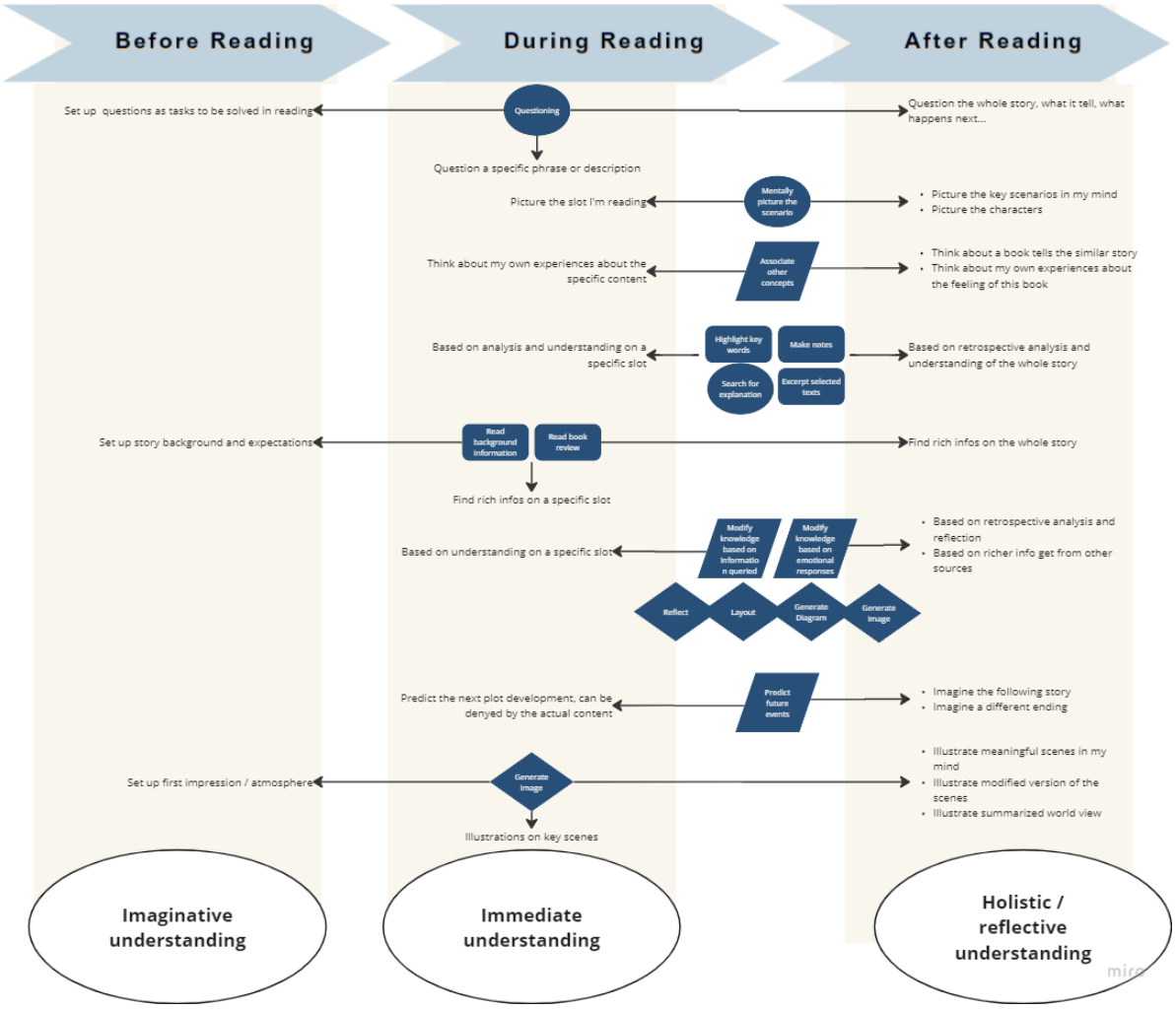


Figure 6: Different purposes for the same activity at different stages

5.3.2 User needs for active reading

From thematic analysis of the interview transcripts, the knowledge construction process in narrative reading can be categorized into several cognitive activities:

Theme Identification: Readers explore the underlying themes or messages of the narrative, which requires integrating personal experiences and broader world knowledge with the textual information.

Inference Making: Readers make inferences about the causes of events, characters' motivations, and possible outcomes, which are not always explicitly stated in the text.

Imagining the Scene: Readers visualize the setting and characters, which helps in deeper engagement and memory retention.

Emotional Engagement: Readers often become emotionally involved with characters and situations, which can affect their interpretation and understanding of the narrative.

The yellow dots in Figure 5 represent the pain points that participants claimed to have in doing active reading activities. To identify the latent values, I adopted laddering techniques, by asking “why is this important to you” to the pain points. Besides, I also identified some latent needs hidden in the marks and mental model sketches they made. Therefore, the following 2 groups of user needs are found.

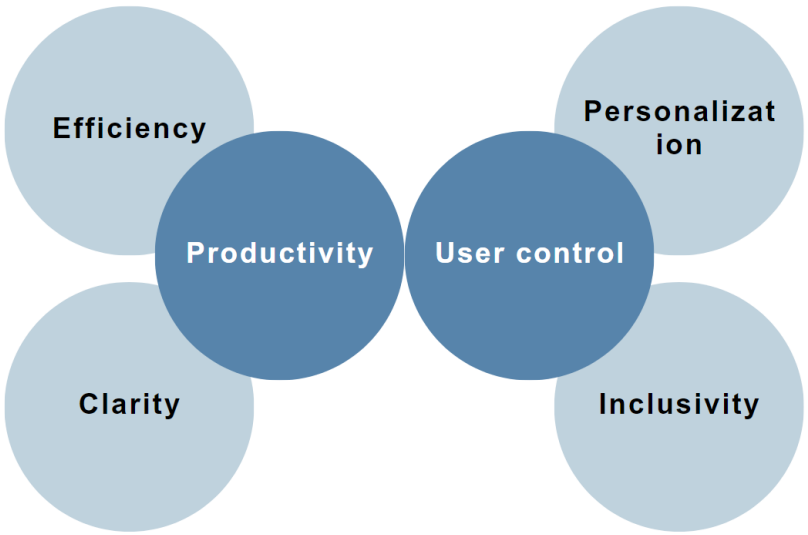


Figure 7: User needs for active reading experience identified from data analysis

Productivity

- Efficiency: In an immersive reading experience, inefficiency will disturb user engagement and cause information loss because of memory becoming blurry over time. Efficiency here means time-saving and effort-saving.
    - Search for answers: get answers without disrupting reading **(A)**
    - Make notes: take less time when making long notes **(B)**

    - Generate images: picture the image mentally quickly and get the appropriate image without too much effort. **(C)**
    - Generate diagram: Integrate massive information quickly **(D)**
- Clarity: One concept can be associated with numerous relevant concepts, it causes difficulty in identifying and organizing knowledge. Clarity here means selecting, simplifying and organizing relevant knowledge.
    - Search for answers: get clear and readable answers **(A)**
    - Summary: integrate complex information in the long story **(A)**
    - Navigation, reflection: link to desired referred contents
    - Generate diagram, layout: simplify/clarify the diagram/layout containing complex information units **(D)**

User-control

- Personalization: people have different preferences and understandings based on the same information, while current digital reading materials only offer fixed contents. Personalization here means generating contents from human input.
    - Generate images: integrate human input **(C)**
    - Generate layout / share: generate output in multiple formats **(E)**
    - Customized reading experience to facilitate personal active reading process **(G)**
- Inclusivity: People have different cognitive and mental imagery abilities, which influences their mental model construction results. Inclusivity helps boost active reading for people at different levels.
    - Mentally picture scenarios / sketch out mental model: support people with different mental imagery abilities and visualization abilities **(F)**
    - Associate relevant concepts: enrich alien concepts **(A)**

The summarized user needs for active reading are numbered from A to F, which will be addressed in the design round 2 (Chapter 8.3).

5.3.3 Elements of Mental model

The 9 participants were asked to sketch out their mental models or describe their mental models to the researcher(me) who helped sketching out.

Generally, look into their mental models sketching, the rephrased text paired with images in a structure constitutes the main body of their mental models. The rephrased texts include key slots in their memory, summaries, self-reflections, associated personal experiences etc. The images include specific scenarios depicted in articles (Figure 8-a), character interactions (Figure 8-b), key

objects/characters through the story(Figure 8-c), graphics of personal experience(Figure 8-d), etc. The structures include timeline, logical analysis, cause and effect, character interaction or a combination. Only one of the 9 participants has only text info in his mental model.

a.
Trees in front of thehouse

b.
The man thought: I'm not depressed anymore

c.
The magic box

d.
I used to have fun with my dad in our yard

Figure 8: The different contents of the images in participants' mental model

By deconstructing all the mental model sketches, the following elements (Figure 9) are detected: text, dialogue, rhythm curve, image, arrow and information level. Among them, text, dialogue, image and arrow are normally served as components which are organized on purpose in the rhythm curve or logical lines shaped by arrows. And different groups of curves or lines are divided into different information levels. For example, in the mental model sketch from participant 3, level 1 is the storyline shaped with scenario images and keywords of specific slots, level 2 is the summary of the storyline, with the value proposition deducted from the summary. In level 3, he associated related concepts from the overall knowledge from level 1 and 2, which links to mental images of his own experience.

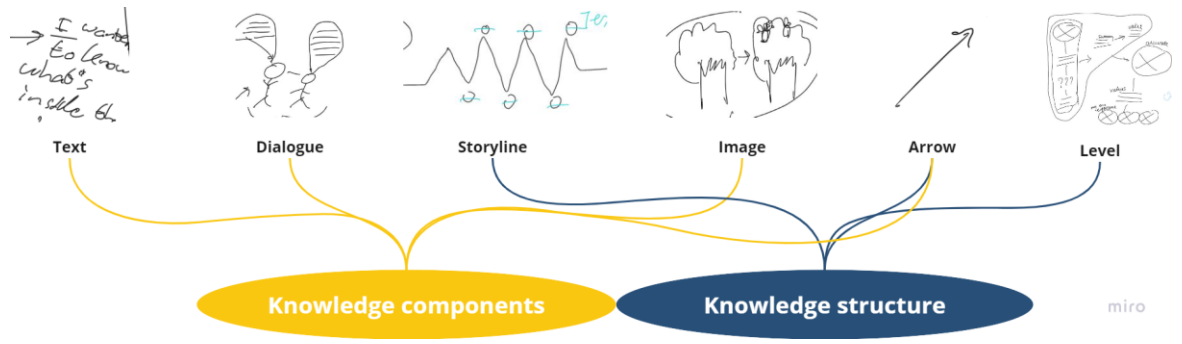


Figure 9: The categorized components of participants' mental model

As the existing research demonstrated, mental models are described as internal representations of external realities. They are constructed based on an individual's unique experiences and understandings (Jones et al., 2011). Mental models represent a logical arrangement of knowledge components in knowledge structures. Knowledge structure is a form of schema such as those that learners use to represent knowledge in memory, which can help learners manipulate knowledge components for various cognitive tasks, thereby forming the meta mental models. A mental model is a schema plus cognitive processes for manipulating and modifying the knowledge stored in a schema. These models assist in the acquisition and problem-solving processes involving conceptual and causal networks (Merrill, 2000). Combined with the sketches done in user research, the mental model by knowledge construction is a combination of texts, diagrams and images.

5.4 Competitive product analysis

In the digital era, people will be increasingly comfortable with reading books on screen (Masson, 2023; Baron, 2015). Reading is currently undergoing significant changes due to the continued development of digital technologies and devices, making it necessary to continually study the cognitive effects and possibilities of implementing innovative technical tools (Daniel & Woody, 2013).

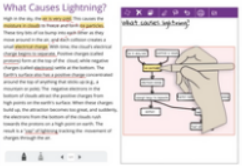
I researched the mainstream book readers and the emerging AI-powered book analyzers through website introduction and personal trial, to try and compare their functions and features in supporting active reading activities. The following 8 products are representative cases which illustrate rich functions.

5.4.1 Book readers

E-book reading platforms have been well developed across the world, such as KITABOO, Apple Books, Instapaper, Nook, and Kindle. By embracing touch screens and stylus, platforms like LiquidText and GoodNotes allow active readers to map out their thoughts on the book.

1. **texSketch** prototyped by Subramonyam et al. (2020) combines pen-and-ink interactions with natural language processing, offering an active reading experience in terms of diagram construction. It can visualize personal mental models through coherent diagrams (Subramonyam et al., 2020).
2. **LiquidText** is a computer-based active reading system, offering a flexible, fluid document representation built on multitouch input, with a range of interactions and gesture vocabulary designed to facilitate the activities, including highlighting, writing notes, mind-mapping etc.
3. **Kindle** is a series of e-readers designed and marketed by Amazon. Kindle devices enable users to browse, buy, download, and read e-books, newspapers, magazines, and other digital media via wireless networking to the Kindle Store. The primary appeal of Kindle devices lies in their long battery life and electronic paper (e-ink) display.
4. **Weixin Read** offers users access to a vast library of digital books, including novels, non-fiction, comics, and more. It emphasizes a social reading experience, allowing users to share their reading progress, recommendations, and reviews directly with their WeChat friends and on social media.

Table 1: Comparison of e-reader products' features that support active reading behavior

Product	LiquidText	texSketch	Kindle	Weixin Read
Active reading activities				
Set up reading task	Supports creating specific reading tasks	Not supported	Supports creating specific reading tasks	Supports creating specific reading tasks
Background information	Links to background info	Not supported	Provides X-Ray for additional info	Provides context for content
Annotation	Extensive annotation tools	Supports simple annotations	Allows highlighting and notes	Allows highlighting and notes
Content excerpt	Ability to pull excerpts and view in context	Ability to pull excerpts and view in context	Limited excepting capabilities	Allows content extraction
Navigation	Advanced navigation through documents	Easy navigation through content	Easy navigation through content	Structured reading paths
Summarizing	Not supported	Not supported	Not supported	Not supported
Image generation	Not supported	Not supported	Not supported	Not supported
Connection	Links annotations and excerpts across documents	Links annotations and excerpts across documents	Not supported	Not supported
Diagram generation	Not supported	Primary function	Not supported	Not supported
Question and Answer	Not supported	Not supported	Only supports word explanations	Only supports word explanations
Reflection guide	Not supported	Not supported	Not supported	Not supported
Free Layout	Flexible workspace for organizing content	Supports creative layouts	Not supported	Not supported
Share	Supports exporting and sharing documents	Supports sharing creations	Content can be shared with limitations	Allows sharing of highlights and notes

Most online book readers have been developed for knowledge gathering and comprehension, by providing basic functions of annotation, navigation, fixed Q&A etc. Some stylus-linked platforms allow users to diagram and make notes freely in a whiteboard to map out their mental model during reading, such as LiquidText and texSketch.

However, summarizing, image generation and reflection guides as part of fundamental activities in active reading experiences are missing in the mainstream online book readers. For Q&A, existing popular book readers can only show explanations or comments based on selected texts from the book, instead of allowing users to explore emerging questions in mind while reading.

5.4.2 AI-powered book analyzers

As AI technology is applied into all aspects of modern lifestyle, many emerging online reading platforms have embedded AI into their systems to efficiently analyze books, such as MyReader AI, iChatBook, StoryBooks, Pooks, ReadingClub, and StoryWiz. They are quite useful in generating summary and navigating information.

1. **Marlowe** from authors.ai is an AI tool to analyze scripts to provide detailed feedback on various elements of the book, including plot structure, pacing, character development, and more.
2. **MyReader AI** is an advanced tool that reads and dissects PDF, EPUB, and DOC files. It features question-answering, in-depth text search, cloud storage, and cross-file upload for seamless research and insight extraction.
3. **Emdash** is a software application that assists readers in comprehending and learning from challenging texts. It uses AI to analyze book passages and find conceptual connections between different authors and ideas.
4. **Oddbooks** is an AI tool that automatically extracts dialogue, characters, emotions, and scenarios from novels. It simplifies scenario creation and generates script ideas using AI-powered examples for derivative works.

Table 2: Comparison of book analyzer products' features that support active reading behavior

Product	Marlowe	Myreader	emdash	oddbooks
Active reading activities				
Set up reading task				
Background information	Not supported	Generate background info form prompts	Not supported	Not supported
Annotation				
Content excerpt	Not supported	Automatic excerption from prompts	Automatic excerption from highlights	Not supported
Navigation	Keyword navigation	Keyword navigation	Keyword navigation and association	Keyword navigation
Summarizing	Automatic summarization	Automatic summarization	Automatic summarization	Automatic summarization
Image generation	Not supported	Not supported	Not supported	Not supported
Connection	Not supported	Not supported	Connect asscoiated concepts by providing options	Not supported
Diagram generation	Automatic diagram analyzing generation	Not supported	Not supported	Automatic diagram analyzing generation
Question and Answer	Not supported	Generate answers form prompts	Not supported	Not supported
Reflection guide	Fixed reflection, without personal input	Generate reflection form prompts	Fixed reflection, without personal input	Fixed reflection, without personal input
Free Layout	Not supported	Not supported	Not supported	Not supported
Share	Supports exporting and sharing documents	Supports exporting and sharing documents	Supports exporting and sharing documents	Supports exporting and sharing documents

** Functions of setting up reading task and annotation are not applicable for AI book analyzer's function, because they are user behavior in the reading process.*

Most AI-powered book analyzing tools are capable of generating summaries of storyline, character interaction etc. They typically generate fixed narrative analysis reports based on fixed question templates and analysis methods. Some platforms embedded with ChatGPT can also make dialogues with readers to navigate keywords and answer questions instantly, which has shown their effect in reducing reading time, and integrating information.

However, image generation based on the human mind map has been significantly missing. And even though AI provides fruitful diagram analysis of book contents, the integration of personal input and preferences is missing too. It makes the current AI-powered Book readers are still dominated by the book and AI, which leads to passive reading experience as well.

5.4.3 Limitations of existing online reading platform

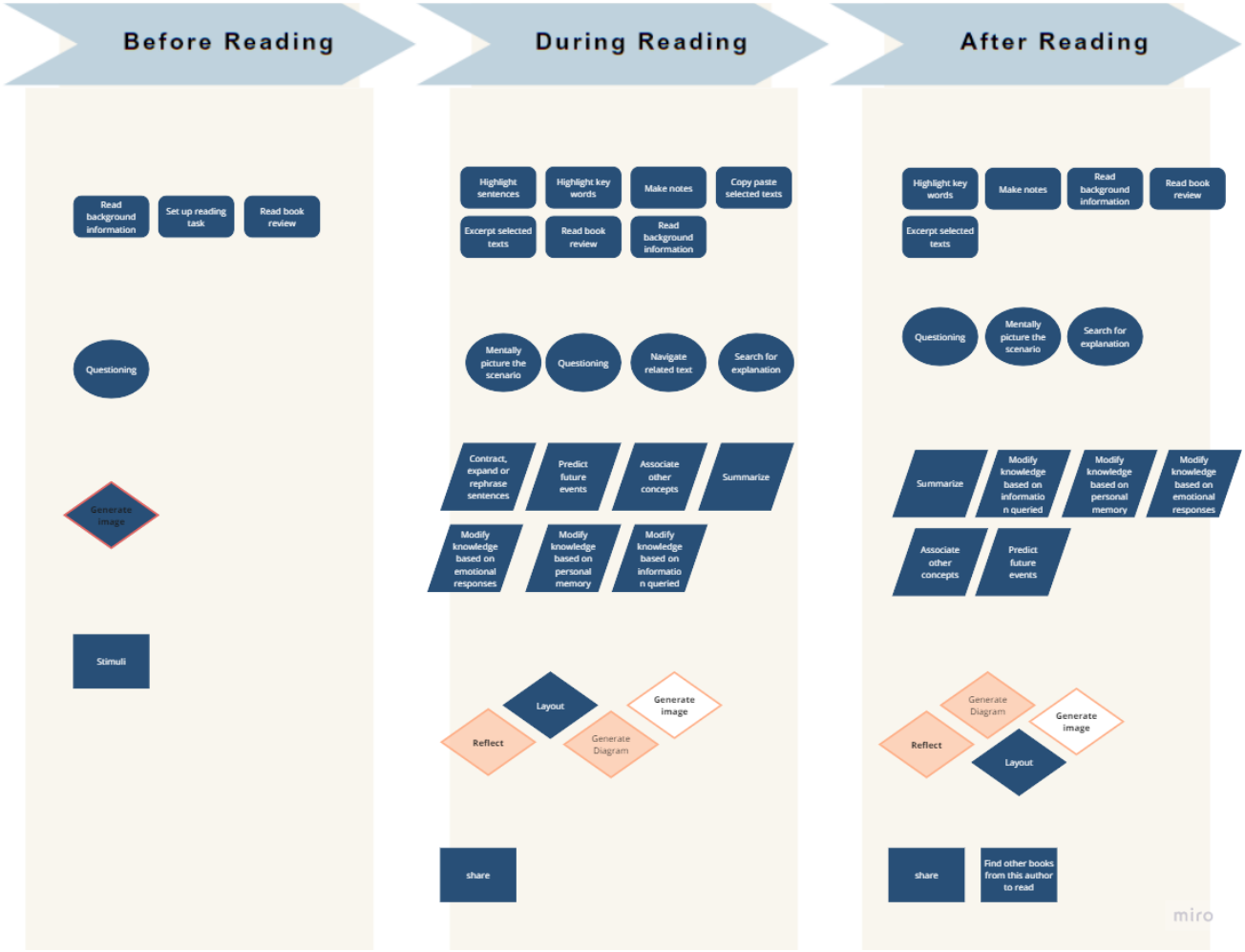


Figure 10: The incompatibility between users' active reading behavior and knowledge construction output of narrative books and existing e-reading products

The existing online book reading platforms can meet most requirements of the active readers, as shown on the figure above. However, the visualization functions and personalization in a toolkit level are rarely developed.

- **Lack of image content personalization (Design round 1)**

  Current online reading tools focus more on annotating and summarizing to reinforce knowledge construction (and reflection). However, for narrative books, storyline clarifications and scenario descriptions are more applicable than making text-based notes, which are also key user needs in the research outcomes. Neither book readers nor book analyzers well developed the visual functions, especially the image generation functions.

- **Limit Diagram and reflection generation (Design round 2)**

  Current online reading tools can provide a user workspace to allow active users making notes and sketching mind maps, based on the assumption that the user is already an experienced and active reader. However, AI tools are outputting data

based on the given book data and algorithms, without actively integrating human thinking to make the user dominate the knowledge construction.

- **Lack of toolkit personalization (Design round 2)**

  Current online reading tools have touched upon personalized reading experience in generating user-controlled contents by user input, e.g. Q&A powered by ChatGPT. However, they provided fixed functional features to support users' active reading, which may not be applicable for people with different reading processes. From previous user research, everyone's active reading and mental model construction process is different. Besides, the fixed function list is not applicable for people at different learning levels and diverse visual imagery abilities.

5.5 User journey in visualizing mental model

This is a fiction/persona and that it is built upon literature and mental model tests. The user journey is visualized to better illustrate the incompatibility between users' active reading behavior and output of narrative books and existing e-reading products.

Name: Timo

Age: 24

Book for actively constructing knowledge: Fiction with educational value

Tools for reading: book, laptop, notebook

This project will meet this gap between user needs and function offerings to design for visualized narrative reading experiences, using AI technology to facilitate efficient and personalized mental model visualization process of the narrative reading.

While Timo was immersed in reading novels, he constructed many scenes and images in his mind that were associated with the text descriptions.

When Timo recalled the knowledge model of the narrative book in his mind, he thought of images of key plots, logic of cause and effect, and reflections of his own experiences.

However, when he recorded his thoughts on the whiteboard, the knowledge model could only be presented in the form of words and stick figures.

Figure 11: User journey of actively reading narratives

Chapter 6

Technology Research

In this chapter:

6.1 The development of AIGC

In recent years, Artificial Intelligence Generated Content (AIGC) has gained much attention beyond the computer science community, where the whole society begins to be interested in the various content generation products built by large tech companies. AIGC refers to content that is generated using advanced Generative AI (GAI) techniques, which can automate the creation of large amounts of content in a short amount of time. This generation process usually consists of two steps: extracting intent information from human instructions and generating content according to the extracted intentions (Cao et al., 2023).

AIGC technology mainly involves two aspects: natural language processing (NLP) and AIGC generation algorithms. Natural language processing is a means of realizing how humans and computers interact through natural language. The mainstream AIGC generation algorithms include generative adversarial networks (GAN) and diffusion models.

Large datasets are used to train AI in generating accurate results. The larger sample datasets often result in more accurate models. Despite being trained on large-scale data, the AIGC may not always produce output that aligns with the user's intent. In order to better align AIGC output with human preferences, reinforcement learning from human feedback (RLHF) has been applied (Cao et al., 2023).

Jumping into the Web 3.0 era, artificial intelligence, linked data and semantic network construction form a comprehensive link between humans and machines. Content consumption demand is growing rapidly. AIGC will surpass professionally generated content (PGC) and user-generated content (UGC) and become the content generation tool of the future. Not only can it help improve the efficiency of content generation, but it can also increase the diversity of content. (Wu et al., 2023)

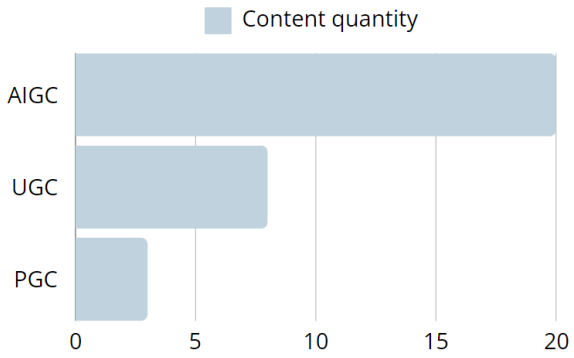


Figure 12: The content quantity produced under different content generation modes

AIGC business models are emerging in text, images, audio, games and code generation. Through research on dozens of AIGC-related companies at home and abroad, applications have gradually matured, especially in fields with highly repetitive tasks and low accuracy requirements. At present, image production and text generation are more common, and most of these AIGC services are monetized in the form of providing SaaS services.

Table 3: The AIGC products and major technology companies (Wu et al., 2023)

Company	Product	Applications
OpenAI	ChatGPT	Text generation, chatbots, and text completion
Google	LaMDA	Question answering and chatbots
NVIDIA	StyleGAN	Image generation, art, and design
Microsoft	Turing-NLG	Summarization, translation, and question answering
DeepMind	DVD-GAN	Video generation
Stability.AI	Stable Diffusion	Text to images
EleutherAI	GPT-Neo	Text generation
Baidu	ERNIE	Question answering and chatbots

6.1.1 Opportunities of AIGC in narrative reading

Artificial Intelligence in Generative Content (AIGC) technologies can significantly enhance the digital reading experience by generating images and diagrams that are contextually relevant to the text.

Contextual Imagery: AIGC can analyze text and automatically generate images that match the content, providing visual support for complex ideas or narratives.

Adaptive Content: Images can adapt to the reader's comprehension level or personal preferences, changing in complexity or style to suit individual needs.

Diagram Creation: For logical contents, AIGC can create diagrams, charts, and graphs that help explain storyline or concepts, making abstract information easier to comprehend.

Dynamic Updates: Diagrams and images can be updated in real-time based on new data or evolving stories, ensuring that the reader always has access to the most current information.

Concept Linking: Images can include embedded links to additional resources or related texts.

6.2 Principles for design with AIGC

Designers and developers may leverage the full potential of AI to increase their creative powers by understanding and utilizing these ideas. Generative AI is more than simply a tool; it is a collaborator in the creative process, opening new channels for creativity and expression. As technology advances, the AI future of creativity is only limited by our imagination and our ability to properly apply these concepts. The Generative AI Design principles give a framework for traversing the creative border where artificial intelligence and human creativity interact (Leslie, 2019; Leonhard, 2016).

Principle 1: Data-Driven Creativity

Generative AI relies heavily on vast amounts of data. The quality and diversity of this training data directly influence the quality and relevance of the output, whether it be text, images, or music. Thus, curating, cleansing, and organizing data are essential tasks in designing Generative AI.

Principle 2: Creative Constraints and Guidelines

While AI can generate diverse content, it is important to set creative limits and guidelines. These criteria help define the AI's creative boundaries and align its output with human objectives. For instance, in graphic design, specifying color palettes, typographic rules, or composition principles ensures that AI-generated material complements human creativity rather than replacing it.

Principle 3: Ethical Considerations

Designers and developers must ensure that AI-generated content adheres to ethical standards and social values. This includes addressing biases in training data, preventing the creation of harmful or offensive content, and being transparent about the AI's role in the creative process.

Principle 4: Collaboration Between Humans and AI

Generative AI is most effective when working alongside human creators. This principal views AI as a tool to enhance rather than replace human creativity. Designers, authors, and artists can use AI for inspiration, as a creative assistant, or to automate repetitive tasks. The synergy between human intuition and AI capabilities often yields remarkable results.

Principle 5: Explainable AI

As AI systems become more sophisticated, the need for explainable AI grows. This principal advocates for transparency in AI decision-making processes. Understanding how and why AI generates specific content allows designers and users to trust and control the AI. Explainable AI enables designers to fine-tune outputs, correct errors, and make informed creative decisions.

Principle 6: Continuous Learning and Adaptation

Generative AI is dynamic and evolves with experience. AI systems should be designed to adapt to changing trends, user feedback, and emerging creative needs. Regular updates and improvements ensure that AI remains a powerful and relevant creative tool.

Principle 7: User-Centric Design

Generative AI should be designed with the users' needs and preferences in mind. A user-centric approach ensures that AI-generated content resonates with the intended audience.

Chapter 7

Design round 1: Facilitate personalized image generation function

In this chapter:

7.1 Design task

Design round 1 focused on the main function gap of the active reading process in facilitating user-desired mental model construction, which is the mental image generation function. This design round aims to build up content personalization in an active reading experience. Therefore, the design task is:

A digital service powered by AI image-generating tools to visualize mental images in Narrative books based on human inputs.

It provides editorial guidance to the narrative readers drawing on knowledge construction done by AI supported information visualization to generate storyline and images with preferred layouts, helping each narrative's message be communicated most effectively (Hohman et al, 2020) and each reading experience be personalized to people with different cognitive experiences and visual abilities.

7.2 Existing AI Image generating products research

7.2.1 Mainstream image generation products

With the rapid development of deep learning, AI image processing has made great progress. The Generative Adversarial Networks (GANs) have achieved impressive results in image processing, which has been applied to various research and applications, such as image generation, image inpainting, text generation, medical image processing, semantic segmentation, image colorization, image-to-image translation, and art generation (Wang et al., 2023). In the AI-generated image category, the following 3 applications are popular in recent years.

1. **DALL·E** is an artificial intelligence program developed by OpenAI that generates images from textual descriptions. This technology builds upon advancements in deep learning and neural network architectures. DALL·E's underlying technology involves training a model on a vast dataset of images and their associated textual descriptions. The model learns to understand the content and context of these descriptions and how they relate to the visual elements of the images.

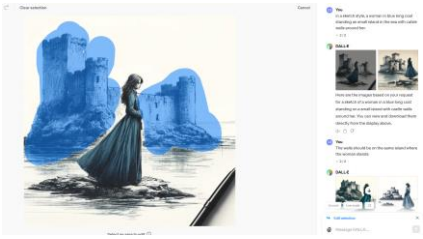


Figure 13: DALL·E's image generation and modification interface

2. **Stable Diffusion** is a state-of-the-art machine learning model developed by CompVis Lab and Stability AI that generates images from textual descriptions. It's part of a broader category of AI known as generative models, which are capable of creating new content based on the patterns they've learned from their training data. Stable Diffusion, specifically, utilizes a technique called "diffusion," which starts with a pattern of random noise and gradually shapes this noise into an image through a series of steps

that refine the details and structure based on the given text input.

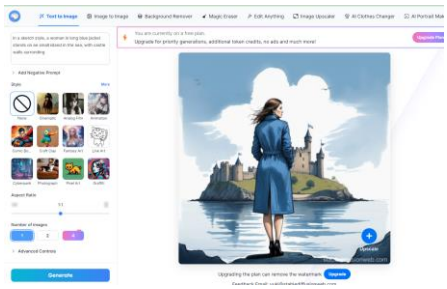


Figure 14: Stable Diffusion's image generation and modification interface

3. [Midjourney](#) is an AI research lab that has developed a namesake AI tool focused on generating images from textual descriptions and image references. The tool leverages advanced machine learning techniques, specifically deep learning models, to interpret textual inputs and convert them into visually compelling artwork and images.

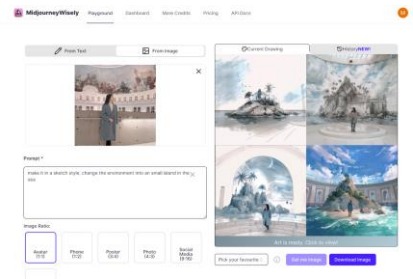


Figure 15: Midjourney's image generation and modification interface

Table 4: An overview of the functions of three mainstream AI image generation models: DALL-E, Midjourney, and Stable Diffusion (Versions before 2024.4）

Function	DALL-E	MidJourney	Stable Diffusion
Text-to-Image	Generates images from textual descriptions.	Generates artistic and abstract images from text prompts.	Generates images from text descriptions using diffusion models.
Image-to-Image	Capable of inpainting and outpointing from existing images.	Supports stylizing or transforming existing images.	Capable of inpainting and outpointing from existing images.
Styles & Aesthetics	Supports various artistic and photorealistic styles.	Focuses on high-quality, often surrealistic art styles.	Offers diverse styles, including photo realistic and artistic.
Control	Limited guide for style options, high degree of control.	Limited direct control but offers guided style options.	High degree of control on guided style options.
Follow up editing	Provides partial editing on selected areas.	No editing after generation finished.	No editing after generation finished.
Commercial Use	Limited commercial use allowed under certain conditions.	Allows commercial use with different subscription tiers.	Open source and free for commercial use with proper attribution.
Model Availability	Closed access, available through OpenAI API.	Subscription-based with limited access for trials.	Open-source, widely accessible for anyone to use.
Community & Support	Supported by OpenAI with a large community and resources.	Active community, especially strong in creative fields.	Large and active open-source community with extensive support.
Customization	Limited customization options.	Limited customization beyond style control.	Highly customizable due to open-source nature.

7.2.2 Extended features based on Generative AI on image

1. **ControlNet** (Zhang et al.2023) supported by Stable Diffusion significantly reduces the non-deterministic outcome of generative models by allowing users to provide a sketched outline as an input to image generation.

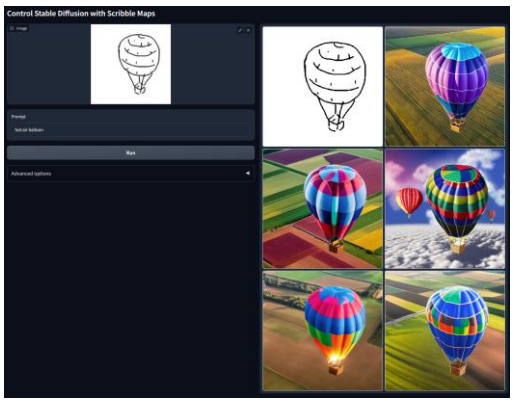


Figure 16: ControlNet's image generation and modification interface

2. Research team from Osaka University, invented a new method based on a diffusion model (DM) to **reconstruct images from human brain activity** obtained via functional magnetic resonance imaging (fMRI). Specifically, the researchers mapped brain regions as input to image and text encoders. The lower brain areas are mapped to image encoders, and the upper brain areas are mapped to text encoders. This allows the system to use image composition and semantic content for reconstruction (Takagi & Nishimoto, 2023).

   Similar research has also been done by Ferrante et al. (2023). Their model utilizes fMRI measurements to extract features for GIT captioning and VDVAE initial and depth image estimation. They proposed a method to generate image captions from brain activity using a multimodal large language model and introduce a novel image reconstruction pipeline based on predicted text and estimated initial and depth maps from brain activity.

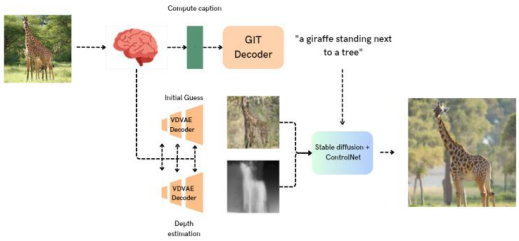


Figure 17: Workflow of reconstructing image from human brain (Ferrante et al.2023)

3. **Storybooks** is an AI-powered tool that allows users to easily create personalized stories in multiple languages. It can customize storylines and illustrations to craft truly unique stories based on human input. By uploading pictures, it can even bring family members and pets into the user's story adventures. Similar platforms like Magic Tales, CreateBookAI, StoryWiz etc make imagination become actual children's books with AI-generated contents and illustrations.

Figure 18: an example generated by Storybooks

7.2.3 Platform performance test

With the function analysis of mainstream image generation models, DALL.E and Stable Diffusion are two comparable options with controlling output through user input, following

up editing and platform customization. I further tested the platform performance with the same text and image prompts. In the section of sketch to image, ControlNet was tested since it's an extended platform based on Stable Diffusion, which specializes in following sketch prompts.

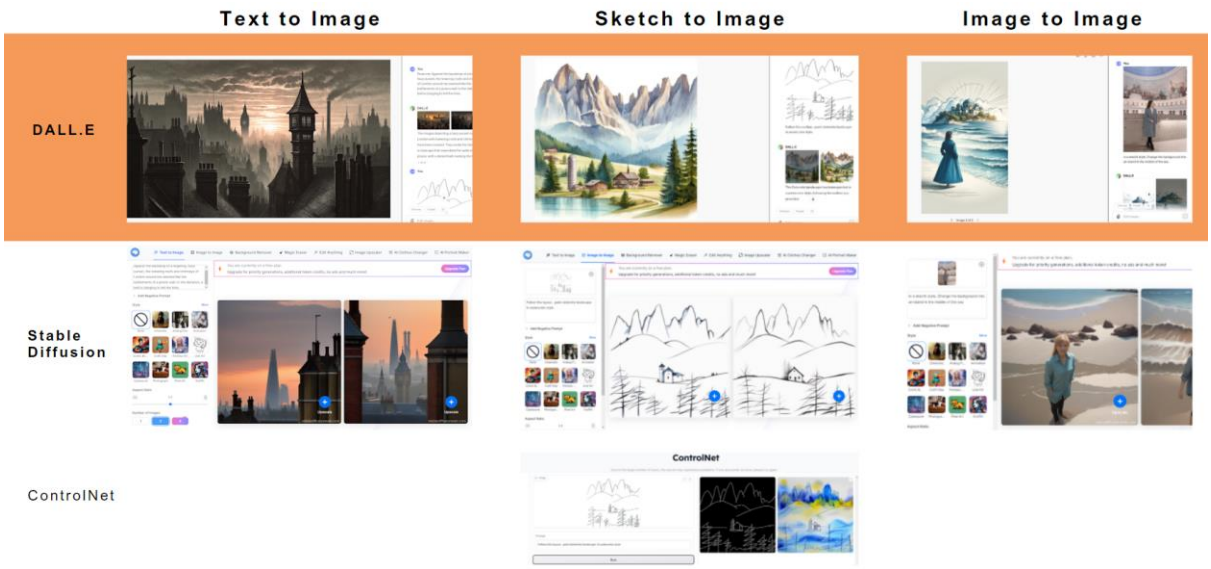


Figure 19: Performance comparison of DALL.E, Stable Diffusion and ControlNet

In practice, DALL.E demonstrates comparable performance with Stable Diffusion in generating images from text, while has a much better performance in generating images from sketch and images, compared to Stable Diffusion and ControlNet. In addition, DALL.E also has the function of selectively modifying generated pictures.

7.2.4 Conclusion

The quality of AI output will seriously affect the user's satisfaction with my product design. Therefore, it is particularly important to choose a high-efficiency /high-precision /high-controllability AI model as a tool to implement this design.In AI-generated images, ChatGPT embedded with DALL.E is considered to be the ideal platform to help visual mental models for reading narratives.

7.3 Pain points on image generation

To design for content personalization, I further researched existing study on human behaviors about mental picturing, and did user interviews with 3 Master students majoring in Design, all of them are familiar with using DALL.E to generate images in mind. Although Generative-AI can significantly improve the efficiency of visualizing contents, there are still significant concerns about human behaviors and technological factors that cause negative effects. The following insights are integrated from literature review, first user research (Chapter 3.4) and further user interviews on AI visualizing.

7.3.1 User behavior of mental image

1. Deliberate visualization creates disturbance

During the process of visualizing the mental model, when imagining text content, learners must mentally transform verbal information into pictorial information which fosters deeper processing. Extraneous cognitive load refers to the cognitive load caused by the extraneous environment that is ineffective for learning. For learning to be maximized it's essential to reduce extraneous cognitive load (Phelan, 2020). When visualizing pictures of text content on screen, however, learners have to mentally transform pictorial information into verbal information and evaluate the pictorial information, which is expected to require additional cognitive resources. Leutner et al. (2009) experimented with 111 students to research this phenomenon. The results show that **constructing mental images seems to optimize cognitive load and to increase comprehension and learning outcome when the mental visualization processes are not disturbed by externally drawing pictures on paper or by AI,** whereas generating pictures seems to increase extraneous cognitive load resulting in reduced learning outcome.

2. Ambiguity of mental image

And yet, visualization is an elusive phenomenon, **for the most part not even retrievable by the individual reader herself.** It emerges from an interaction between a particular text and the reader and a particular situation and context for both. This triangulation is difficult, perhaps impossible to disentangle. Moreover, **most of what is imagined during an absorbed reading of a fictional narrative hovers on the threshold of consciousness and cannot be remembered in total once the reading is completed.** It is a necessary feature of prototype images to be vague. As established on the basis of empirical studies, indeterminacy, uncertainty, blanks and gaps are an ordinary feature of default visualization and do not so much challenge readers to continuous 'filling-in' as leave them leeway to establish 'good continuation' (Brosch, 2018).

3. Impact of Speed on Mental Imagery Vividness

In empirical studies, an important factor for vividness in mental imagery is the speed of emergence. Amedeo D'Angiulli et al (2013) conducted elaborate tests relating vividness in mental imagery to latency. The results demonstrate that a slow generating process does not improve mental imagery and that faster image generation is accompanied by higher vividness. Applied to visualization, this must mean that complicated narrative and linguistic structures delay or hinder the production of visualizations (Brosch, 2018).

Participants also declare that continuously stacked scene descriptions will keep refreshing their mental images, which sometimes lead to a complex overview with low vividness. The vividness of mental images

often results from their existence or observation in actual scenes, which shaped their intuitions to some extent.

4. Variability in Mental Imagery Among Readers

In terms of the vividness of mental images, people vary greatly in their ability to create vivid mental images. Some readers can easily visualize detailed scenes and characters as they read, while others struggle to generate any visual representation. This variance can affect the level of engagement and understanding of the narrative, potentially leading to different experiences and levels of comprehension among readers (Mackey, 2019; Williams, 2023).

5. Emotional Effects by images

Intense or vivid visualizations can have strong emotional impacts, which can be overwhelming or distressing for some readers, particularly when narratives involve themes of violence, loss, or trauma. This can make reading a challenging experience that some may wish to avoid or manage (Brosch, 2017).

Participants also mentioned that realistic pictures are more visually impactful than imagery pictures, which can quickly amplify or change emotions. For example, when they turned the page and saw the illustration of the dark castle described in the previous text, the sense of predicament was suddenly magnified. But it was not uncomfortable if it didn't deviate too far from imagination.

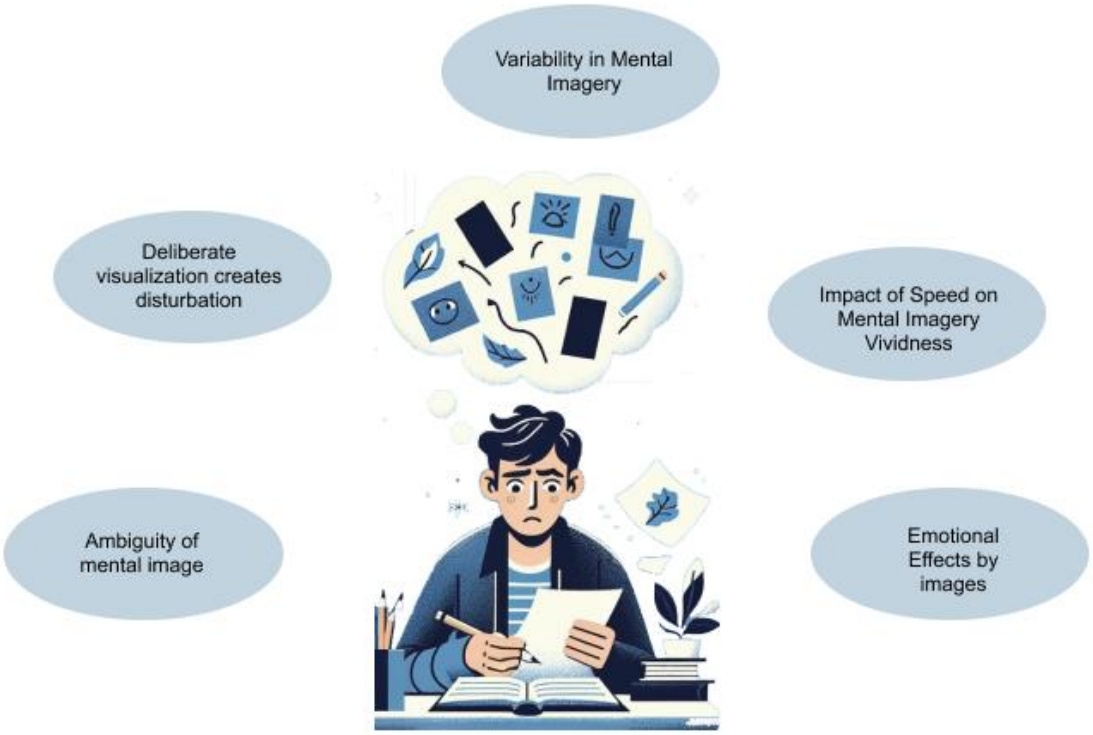


Figure 20: Pain Points of users visualizing mental image

7.3.2 AI operation of mental images

1. **Uncontrollable variables**. There are always uncontrollable variables showing up in the AI generated images, because it's not possible for users to limit every component. AI would automatically fill in the blank spaces that didn't be mentioned by the prompts. To try to control those emerging variables, the user has to elaborate more restriction wording and re-generate the images several times. In the research activity "Dreaming with AI", participants generated an average of 12.6 images (min: 4, max: 31) to make the image meet their satisfaction, which implies the participants have to keep tweaking the prompt and adding more details to get to images they desired. And only 26 out of 34 participants ended up with the images they felt satisfied with. All the finished prompt was longer than the initial prompt, by adding more details to limit the AI-generated image (Ali et al., 2024).

2. **Complexity of text-based describing**. Mental image generation is a fast and intuitive process that sometimes contains many vague or abstract elements. Users find it difficult to accurately describe all the contents of the mental image and its expression in words, especially the intuitive part. For AI to generate pictures that meet psychological expectations, users need to add adjectives to the input multiple times. Frequently adding enrichment or adjusting wording in generating desired mental image will cause user fatigue=, because this process is usually slow and time-consuming and requires a lot of energy to think about detailed elements.

3. **Primacy effect** is a psychological phenomenon. The initial impression formed by the first received information will form the core knowledge or main memory schema in the brain (Van Erkel et al., 2016). The first version generated will set up a foundation/impression in the user's mind. As more details are added and more versions are generated, the image may be twisted away from the first version. The more times they try, the more unsatisfied they feel.

4. For people with weak drawing techniques, it's **difficult to sketch out the pictures in mind** as an input for AI.

5. **Pre-established impression**. For users who have seen the actual scenario from actual experiences or relevant videos or illustrations before reading the book, they have a pre-set template in construction of mental images, which is derived from previous videos or illustrations. This pre-established impression highly dominates user's expectations from the AI-generated images in multiple levels, which is very important for training AI in order to generate the desired outcome.

6. Automation and AI decision-making can sometimes lead to **slowing down human creativities**. The use of AIGC can create a dependency on technology and discourage critical thinking and problem-solving skills (Wu et al., 2023), which is opposite to one of the design goals of promoting active reading. It's important to make sure the human-machine co-creation process powered by image-generating technology is highly controlled by

users, in order to enhance active reading and learning experience.

Figure 21: Pain Points of users using AI to visualize mental image

7.4 Design

7.4.1 Design method and process

This is done using the basic design process, through diverging and converging. Figure 22 depicts the following steps:

1. Put personalization at the core of this design phase, I arrange the pain points based on different levels of mental imagery abilities and experiences.
2. Deducting sub-functions of user requirements from concerns
3. Use the morphological chart to diverge, brainstorm 5-10 directions for each pain point.
4. Strategically integrate directions and synthesize with AIGC in achieving the solutions, generating several thematic solutions.
5. Evaluate solutions through AIGC principles.
6. Look back into sub-functions, prioritize solutions by Harris Profile, select 2 out of 7.
7. Prototyping

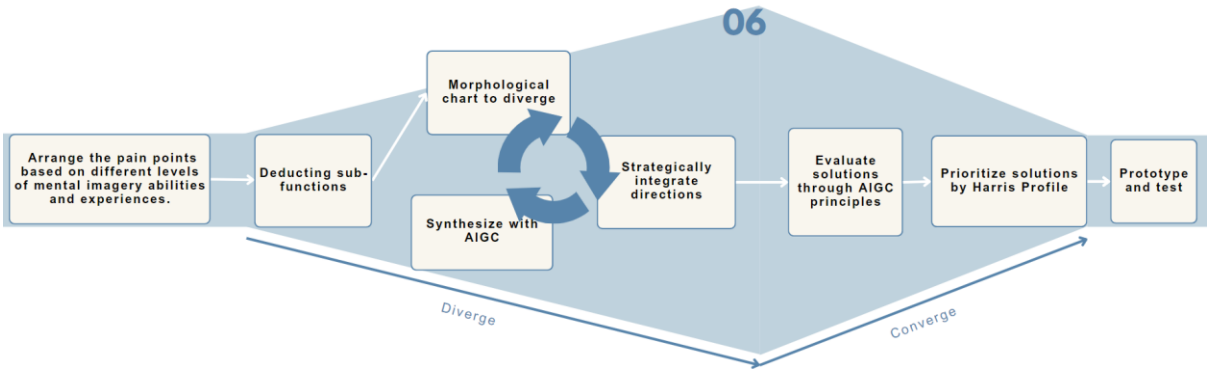


Figure 22: Design process

7.4.2 Deducting sub-functions

Put personalization at the core of this design phase, I arrange the pain points based on different levels of mental imagery abilities and experiences (Figure 23).

Readers with low visual imagery / less experiences means that they feel more alien to the narrative contents, it leads to the low vividness of their mental image. So, the ambiguity of mental image will concern this type of readers more. Without a strong feeling of the desired image, the AI-generated image output can more easily affect their emotions, therefore causing uncomforting and distribution in reading. Besides, the primacy effect happens more as they have no strict expectation of the output, whereas the first version takes up their cognition.

Readers with high visual imagery / more experiences means that they feel more familiar or creative to the narrative contents, it

leads to the high vividness of their mental image. This type of readers has more restrictions on the details of AI-generated images, a uncontrollable variables that are not requested concerns them more.

The rest of the user concerns regarding visualizing mental images by AI has the similar pain on the variable mental imagery types of readers

.

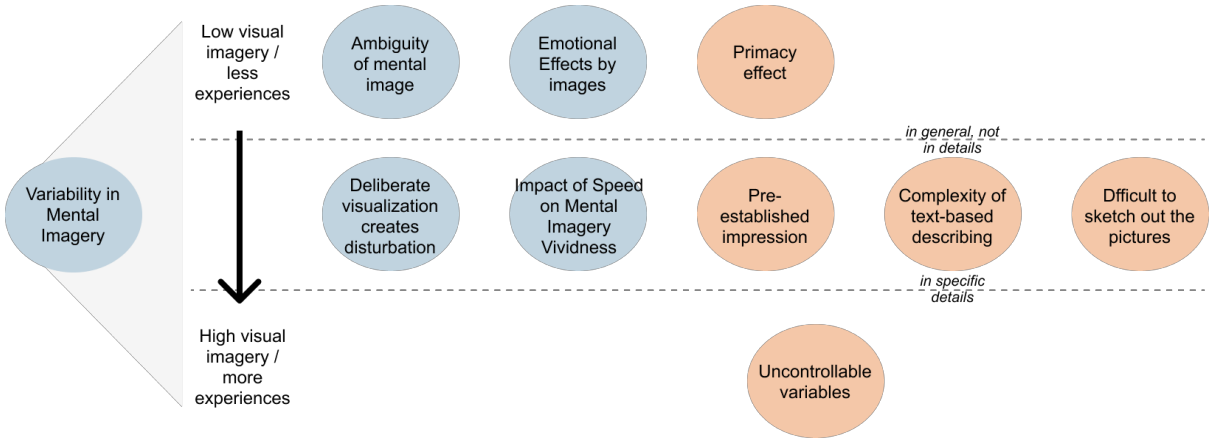


Figure 23: List pain points according to variability in mental imagery

In order to design human-centered solutions in using generative AI to visualize the mental model of narrative reading, minimizing the user's pain points is essential. The following design opportunities as sub-functions (Figure 24) are deducted from pain points by asking “What do users with such concerns need?”

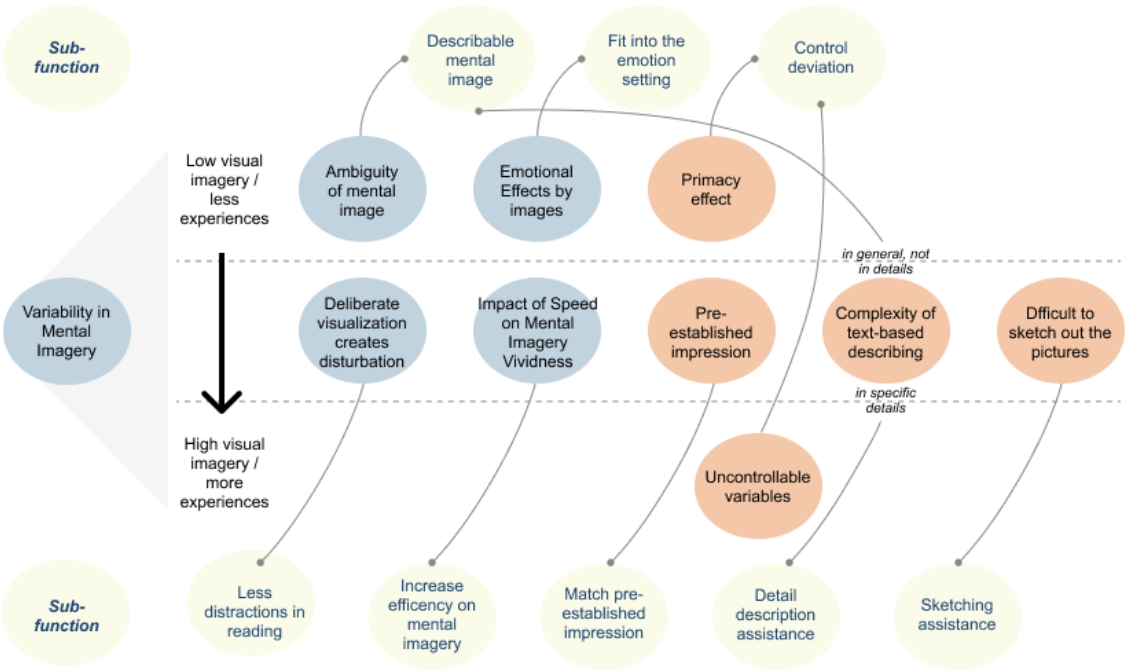


Figure 24: Deduct user needs (sub-functions) from user pain points

Describable mental image: assist individuals with limited cognitive or visual imagery abilities in concretizing describable mental images.

Fit into the emotion setting: align with human emotions and do not overwhelm or disturb readers.

Less distractions in reading: visualize mental images without disrupting the reading flow.
Increasing efficiency on mental imagery: efficiently capture and visualize vivid images from user intuition.

Sketching assistance: help users with weak sketching skills to effectively convey their mental images.
Control deviation: control the variables that users didn't mention in prompts and keep the variables that has been approved by users from deviating
Match pre-established impression: manage and potentially utilize these pre-existing impressions
Detail description assistance: minimize fatigue caused by frequent adjustments or enrichment inputs

7.4.3 Brainstorming with Morphological Chart

Design brainstorming is a diverging step, through the tool called Morphological Chart. The sub-functions are taken as a starting point. To guarantee all the sub-functions are recognized, Possible components are listed on the basis of their functions. The components are concrete and specific, specifying the elements that belong to a category. Functions are listed in columns, and components are the means that realize the functions and are listed in rows. The morphological method thus yields a matrix of functions and possible solutions. By means of the morphological chart, the components are

strategically categorized into sets of sub-functions, which stands for thematic solutions to potentially realize the overall target function.

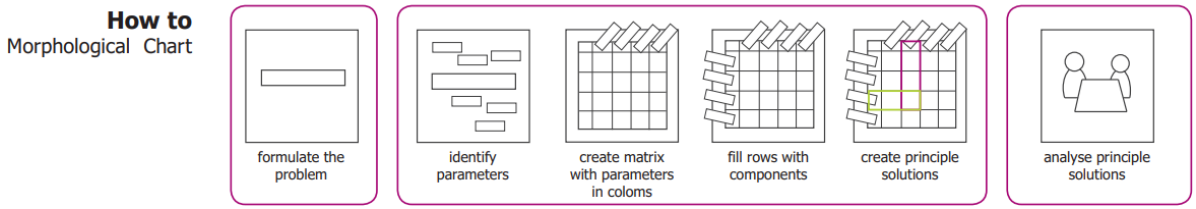


Figure 25: Workflow of Morphological chart (Van Boeijen et al., 2014)

The following figure is the Morphological chart for using Generative AI to visualize mental images when reading narratives. The sub-functions are listed according to the users with different visual imagery abilities and less relevant experiences from low to high. In each sub-function category, the solutions are only targeted at achieving that sub-function.

Function: Use Generative AI to visualize mental images when reading narratives

	Sub-functions	Solutions						
Low visual imagery / less experiences ↓	Describable mental image	Use the text provided and random personal thinking as pormpts for AI to generate description	Stimuli board linked to description guidance	Step by step instruction / generation	Work collaboratively with other readers to express or refine thoughts	Extend given texts into fruitful description	Role play to enhance immersion	Multi-sensory presentation to stimulate imagination
	Fit into the emotion setting	Sentiment analysis and sdjust the images based on the reader's real-time emotional responses	Emotional Filters pre-setted by users	Cater to emotion arc from understanding the context	Match narrative pace	Safe Viewing Modes which shows warnings	Learning from massive feedback	Ask for user's emotion before showing outputs
	Increase efficency on mental imagery	Speech to text	Summarize the long texts as visualization prompts	Maintain a buffer of pre-generated images based on upcoming text	Brain-Computer Interfaces	Sketch out overview before thinking through details	Present massive images to set up vividness	
	Less distractions in reading	Encourage users to do visualization after reading	Image generating is blended with the book	Predict images based on contexts and user settings	Focus Mode to control the time spending on image generation	Show several options at once	Asynchronous Review Mode	Fix elements set up before reading
	Sketching assistance	Stimuli board	Basic shape offering	Quickly convert text elements into graphics	Insert photos into sketch	Automatically correct random sketches	Take sketching trainings	
	Control deviation	Explainable AI in displaying every variables adding into the original prompts	Give refrences to prove the deviation is reasonable	Request users to input missing variables	Analyze narrative elements and provide options	Selected area modification	Don't show the elements thar are not in prompts	Use sketch as input
	Match pre-established impression	Pre-train AI with existing images	Request checklist from the user	track user activities on platform	Refresh their impression by presenting images before reading			
High visual imagery / more experiences	Detail description assistance	Predict descriptions and offer examples	Offer framework for user to fill in	Modify provided images instead of starting from scratch	User gives feedback in batches	Use detailed sketch as input	Explainable AI in displaying every variables	

Figure 26: Brainstorm for each sub-function

After brainstorming the solutions to a certain point, the similarities among solutions across categories, which formulated the thematic solutions.

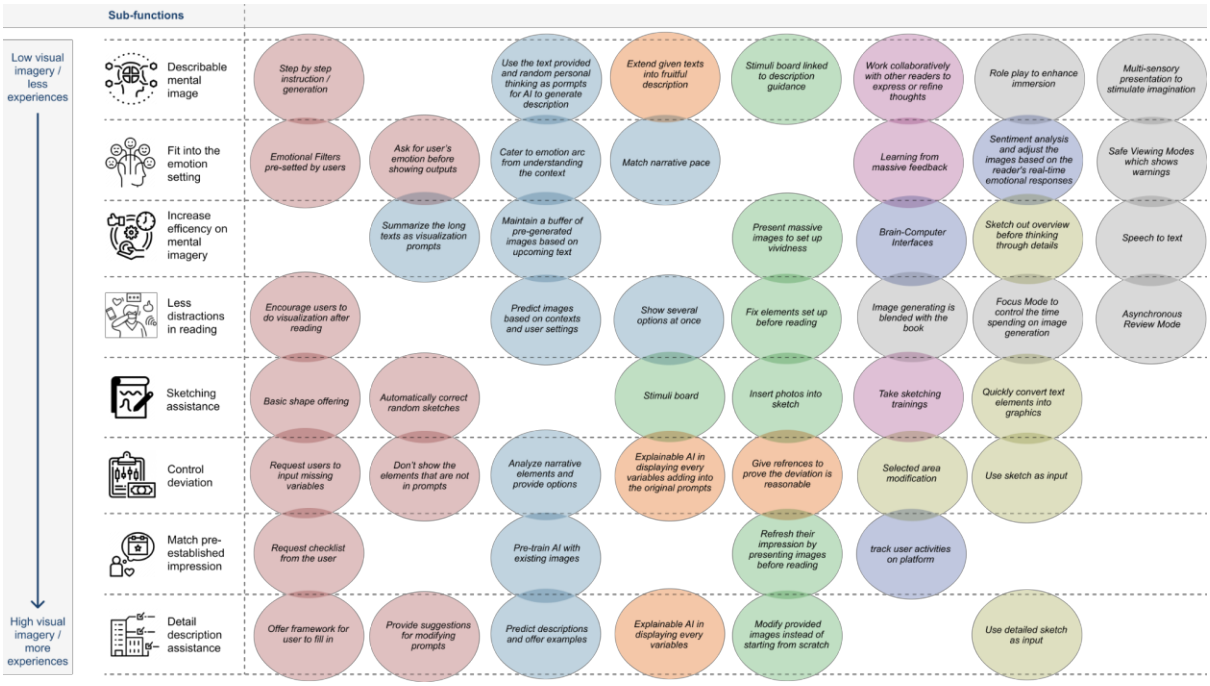


Figure 27: Categorized sub-solutions into thematic solutions

7.4.3.1 Thematic solutions

Instructive AI:

Instructive AI, which represents AI technologies designed to assist, guide, and enhance mental image visualizing processes. It provides all-round guidance and verification before presenting the AI generation input, which can significantly support users with low mental image abilities or less relevant experiences.

Predictive AI:

Predictive AI can leverage advanced machine learning models to anticipate and generate visual content that aligns closely with the unfolding story and user expectation. It maintains a buffer of pre-generated visual/textual aids by analyzing narrative contents, massive databases, and the user's emotion and preference, therefore increasing the mental image generation efficiency.

Explainable AI:

Explainable AI (XAI) refers to techniques in AI that provide human-understandable explanations of how the AI reached its decisions or outputs. It can provide detailed explanations on how certain aspects of the narrative prompted the AI to generate specific images, and allow users to seek further clarification and adjustment based on each variable. Explainable AI plays a critical role in enhancing the reader's understanding and trust in the AI-generated visual content.

Expectation establishment:

Expectation establishment means setting up expectation in advance before the user starts mentally picturing the narrative slot, to add default values as training materials into the AI image generation system. It can be done by presenting users with existing inspirations collected by AI / other readers / authors etc. to refresh the user memory, as well as get early feedback of existing inspirations as AI training material.

Reading community collaboration:

Collaborative community platforms enable readers to share insights, interpretations, and visualizations, fostering a collaborative filtering and recommendations. It can facilitate a channel for people with different mental imagery abilities and cognitive experiences to reach a balance by exchanging complementary skills and knowledge, which can improve the quality and relevance of AI-generated visualizations.

User behavior monitoring:

User behavior monitoring will closely observe and analyze how users interact with narratives and visual content, including actions, facial expressions, speaking and even feelings. AI systems can adapt and refine their processes to better meet individual preferences

Sketch-based generation:

Sketch-based generation, where AI interprets and transforms simple user-generated sketches into detailed visualizations, offers a unique and interactive way to enhance the visualization of mental images when reading narratives. Sketch-based AI systems can be designed to interpret very basic drawings and collaborate with AI correcting, making it accessible to people with different sketching levels.

Multiple reading modes

Integrating multiple reading modes into AI-driven systems for visualizing mental images while reading narratives can adapt to different user rituals and needs. The modes could include Focus Reading Mode, Audio-Visual Mode, Augmented Reality (AR) Mode, Interactive Mode, Asynchronous Review Mode, Collaboration Mode etc.

7.4.3.2 Evaluate with design principles for generative-AI

Generative AI design principal / *Thematic solution*	Data-Driven Creativity	Creative Constraints and Guidelines	Ethical Considerations	Collaboration Between Humans and AI	Explainable AI	Continuous Learning and Adaptation	User-Centric Design
Instructive AI							
Predictive AI							
Explainable AI							
Expectation establishment	⊗						
Reading community collaboration				⊗			
User behavior monitoring			⊗				
Sketch-based generation							
Multiple reading modes							

Figure 28: Evaluate thematic solutions based on AIGC design principles

The design principles for generative-AI (Chapter 6.2) offers basic considerations for designers to create responsible, user-centered and ethical AI. Incorporating design principles into the evaluation of thematic solutions ensures that these solutions are not only technically sound but also user-friendly, ethical, and innovative.

When checked against the design criteria, most of the thematic solutions met the general requirements. Three of them may cause side effects in whole or in part.

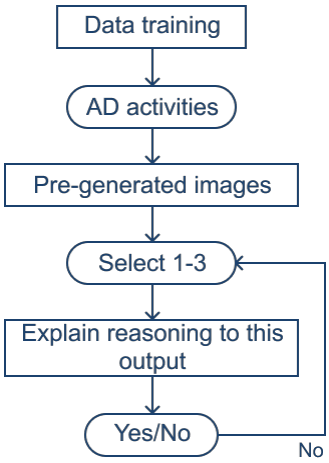
1. Expectation establishment can weaken the data-driven creativity since it will influence users' memory and perception, making the mental image of the user's mental image be manipulated by the pre-displayed pictures. Although this method can achieve strong correlation between AI-generated images and user expectations, AI cannot empower human thinking's individuality by generating a wide variety of images based on personal experiences and understandings.
2. Reading community collaboration can weaken the collaboration between human and AI, since human collaboration may dominate the visual reading experience when the creativity is mainly from human interaction and judgements. AI systems excel in processing large volumes of data and identifying patterns that might be invisible to humans.
3. User behavior monitoring can significantly weaken the ethical AI by damaging trust, data privacy and safety. Even though we can obtain genuine informed consent from users for behavior monitoring, users won't fully read the content or still feel uncomfortable and safe when being monitored.

Based on the considerations above, there are several rules that need to be addressed. First of all, the influence and establishment of user expectations must be moderate. While helping users improve the clarity of visual imagery, it cannot interfere with or control the user's subjective initiative. Give full play to the big data advantages of AI to help users diversify infinitely within a certain creative framework. Second, striking a balance where human input enhances rather than hinders AI performance is crucial. This requires careful management of the collaboration process, clear alignment of goals, and ensuring that the AI maintains its strengths in data analysis and pattern recognition while incorporating meaningful human insights. Thirdly, to maintain ethical integrity, AI systems must prioritize user rights, transparency, fairness, and autonomy over the extensive monitoring of user behavior.

7.4.3.3 Brainstorm Solutions by combining thematic solutions

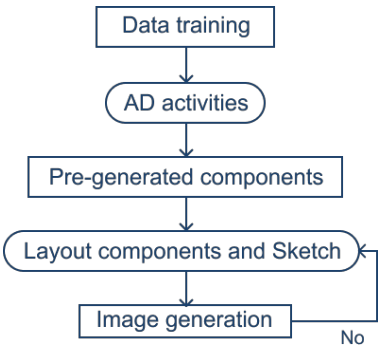
After carefully analyzing and evaluation, in the further brainstorming session, the thematic solutions can be used as foundational ideas of system building. By strategically arranging the thematic solutions in a workflow, the simplified system flow for visualizing mental image when reading narratives can be elaborated. The following 7 are reasonable directions to be continued after the brainstorm session.

1. Predictive AI + Explainable AI



First, after the user selects a book, train the AI with massive data on the existing visual materials about this book. The visual materials for training the AI can include book covers, illustrations, fan art, movie adaptations, and related imagery. These materials can be sourced from various platforms such as Google Images, Pinterest, DeviantArt, and fan forums. When the user does the active reading activities, maintain a buffer of pre-generated images. In the visualization stage, the user will select 1-3 pictures from the images pool, with detailed generation logic explanation displaying. The users can choose if he/she is confirmed by the explanation.

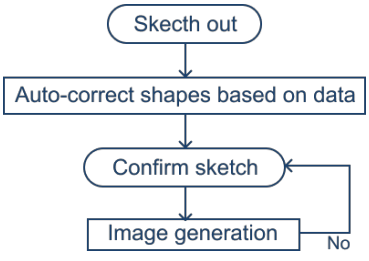
2. Predictive AI + Sketch-based generation



First, after the user selects a book, train the AI with massive data on the existing visual materials about this book. When the user does the active reading activities, maintain a buffer of pre-generated visual components dismantled from the visual data.

To achieve this, the AI uses techniques like image recognition and segmentation to dismantle visual data into components. Pre-trained models, such as those based on convolutional neural networks (CNNs) and generative adversarial networks (GANs), analyze and break down images into distinct elements. These components are stored in a buffer for quick retrieval during active reading. In the visualization stage, the user can draw the sketch of their mental image, they can also play with the components by adding them into his/her sketch. AI will use the final sketch as a prompt to generate an image.

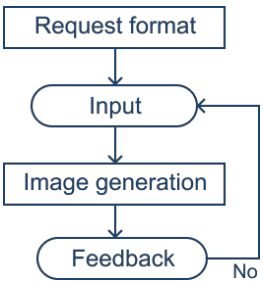
3. Sketch-based generation + Predictive AI



In the visualization stage, the user will sketch out the mental image, AI will automatically correct the shapes based on data and analyze to assist people with weak sketching ability. The user has to confirm the final sketch in order to put it into AI generation

In this case, the training data includes a large dataset of hand-drawn sketches paired with their corrected versions, sourced from online drawing platforms, art tutorials, and user-generated content. Additionally, the dataset incorporates various shapes, objects, and scenes to enhance the AI's ability to recognize and correct diverse elements.

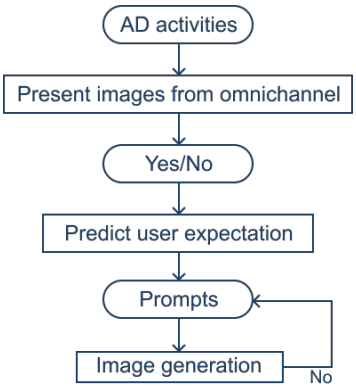
4. Instructive AI



Before generating an image, AI will ask the user to describe every variable of their mental image into a template. AI cannot display any output if any variables are missing. Then, the user will give feedback or changes in the template to further generation.

In this case, the training data consists of a diverse set of annotated images and corresponding descriptive templates detailing various visual attributes. This data helps the AI understand how different descriptions map to visual elements. It includes information on object types, colors, shapes, spatial relationships, and other contextual details, enabling the AI to generate accurate images based on user-provided descriptions.

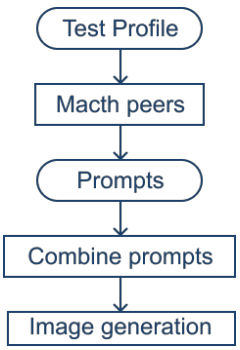
5. Expectation establishment + Predictive AI



When the user does the active reading activities, AI will detect the selected texts and present relevant images from omnichannel. The user needs to choose Yes/No for each image, through which, AI can predict user preference and expectation. In the visualization stage, AI will base on the prompts and predicted user expectation to generate images.

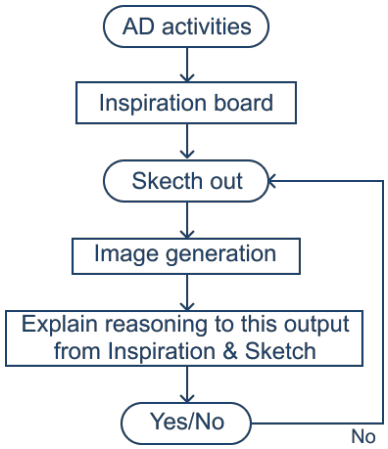
To realize this, the technology involves a combination of natural language processing (NLP) and image retrieval systems. NLP algorithms analyze and understand the selected texts to identify key concepts and themes. These insights are then used to query a vast database of images from various sources, known as omnichannel, to find and present relevant visuals.

6. Instructive AI + Community collaboration



Before the user starts reading, the system will ask him/her to do a test to establish a reader profile. With a bunch of profiles, Machine learning models then analyze this data and match users with similar profiles based on factors such as reading interests, styles, and preferences. AI will match the user with other peers to co-read a book. In the visualization stage, AI will combine the prompts from the user and his/her peers, as final prompts to generate an image.

7. Expectation establishment + Sketch-based generation + Explainable AI



During the active reading activities, AI will generate an inspiration board with existing images based on user-selected texts. In the visualization stage, the user will sketch out his/her mental image by referring to the inspiration board. After generating an image based on the final sketch, AI will explain reasoning to this output from inspiration and sketch. The user can further edit the output by modifying the variables in the explanation.

For reasoning, the system employs interpretability techniques such as attention maps and feature importance analysis to trace how specific elements from the inspiration board and the sketch influenced the generated image, providing a clear, logical explanation of the design choices.

7.4.4 Concept converging with Harris profile and KANO Model

After the diverging process, the 7 concepts need converging and filtering. Whenever several alternatives of product concepts need to be compared and consensus/an intuitive decision needs to be reached/made, the Harris Profile can be used.

The Harris Profile is a graphic representation of the strengths and weaknesses of design concepts. Originally, one product profile is applied as a base to evaluate and compare other alternatives. Criteria are set up to be applicable to the alternatives on the specific level of development. The designer will interpret the meaning of the scale positions (i.e. - 2 = bad, - 1 = moderate, etc.) to select several concepts out of the seven.

The criteria I set up are sub-functions to achieve in using Generative AI to visualize mental image. Since there is no existing product specifically focusing on visualizing mental image in narrative reading, I combined Harris profile and the Pugh method by taking the first idea (Predictive AI + Explainable AI) as the baseline, the other 6 concepts will be graded by comparing with it. As it is shown in figure 29, all the alternatives have been evaluated.

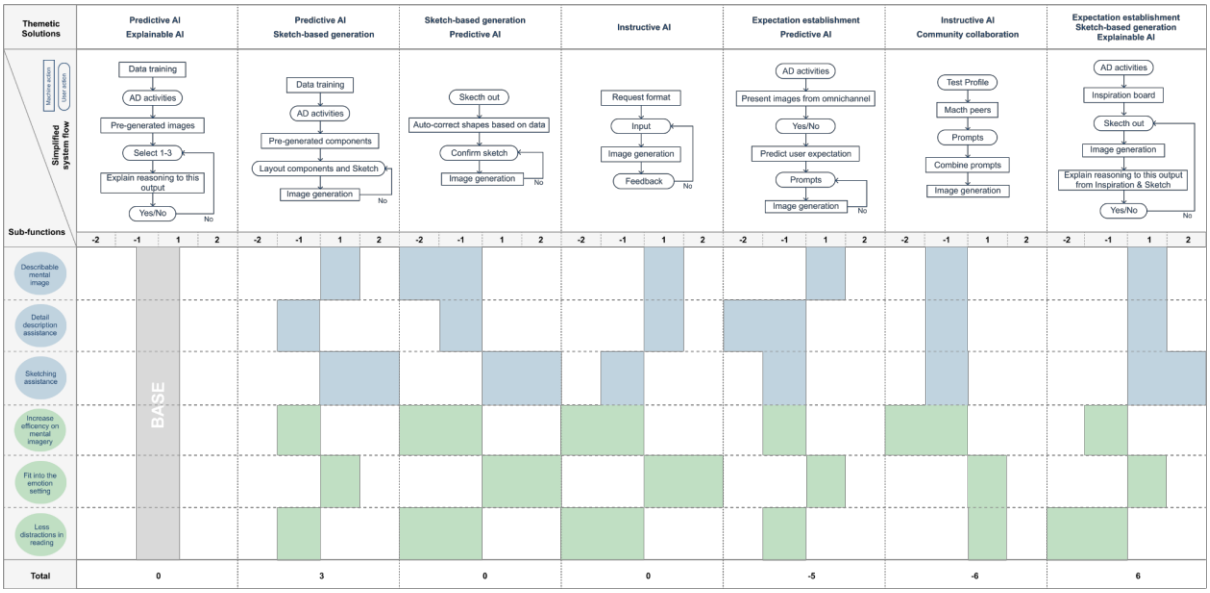


Figure 29: Evaluate concepts with sub-functions

The alternative concepts show different weaknesses and strengths when compared to the base concept. For example, in terms of describable mental image function, Sketch-based generation + Predictive AI gets a -2 when compared with Predictive AI + Explainable AI, because it doesn't provide any detail description or cues leads to vivid mental image, while Explainable AI can describe the details of an image to the user to the greatest extent to help users build thinking about image details. In terms of less distraction in reading, Instructive AI gets a -2 when compared with Predictive AI + Explainable AI, because it highly relies on user input before generating output, which requires users a lot of time to think and fill in.

To calculate the total score, the sub-functions should be added weight depending on actual user needs. A modified KANO model is applied to distinguish whether the sub-function comes from user pain points or user excitement points. Usually, needs that come from users' pain points are rigid needs and are given a score of 2 points, while needs that come from users' excitement points are icing on the cake and are given a score of 1 point. In order to focus the concept converging on user experience design, the independent variable of the X-axis is modified to If technology can achieve it. As the following figure 30 depicts, 8 sub-functions are classified according to three quality types.

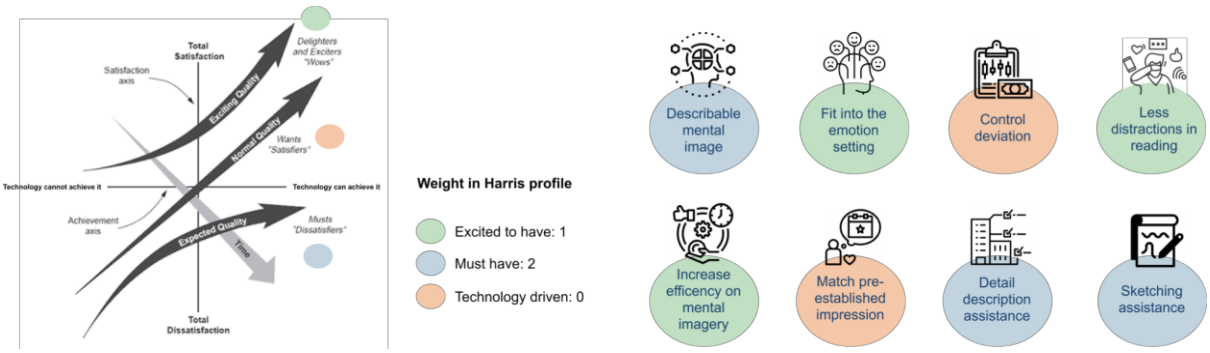


Figure 30: Use KANO model to assign weighted scores to sub-functions

Describable mental image, detail description assistance, and sketching assistance are expected quality, because they are the foundation of image generation for specific types of users, but it doesn't mean they will create images with high-satisfaction, since it still depends on other factors like emotion setting, etc.

Control deviation, match pre-established impressions are normal quality because as long as the AI technology becomes more mature in precisely controlling various parameters in image generation, these two sub-functions can be easily achieved. Therefore, those two are out of scope for the human-AI collaboration experience design.

Fit into the emotion setting, less distractions in reading, and increase efficiency on mental imagery are exciting qualities, because they can increase the quality of the image output and the user perception of inputting. However, they only take effect when the image can be correctly generated based on human input.

With the weight applying on all the criteria, I calculated the total score for each alternative in Harris Profile. The 2 concepts in the following figure get higher scores than the others. The differences between these two lie on Detail description assistance and Less distractions in reading. Explainable AI can provide more detailed description in output elements and generation logic. But expectation establishment causes more distractions in reading because it requests users to pre-read some images to refresh their memory.

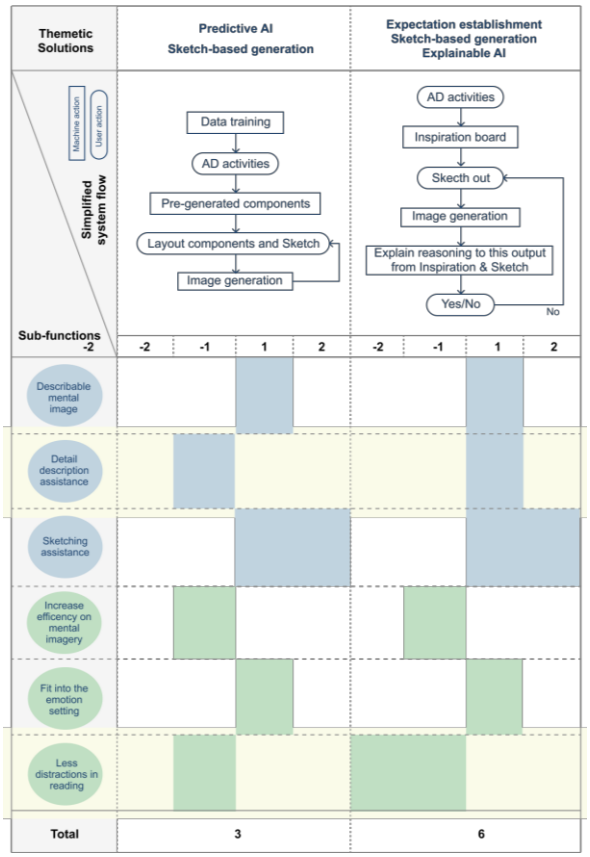


Figure 31: The two solutions with high scores

7.4.5 Concepts further elaboration

7.4.5.1 Solution flow

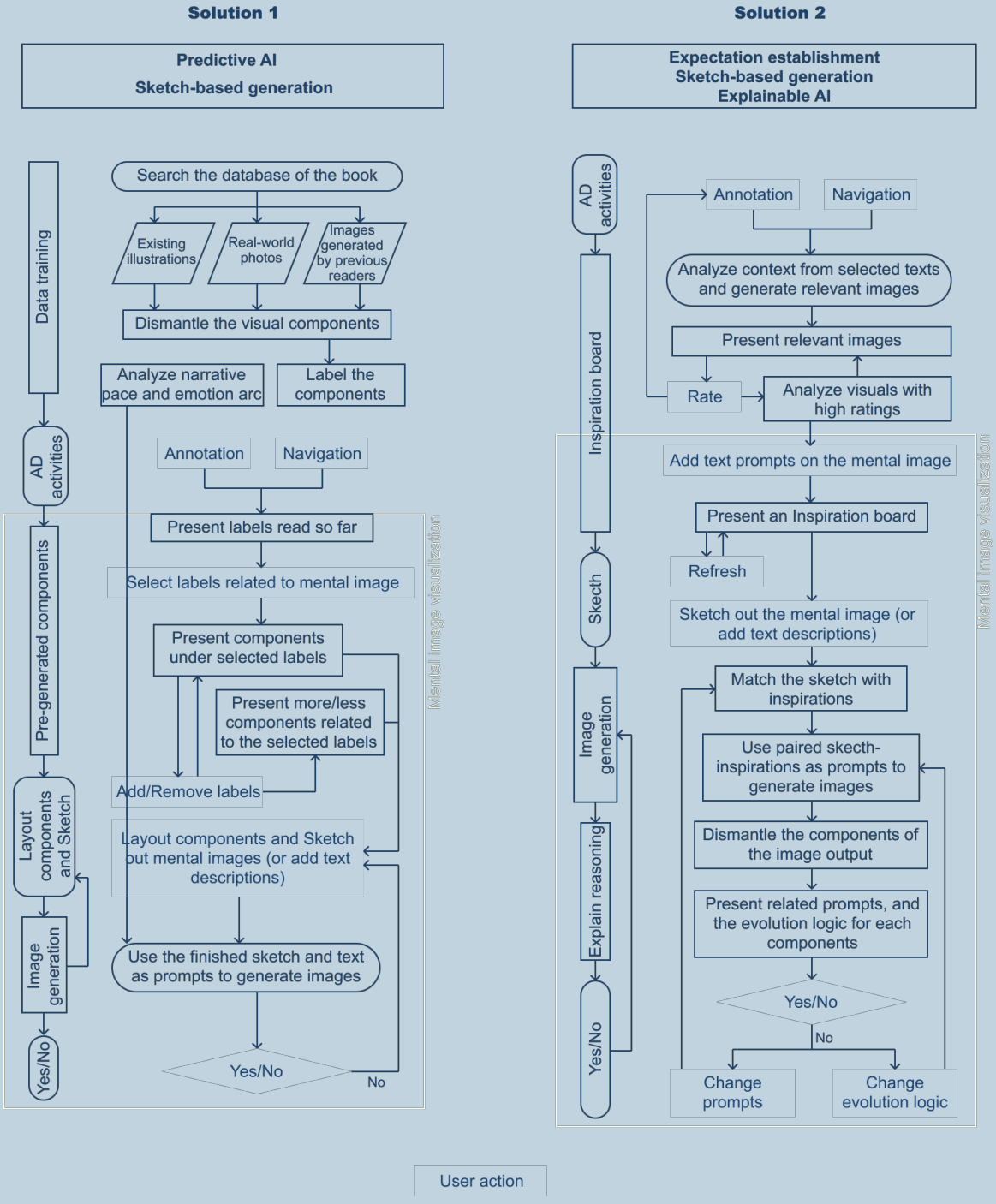


Figure 32: Solution flow of the two selected concepts

Solutions 1 first utilizes Predictive AI for information analysis and output predicting. This is achieved by searching the database of the selected book on existing illustrations created by authors or illustrators, real-world photos regarding the context of the narratives, visual works such as movies/comics based on this book, and images generated by previous readers if the AI system has been applied for other readers. After gathering sufficient data,

AI will analyze the components of those visuals and dismantle the minimum object units. Then AI will identify those objects which are described within the content of the book, and label them with the terms in the book. This dataset of visual components is the foundation of predicting and assisting user's mental image visualization. Secondly, solution 1 empowers the mental model visualization process by presenting visual components for users to manipulate and create sketches. It will first present all the labels the user has read so far, and the user can select relevant labels towards their mental images. Under the selected labels, AI will predictively present 4-5 visual components in easy sketch style. The user is able to add or remove labels if they realize more or less components mentally. Afterwards, the user can freely drag or resize the components in the sketching area. Together with manually sketching, the prompts for AI to generate a final mental image output will be a combination of presented visual components and manual sketches. The text prompt is also required but not necessary to be detailed. Thirdly, the modification function is built as a non-instructive process. The user can intuitively relate their sketch to the final AI output, then through adjusting their sketch, components layout and text prompts to modify the AI output.

Solution 2 begins by navigating and annotating the text to highlight sections they want visualized. The AI analyzes the selected text to understand its context and presents a set of relevant images. Users rate these images based on their relevance and accuracy to their mental image. The AI further analyzes images that received high ratings to understand user preference for further visualization. Therefore, AI can capture the instant user mental images efficiently before they become vague. During mental image visualization, users can add specific text prompts to generate an inspiration board. Then users sketch their mental image or add detailed text descriptions by referring to this inspiration board. After finishing the sketch, AI will use the paired sketch and inspiration elements as prompts to generate more refined images. At the same time, the system presents related prompts and explains the evolution logic behind each component of the image, providing transparency and understanding of how the image was generated through AI. Users decide if the generated images meet their expectations, if not, users can modify the text in the explanation table to regenerate image outputs.

The two solution flows differ on user perception in 3 variables: distractions during reading, sketching assistance and subsequent modifications. The following figure explains the difference, which will be evaluated through user testing afterwards.

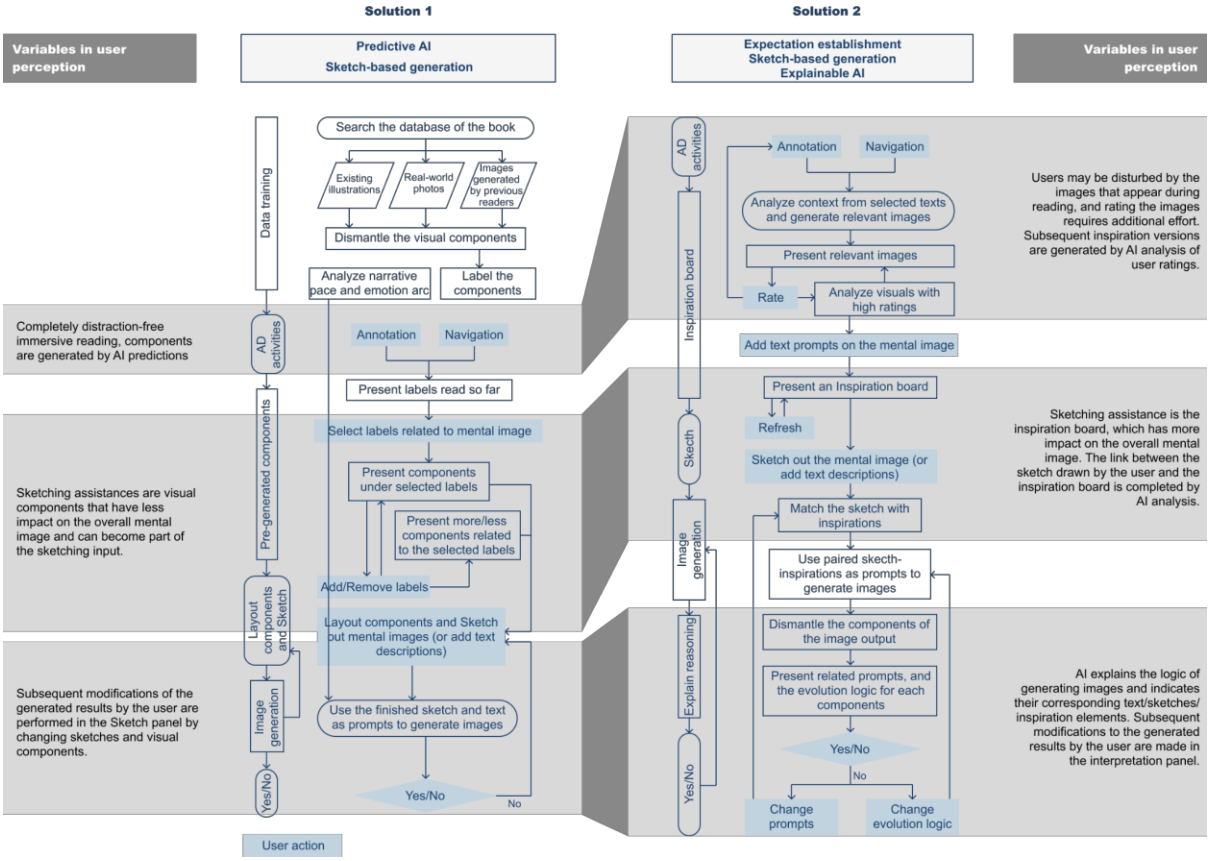


Figure 33: 3 Variables of the differences in user perception between the two solutions

7.4.5.2 Input/Output System Flow

The following input/output workflow explains how the user action leads to the desired output through data processing and computing.

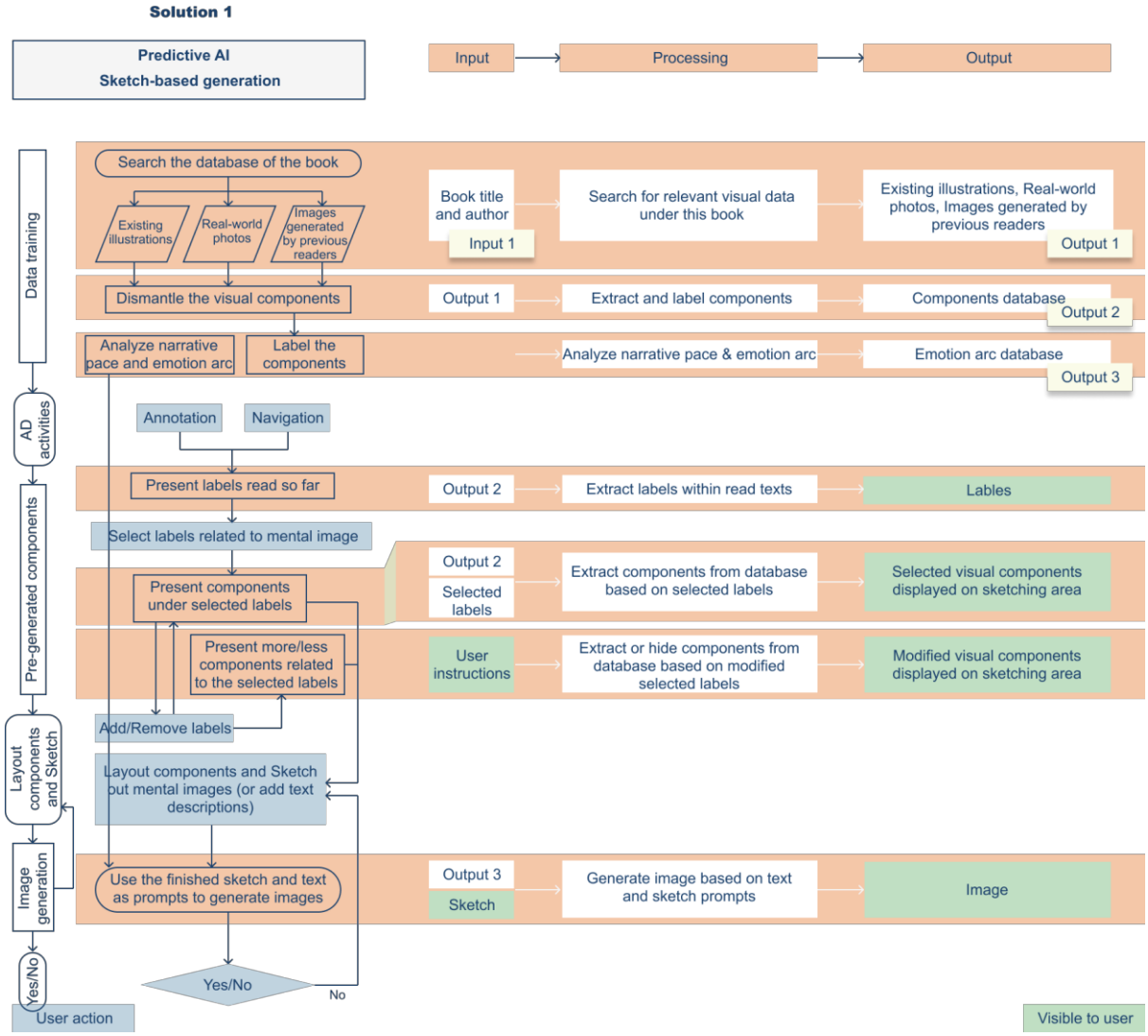


Figure 34: Input/Output System Flow of Solution 1

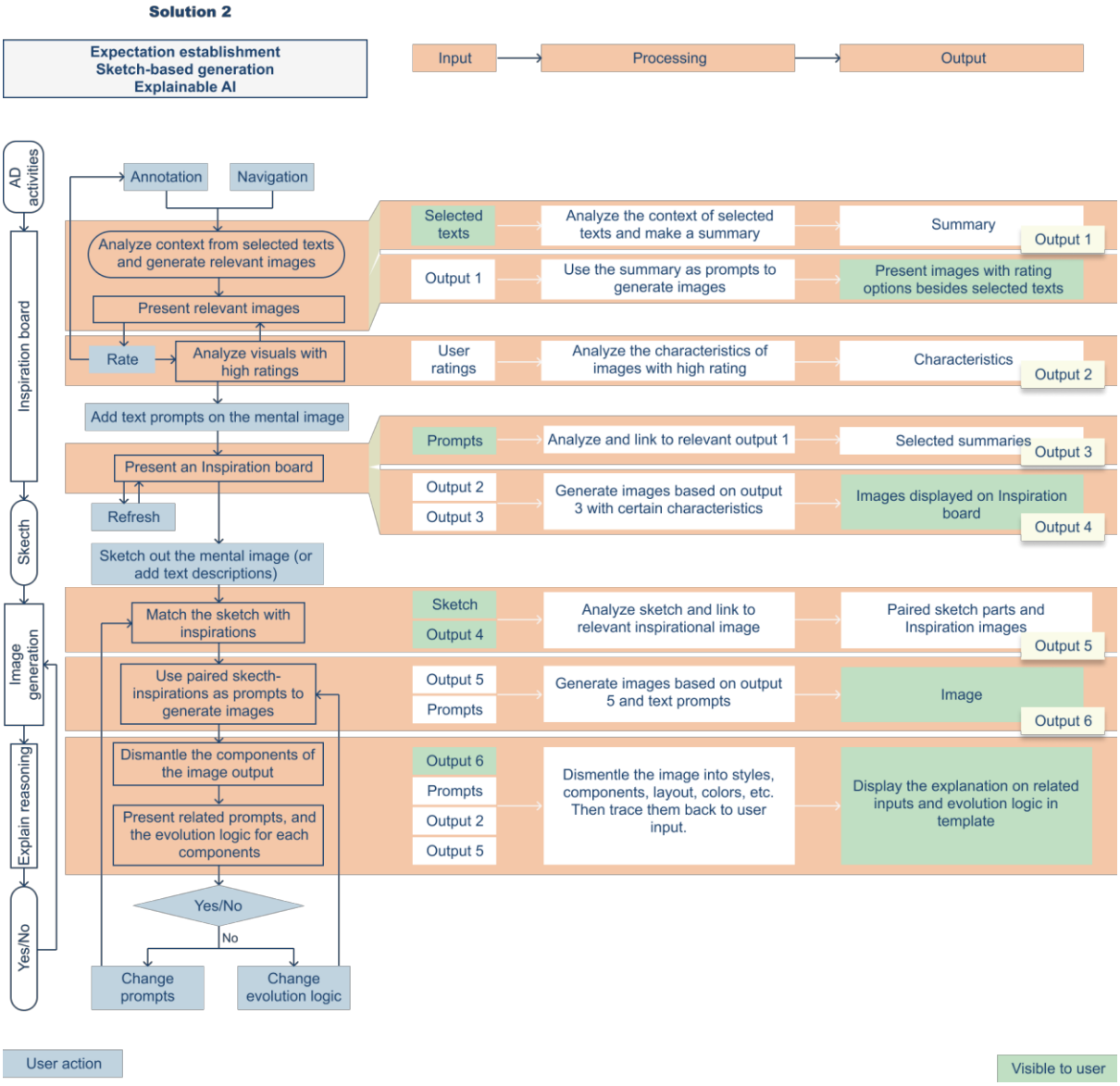


Figure 35: Input/Output System Flow of Solution 2

7.5 Prototype and User Test

The purpose of prototyping the two solutions is to evaluate the effectiveness of mental image generation with AI, and the variables of user perceptions that influence user experience. Based on the Input/Output system flow analysis, the visible interface elements (green squares in Figure 34,35 are the focus of the prototyping. The complete prototype design is in Appendix 3.

7.5.1 Purpose of user test

The purpose of the user test is multi-faceted, aiming to validate the feasibility of the AI-assisted mental image visualization workflow and evaluate the user experience of the selected solution flow. This evaluation focuses on user satisfaction with the image output, specifically assessing 1) how well the images align with users' mental images and 2) whether they exceed expectations. Additionally, the user experience of the image generation process is scrutinized through various metrics, including 1) extraneous cognitive load, 2) efficiency, 3) disturbance avoidance, and 4) inclusivity. The test also aims to select user perception variables to enhance reading and visualization experiences based on the vividness of users' mental images. Based on user input, the system flow will be iterated and refined, and user personas will be created to inform further system design.

The evaluation metrics are highly relevant to the design of a digital service powered by an AI image generation tool that visualizes mental images in narrative books based on human input. **Alignment with mental images** ensures that the generated images match the user's expectations and mental structure, which is critical to user satisfaction and engagement with product design. **Excitement beyond expectations** measures the pleasure factor, indicating the extent to which the service surprises or shocks the user. **Extraneous cognitive load** assesses the mental burden of the process, and this metric is prominent in the results of user pain point research (chapter 7.3.2). **Efficiency** is identified as a key user need in user ritual research (chapter 5.3.2), which reflects the usefulness of the tool. **Avoiding distractions** ensures that the process is seamless and undisturbed, which is critical to maintaining user attention and satisfaction (chapter 7.3.1). **Inclusiveness** is another key user need (chapter 5.3.2), assesses whether the tool meets the personalized needs of different users, regardless of their abilities or backgrounds. Together, these metrics provide a comprehensive assessment of the user experience, guiding the design and improvement of digital services.

7.5.2 Data collection

The participants of this user test are narrative readers who seek for active reading experiences. The sampling method adopted a random selection of users who fit the reading demand.

The user test is done by **between-subjects design** testing. In a between-subjects design, different groups of participants are exposed to different conditions or solutions. Each participant only experiences one condition, allowing for a comparison between groups rather than within the same group (Charness et al., 2012). Each solution will be tested individually by a group of 5 people. I

randomly assigned participants to one of the two groups to avoid selection bias.

Within each group, the VVIQ test (Marks, 1973) will be required for every participant, and I will **select participants with different mental imagery abilities on purpose**. For each group, make sure there are at least one high visual imagery user and one low visual imagery user.

The data is collected through a **mixed-methods approach**, combining quantitative rating scales with qualitative follow-up interviews (Schensul & LeCompte, 2012).

Quantitative Rating (0-5 Scale) provides a numerical measure of their responses.

- Participants are presented with specific criteria or questions.
- They are asked to rate their experience, satisfaction, or other relevant aspects on a scale from 0 to 5 (where 0 might indicate very negative and 5 very positive).

Qualitative Follow-Up Interviews provide deeper insights and context to the numerical data collected.

- After the rating, participants are interviewed to explore the reasons behind their ratings.
- Open-ended questions are used to delve into the participants' thoughts, feelings, and experiences.

The complete user testing plan is in Appendix 4.

7.5.3 Results

The user test is conducted through a combination of qualitative interview and quantitative rating. 5 participants for each solution (10 in total) are recruited. After participant selection by finishing the VVIQ test, 2 solutions were both tested by people with high and low mental imagery abilities. Table 5 provides a clear overview of the distribution of users with varying levels of visual imagery abilities for each solution tested.

Table 5: Participant inventory of User test

Users	Extreme High Visual Imagery Ability	High Visual Imagery Ability	Low Visual Imagery Ability	Extreme low Visual Imagery Ability
Solution 1	0	4	1	0
Solution 2	1	3	1	0

The output of user testing is in Appendix 5.

7.5.3.1 Comparison of the ratings of the two solutions

The following figure is the quantitative data analysis results, showing the average of the user perception scores of different criteria in the two solutions. A score of 3.5 is above the midpoint (2.5) of the scale, indicating a performance level that is slightly better than average. Setting the dividing line as 3.5, we can see that 12 out of 21 criteria can relatively

meet user requirements. The following qualitative interview explains the satisfaction and insufficiency of each solution pointed out by users.

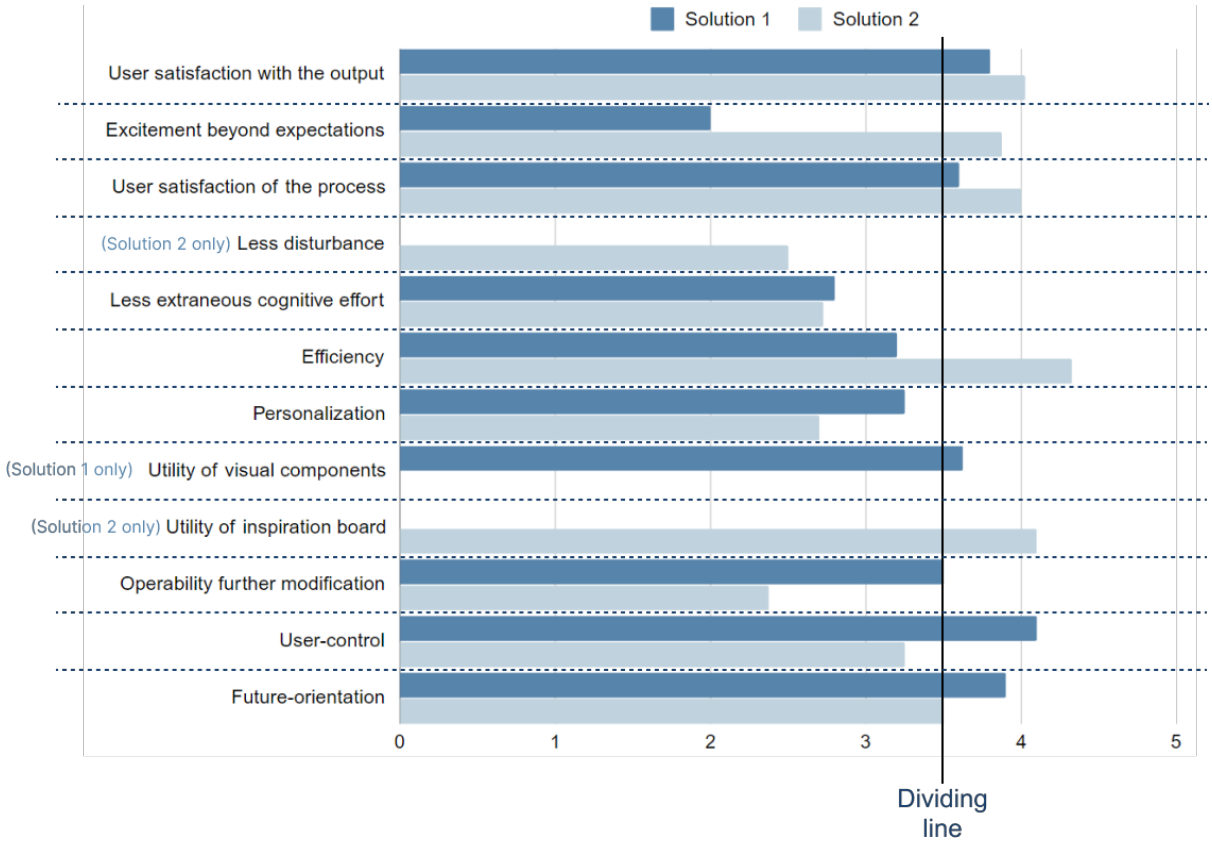


Figure 36: Results of quantitative rating for user testing of 2 solutions

The following concluded user needs are numbered from **a to p**, which links to the design iteration (Chapter 7.6).

Criteria scored over 3.5

User satisfaction with the output:

The overall user satisfaction with the output is high. Most participants consider the output is aligned with their mental image to a large extent. The AI tool can capture most of the requirements from text and sketch inputs in both solutions. However, the output still has deviations from what they desired in terms of perspective **(a)**, mood and contents. And this deviation is magnified in the eyes of high visual imagery users.

> *"The layout and content are what I wanted, but the atmosphere and colors are weird."* — a participant of solution 1

Excitement beyond expectations:

DALL.E (the AI Model I used in my prototype) has been trained on a vast dataset containing numerous images and their descriptions. In the absence of specific instructions, DALL.E may fall back on default assumptions based on typical patterns seen in the data. For example, if a prompt mentions "a room," DALL.E might add furniture and decor items

that are commonly found in rooms. Therefore, the final output will automatically be added with details not mentioned in the prompts to increase the richness and realism of the picture **(b)**. Some users find it nice to have those extra details, while some users identified those as extra or incorrect visual elements. To minimize the added elements in the output, users must spend extra time on refining the prompts and regenerating images.

> *"I didn't ask for the books, the candles, but it really adds to the scene, and it's quite nice."* — a participant of solution 2

> *"I sort of expected a little bit less than this."*— a participant of solution 1

User satisfaction of the process:

In general, most participants feel satisfied with the reading and mental image visualization process. It offers understandable guidance for users to generate desired images through AI step by step. Since the two solutions both have the step of sketching, people have different willingness to sketch. Some people feel it's as interesting as making a collage. Some other people consider sketching as a big burden and they don't think they use this process in their normal reading **(c)**.

> *"It feels like making a collage."* — a participant of solution 1

> *"It's a big pressure for me to sketch, I'm bad at sketching and I don't enjoy sketching too."* — a participant of solution 1

Efficiency of solution 2:

Most participants agree that it's efficient to generate images through the process of solution 2. Because the user preference analysis was done during reading, which also influenced their impression and memory of the provided narrative article. And the sketching step doesn't need to evaluate and layout visual components as it is in solution 1. I believe users already suspected what the final output will look like through the inspiration board, so they think it's efficient. However, this efficiency may lower the user control, as I will mention in the criteria of user control.

Utility of visual components of solution 1:

Most participants think the visual components as sketch assistance in solution 1 is helpful. By selecting the relevant labels in the list, they can reflect on the narrative article and enhance their memory. And drag-and-drop elements take some of the stress out of drawing. Especially for those people who like sketching, it's fun to draw their mental images quickly with assistance.

However, some participants mentioned that the ideal visual components are not in the list, and it takes extra effort to select one visual out of five **(d)**, especially when they look almost the same.

> *"Sketching is my hobby, now I have these tools to help me do it, I think it's fantastic. And those components can also help me recall the story."*

> *"It does help me with sketching by increasing efficiency, but some of them are not what I want, it takes effort to select."*

Utility of inspiration board of solution 2:

The inspiration board does help users formulate their mental image and sketch very well, by offering color palette, visual

components and different layouts. Besides, users with low sketching skill can learn how to draw something from the examples in the inspiration board. However, they hold a negative attitude towards those very complete visual pictures. On one hand, those pictures sometimes are not aligned with their mental image, and will affect or even overwrite their original mental images**(e)**. On the other hand, if those pictures already meet their satisfaction, it is meaningless to do the next-step sketching **(c)**.

> *"When I looked more at the inspiration board, I had a slightly more well-lit space in mind. That perception changed the more I looked at these inspiration boards and the intermediate results as well."*

> *"It's a two-edged sword, I think it does help me in imagining that way better but it can also limit me."*

Operability of further modification of solution 1:

Most participants feel satisfied with the further modification step of solution 1. They can easily recognize which parts of the sketch or text prompts need to be refined to regenerate the desired image. However, due to the uncertainty of AI analyzing and generation, sometimes the modified output still misses the accuracy. Some users have to try several times in order to get a satisfactory output **(g)**.

> *"The sketch can be used to modify the layout and the text can be used to modify the style, mood."*

> *"I think the modified output hasn't changed a lot according to my input modification."*

User control of solution 1:

The user's perception of user control is largely determined by whether the final effect is consistent with the mental image. If the final output is aligned with their mental image, those participants will feel the algorithm is under control. On the contrary, no matter how many limitations they have set up through sketch and text prompts, they still feel the algorithm is out of control **(g)**. Luckily most users in solution 1 feel they have decent user control over the AI generation process.

Effects on knowledge construction:

Nearly all the participants are positive about the effects of the AI visualized mental image on knowledge construction and reflection of the given reading material. First of all, the image output helped them feel the story scenes and atmosphere in a concrete way, which can enhance reading immersion. Second, the image output improved the readability and understandability of the narrative, especially for users who feel alien about some settings. Third, the image output as a visual identification strengthened their memory and recall to the given narrative.

> *"It can help me immerse myself in the scene in the book because the combination of the picture and the text gives me a more intuitive sense of immersion."* — a participant of solution 2

> *"Now when I read it again, I'll think more about this particular scene instead of the scene I have in my mind."*

Future orientation:

Participants believe that this visualization service can adapt to the future reading trends as AI is more and more popular among digital reading fields. It's positive to adopt this

service in the near future, especially for some fantasy/fiction novels, or for children's books.

> *"If you are reading a fantasy novel or a science fiction novel, and you really have a huge world in your mind, but you can't draw it, you may want to take a real look at it, and use this to help you construct a new world."* — a participant of solution 2

> *"It can be applied to children's storybooks."* — a participant of solution 1

Criteria from 2-3.5

Disturbance of solution 2:

Users in solution 2 feel the activity of choosing visuals during reading does cause some disturbances for their continuous reading experience. They feel comfortable with looking left and right to jump from visuals and texts. But each option was a very detailed picture, and although the style was very different, the content was basically the same. Therefore, for participants who wanted to pick out the details that matched their mental image, they had to spend a lot of time carefully looking at the picture of each option **(h)**. This review and inspection separated them from reading the story. When none of the options align with their mental image, they feel frustrated.

> *"Usually, I just read continuously so a lot of doing this is unnatural"*

> *"There are a lot of disturbances because most options are not aligned with my mental images, they drag me away from my own thoughts."*

Extraneous cognitive load:

Both solution 1 and solution 2 bring many extraneous cognitive loads that are irrelevant to the knowledge construction of the given reading material. Most of the extraneous cognitive load comes from the pressure of sketching out mental images. They can easily figure out what contents need to be involved in the image, but they cannot easily sketch out the specific contents, even with the assistance of visual components or inspiration boards. And some users feel it is unnecessary to sketch because the inspiration board generated from the text prompt is already good enough **(c)**. Besides, one participant mentioned her extraneous effort comes from the nonlinear order of the visual components list **(i)**.

> *"I feel confused when considering how I should be more descriptive on sketching and text prompts."* — a participant of solution 1

> *"It's a big pressure for me to sketch, I'm bad at sketching and I don't enjoy sketching too."* — a participant of solution 2

> *"The visual components list is not in a linear way as they showed up in the story, it causes concussion."* — a participant of solution 1

Efficiency of solution 1:

As I mentioned in the analysis of user control, the user perception of efficiency also depends on the user satisfaction of final output. If they spend much time on modifying prompts and regenerating images, the efficiency will be perceived as low **(g)**. Besides, participants who don't enjoy sketching would consider sketching a waste of time.

"I spent a lot of time sketching, but the output is not what I want exactly, I still have to spend time modifying." — a participant of solution 1

"The time that I take to sketch, and I sketch very poorly, as you see, I would never do that." — a participant of solution 1

Personalization:

The text-to-sketch-to-image process doesn't support users' needs for a personalized reading experience in a holistic way. Through sketch and text prompts, it does help users in personalizing the content of their mental image. However, some users are not expecting a well rendered image as a fixed output for their reading experience. Some users think the concrete images limit their imagination of the story. They want to maintain the flexibility to imagine different images each time they reflect on the article **(j)**.

And sketching is not a compulsory step for some users to visualize mental images.

"Ideally, I wouldn't use any of this because I do like imagining what a scene would look like in my mind. I feel like if I sketch it down, then my image will always be that. So I try to avoid doing it so that every time I can have a slightly different experience." — a participant of solution 2

"The details, the color tone, the character's appearance and so on may be a little different from what I imagined, and may lose the original appearance in your mind." — a participant of solution 1

Operability of further modification of solution 2:

The further modification of solution 2 is achieved by adjusting the text explanations of the table, which displays how the AI evolves input into the final output. Although the explanation table offers some guidelines by specifying elements in the output image, there are always some elements missing in the table. Otherwise, the table might be too long for users to read. Besides, most users already recognized what they wanted to change once they saw the generated image **(k)**. The further modification happened more in an intuitive way. Under this circumstance, it's extra work to read the explanation table and think about which column they can change.

"I think it helped me specify what I should change in that way. There are certain things like the chair, background, visual style, and mood."

" I don't know what to change or what stuff I want to change is not mentioned in the elements."

User control of solution 2:

In solution 2, the inspiration board is generated by analyzing the text prompts and user preference. It is likely that it already provides a large number of examples of visual elements that match user preferences. Some users think their mindsets are controlled by the inspiration board and their original mental image fades away when seeing the well-elaborated inspiration board **(l)**. So, participants of solution 2 feel less user control in the visualization process.

"Because if the images of the inspiration board are already available, then that's not a lot of room for me to imagine, because I already know what the images could potentially look like."

Discussion

The user research analysis provides valuable insights into the strengths and areas for improvement of the AI-driven visualization tools used in reading experiences. Overall, users show a high level of satisfaction with the output, particularly when it aligns closely with their mental images. The visualized mental images enhanced reading immersion, improved readability, and strengthened memory and recall of the narrative.

Participants felt a sense of control when the output matched their mental image. However, the content of this output is also largely limited by the current development of AIGC technology. Perhaps in the future, through more technical optimization and data training, the output effect will be more in line with user expectations. However, it is currently difficult to optimize technology through user experience design **(b, g)**.

Both solutions added cognitive load, mainly from the pressure of sketching. Even though sketching can help enhance deep thinking into a user's mental image and add more control into AI generation, sketching is not an appropriate step for everyone **(c)**.

While the text-to-sketch-to-image process aids in personalizing content, it sometimes limits users' imagination, as the fixed output can restrict their creative flexibility **(j)**. Solution 2 users feel a lack of control as the inspiration board can overshadow their original mental images, making the visualization process feel less personal and more dictated by the AI.

Looking back on the variables of user perception, the user testing provided insightful directions in comparison to the two solutions.

Distractions during reading

The distractions caused by selecting visual options during reading is acceptable. Sometimes it can even satisfy users' curiosity and promote active reading. The distractions actually come from evaluating the details of the options **(h)**.

Sketching assistance

1. **Utility of Visual Components in Solution 1**: The visual components helped users with sketching, making it easier and more enjoyable, especially for those who like drawing. However, some participants found the selection process cumbersome when ideal components were not available **(d, i)**.
2. **Utility of Inspiration Board in Solution 2**: The inspiration board effectively helped users visualize and sketch, but complete visual pictures sometimes conflicted with their mental images or made further sketching redundant **(e, h, l)**.

Subsequent modifications

Users found it easy to identify which parts needed refinement through intuition. The explanation table is for indicative purposes only, and reading and modifying the table creates an additional burden on the user **(k)**. The user perception of subsequent modification in solution1 is better than that in solution 2.

7.5.3.2 Comparison of ratings of users with low/high visual imagery abilities

For people with different mental imagery abilities, the perceptions on the same feature are also different. The following figure is the mean scores of low visual imagery users (VVIQ hypophantasia) and high visual imagery users (VVIQ hyperphantasia) from the participants.

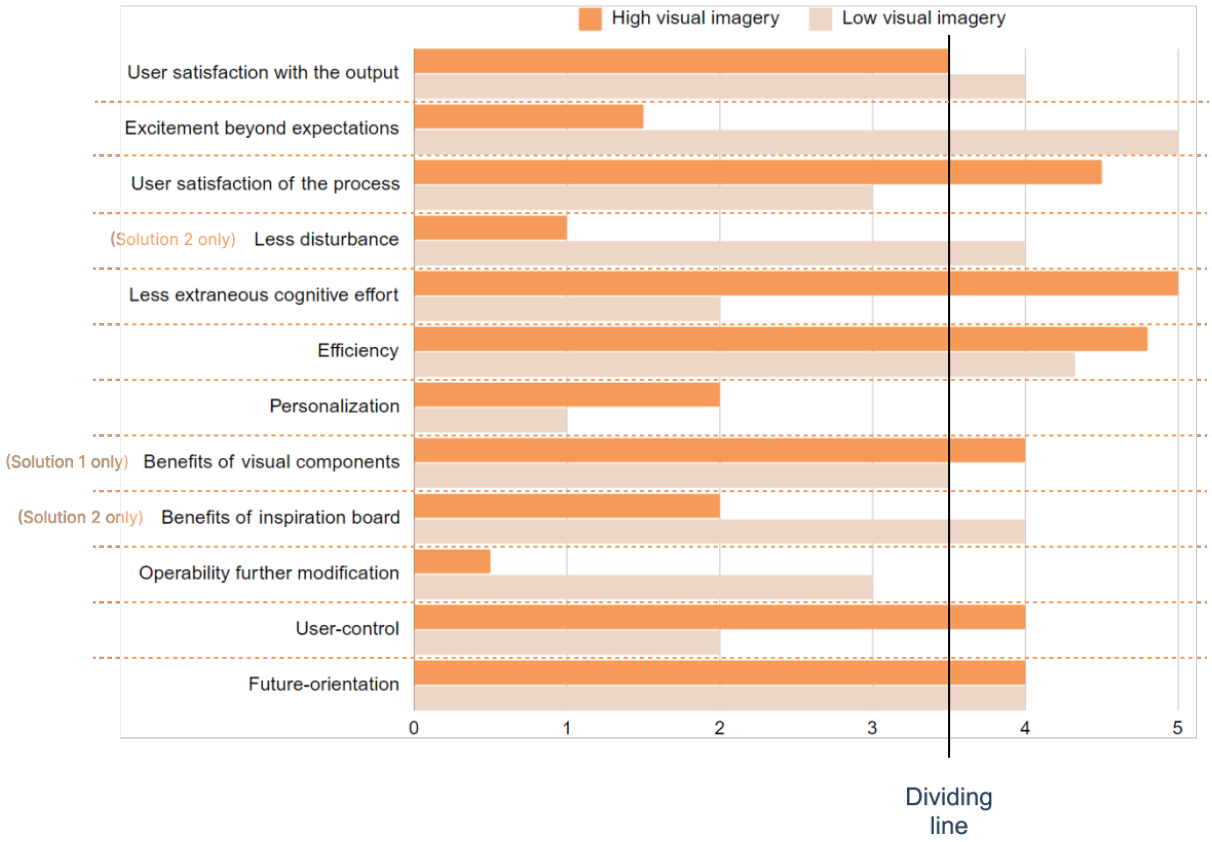
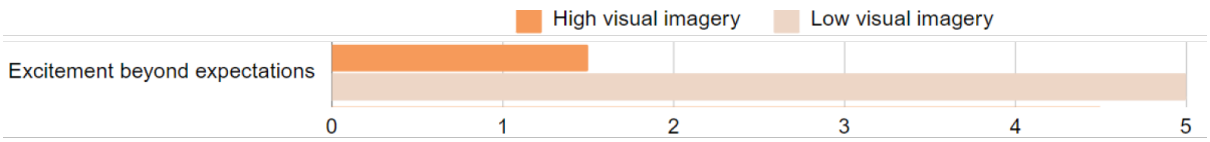


Figure 37: Results of quantitative rating between low and high visual imagery users

Among all the evaluation parameters, the following points showed larger differences.

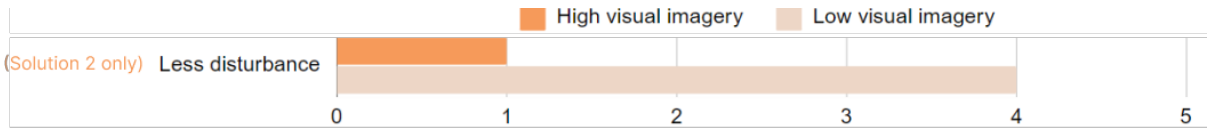
Low visual imagery users feel more excitement beyond expectations



High visual imagery users don't feel the output is beyond their expectation, and normally find some parts of the output are different from expectations. While low visual imagery users feel very surprised by the output with fruitful contents and details. They didn't have a very clear decision on the specific objects but only the overall mood and contents in mind.

> *"So a lot of details in this image I hadn't thought of, it does capture."* — a hypophantasia participant.

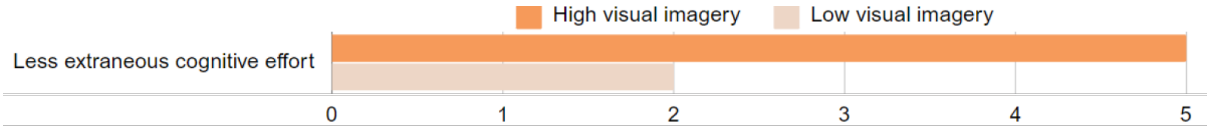
High visual imagery users perceive more disturbance in choosing options:



Most users feel comfortable with reading while selecting relevant images popping up. High visual imagery users thought the reading experience of Solution 2 causes a lot of disturbances. Because they have vivid mental images and they focus more on the details of the options and feel most of them are not aligned with their mental image **(h)**. Those extra efforts can divert their attention away from reading.

> *"There are a lot of disturbances because most options are not aligned with my mental images, I looked at each picture carefully, but it was hard to make a choice."* — a hyperphantasia participant.

Sketching causes extraneous cognitive efforts

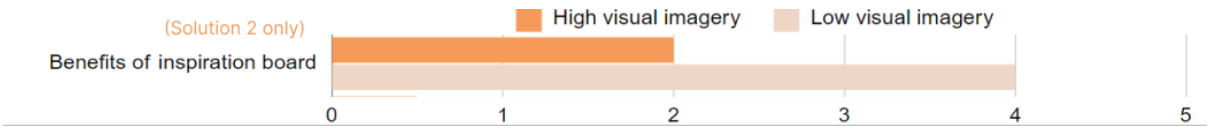


Low visual imagery users and some other users with low sketching skills find there is much extraneous cognitive effort to make during sketching and text describing their mental image. Even though there are many types of visual assistance in helping them sketch, they still find sketch itself is not a natural way for their normal reading experiences **(c)**.

> *"It's a big pressure for me to sketch, I'm bad at sketching and I don't enjoy sketching too."* — a hypophantasia participant.

> *"I'd actually completely forgotten about the story by the time I was messing around with this."* — a phantasia participant.

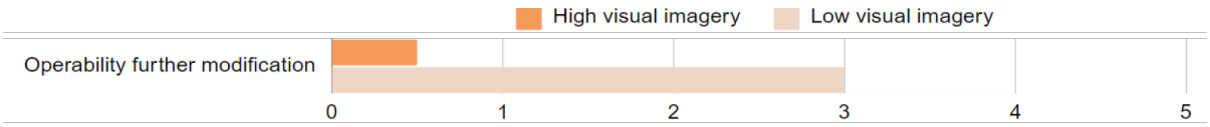
High visual imagery users are correcting the inspiration board



Most people find the inspiration board is insightful for them to sketch out their mental image. However, for some high mental imagery users, they find it less useful because they first need to identify irrelevant and erroneous elements in the inspiration board, then try to correct them in their sketch. So the benefits of the inspiration board for those kinds of users are more like removing elements they don't want **(m)**.

> *"I can learn how to draw the tower from it, but those inspirations are not what I was imaging about, I need to correct them."* — a hyperphantasia participant.

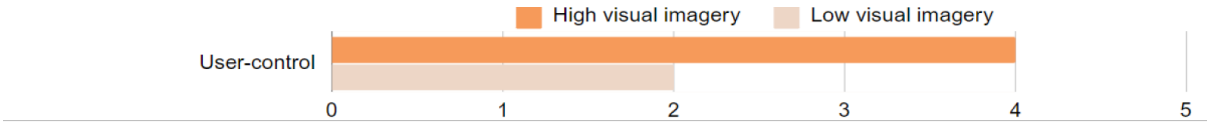
High visual imagery users find further modification less useful



For most users, the explanation table of Solution 2 offers some basic directions on modifying specific aspects. However, it's not very applicable for high visual imagery users due to the lack of parameters they wanted. The table offers an overall description in terms of contents, mood and color palette, but those users focus more on the specific details that are not being mentioned by the table **(n)**.

> *"The things I want to modify are not in this table."* — a hyperphantasia participant.

Low visual imagery users feel less user-control



For those users who generated desired output images, most of them feel the reading and visualization process is in control because of the prompts they input. But some low visual imagery users feel the AI is controlling their thinking because they don't have a precise mental image, and they can only say yes or no to the AI generated output. Therefore, some of them feel frustrated with the excellent image generated by AI, and they don't have the ability to say which part is not desired **(o)**. Some people think the whole visualizing mental image is not a natural behavior in their active reading **(p)**.

> *"Once I use this, I think the control is probably to feel like in some ways it is controlling my thinking more than me controlling the visualization."* — a hypophantasia participant.

Discussion

From the user analysis above, I created 2 user needs profiles by listing the typical characteristics and user needs for low visual imagery users and high visual imagery users. This will guide the design iteration, as well as the participant selection of final design validation.

Low visual imagery users

"Abstract Adventurer"

Reading Preferences:

- Relies on overall feelings and abstract concepts rather than specific details when visualizing
- Appreciates detailed visuals provided by AI to supplement his less vivid mental imagery

Pain Points:

- Finds sketching and detailed descriptions challenging and cognitively taxing **(c, k)**
- Believes that sketching or even visualizing mental image is an unnatural behavior and is not always necessary **(c, p)**
- Feels less control over the visualization process due to the lack of precise mental images **(o)**

Needs:

- Have a supportive and easy-to-use interface that aids visualization without demanding extra cognitive effort
- Simplified tools for generating visuals without requiring detailed input or sketching skills
- Set sketching as an optional input

High visual imagery users

"Detailed Dreamer"

Reading Preferences:

- Enjoys visualizing specific scenes, characters, and intricate details while reading
- Finds illustrations and images helpful but very picky and strict about details

Pain Points:

- Finds parts of AI-generated images often misaligned with her detailed mental images, causing disturbance and breaking immersion **(h)**
- Spends extra cognitive effort to correct or filter out irrelevant elements in visual aids like inspiration boards **(m)**
- Feels less utility in further modifications due to the lack of specific parameters in tools like the explanation table **(n)**

Needs:

- Reduce the distraction of irrelevant details from their reading experience
- Tools that offer flexible parameters for further modification
- Advanced control over AI-generated visuals to ensure alignment with her detailed mental images

The following table lists extracted user pain points from the qualitative interview transcript.

*Table 6: E*xtracted user pain points of the prototype using experience

(a)	The output still has deviations from what they desired in terms of perspective (a)
(b)	The final output will automatically be added with details not mentioned in the prompts
(c)	Sketch is not necessary
(d)	Ideal visual components are not in the list, and it takes extra effort to select one visual out of five
(e)	Those pictures sometimes are not aligned with their mental image, and will affect or even overwrite their original mental images
(g)	The final output cannot always follow every detail of the prompts
(h)	Finds parts of AI-generated images often misaligned with her detailed mental images, causing disturbance and breaking immersion
(i)	Nonlinear order of the visual components list
(j)	The concrete images limit their imagination of the story. They want to maintain the flexibility to imagine different images each time they reflect on the article
(k)	Reading and modifying the table creates an additional burden on the user
(l)	Their original mental image fades away when seeing the well-elaborated inspiration board
(m)	Spends extra cognitive effort to correct or filter out irrelevant elements in visual aids like inspiration boards
(n)	Feels less utility in further modifications due to the lack of specific parameters in tools like the explanation table
(o)	Feels less control over the visualization process due to the lack of precise mental images
(p)	Visualizing mental image is not necessary

7.5.4 Design iteration

The system flow design for the mental image visualization process was heavily influenced by user feedback and needs identified during user testing. The new system flow integrates these insights into a structured process aimed at balancing user control, cognitive load, and the effectiveness of AI-generated visuals.

The system flow for the AI-assisted mental image visualization tool involves several key steps:

1. **Data Training**: Imagery objects from the book are broken down and used to generate visual components in different styles. Image recognition filters unmatched results, and components are labeled and stored in a database.
2. **AD Activities**: These user activities involve annotating and navigating the text, by when the system presents visual options to analyze user preference.
3. **Visual Presentation**: Relevant visuals are presented to the user, who selects and analyzes them. Visual components are organized in a mood board layout, allowing users to sketch or input text prompts.
4. **Image Generation**: The system generates images based on sketches or text prompts, with real-time text analysis and editing suggestions. Users can adjust sketches and prompts iteratively.
5. **Feedback Loop**: Users provide feedback on the generated images, which can be regenerated by adding context, adjusting complexity, or changing text prompts.

As is shown below, the left side of the figure 38 is the iterated system flow, which explains how the system works and how users interact with the system. The right side of the figure 38 is the user needs I responded to in the system flow iteration.

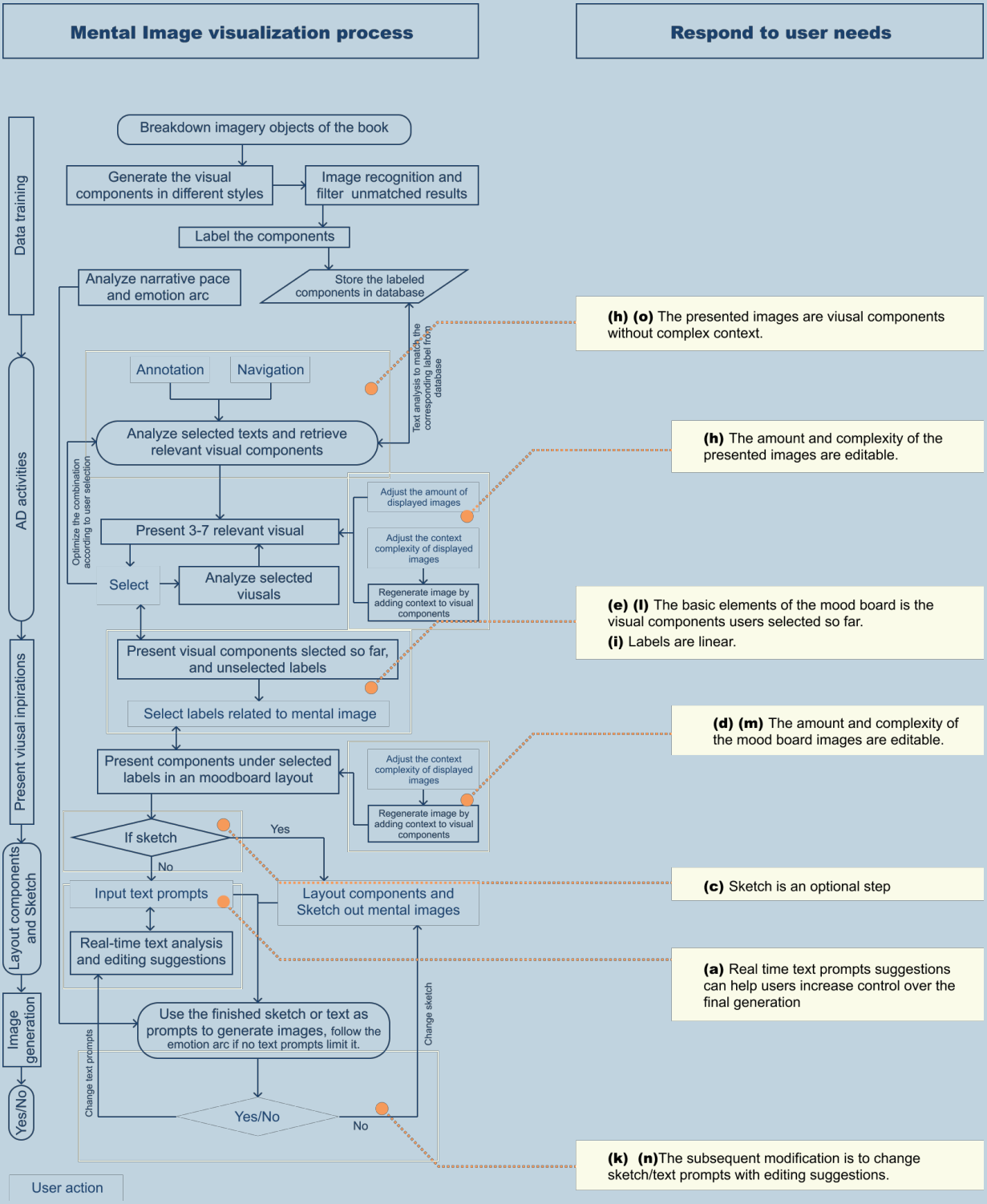


Figure 38: Iterated system flow of visualizing mental image and how the user need are responded

Taking user feedback into consideration, the workflow of the visualizing mental image system is iterated to meet different user requirements.

1. Distractions During Reading

User Feedback: Users found the distractions from selecting visual options during reading to be acceptable and sometimes beneficial, but evaluating the details of the options was a source of distraction.

Design Integration:

- The step of presenting visual options uses visual components of single objects, instead of well-established images, which aims to streamline the process of presenting visuals, reducing the need for users to zoom in and evaluate options extensively **(h, o)**. Besides, it responded to findings of the user ritual research (Chapter 5.3.1), as visual imagination during the reading process is limited to the immediate imagination of the part of the content that has been read. Compared with the reflective imagination after reading, the imagination at this time is more one-sided and vaguer.
- By freely choosing 3-7 visual components to present and choosing the complexity, the system aims to cater to people with different cognitive load requirements while satisfying user curiosity **(h)**.

2. Sketching Assistance

User Feedback: The visual components in Solution 1 helped users with sketching, making it easier and more enjoyable for those who like drawing, but the selection process was cumbersome.

Design Integration: The system provides "Present visual components selected so far, and unselected labels" to help users see available options in a linear order and allow users to refine their selection, which addresses the need for easier component selection and reduces cognitive load **(e, i)**.

3. Utility of Inspiration Board

User Feedback: The inspiration board helped visualize and sketch but sometimes conflicted with mental images, making further sketching redundant.

Design Integration:

- Mood boards are made up of a single visual image object and are completely selected by the user. Avoid images that are too contextual to interfere with the vividness of the user's mental image **(l)**.
- The subsequent "Adjust the context complexity of displayed images" steps allow users to manage and personalize the visual components more effectively **(d, m)**.
- Elements in the mood board can be dragged and resized to assist users in sketching.

4. Subsequent Modifications

User Feedback: Users found it easy to identify parts needing refinement, but modifying these parts through explanation tables was sometimes useless.

Design Integration: Users can change sketch/text prompts in a semi-instructive way. The flow includes step "Real-time text analysis and editing suggestions" to provide intuitive modification guidelines, guiding users to fill in the gaps in text prompts **(k, n)**.

5. For people with diverse visual imagery abilities

User Feedback: User found sketching and detailed descriptions challenging. Low visual imagery users feel less control over the visualization process due to lack of precise mental images. High visual imagery users feel a lot of hassle to correct or filter out irrelevant elements in visual inspirations provided by AI.

Design Integration:

- "If sketch" branch allows users to opt-out of sketching, catering to their preference for less extraneous-cognitively demanding tasks **(c)**.
- "Adjust the number of displayed images" and "Adjust the context complexity of displayed images" steps provide high visual imagery users with more control and flexibility **(h)**.

By integrating steps that provide both simplicity and control, the design aims to cater to both low and high visual imagery users. The flow minimizes extraneous cognitive load, allows for intuitive modifications, and enhances user satisfaction and control by providing a balanced approach to AI-generated visualization.

The user need (j) —- they want to maintain the flexibility to imagine different images each time they reflect on the article, and the user need (p) —- visualizing mental image is not necessary, are responded in the complete platform service design in Chapter 8.4.

7.5.4.1 AI integration in the system

This figure explains how AI participants in the visualized mental image creation journey. The right side of the figure 39 is the prompts for customized GPT, indicating which parts need AI involvement and how AI should be involved. Through actual test in ChatGPT-4o, the prompts work well in returning appropriate feedback.

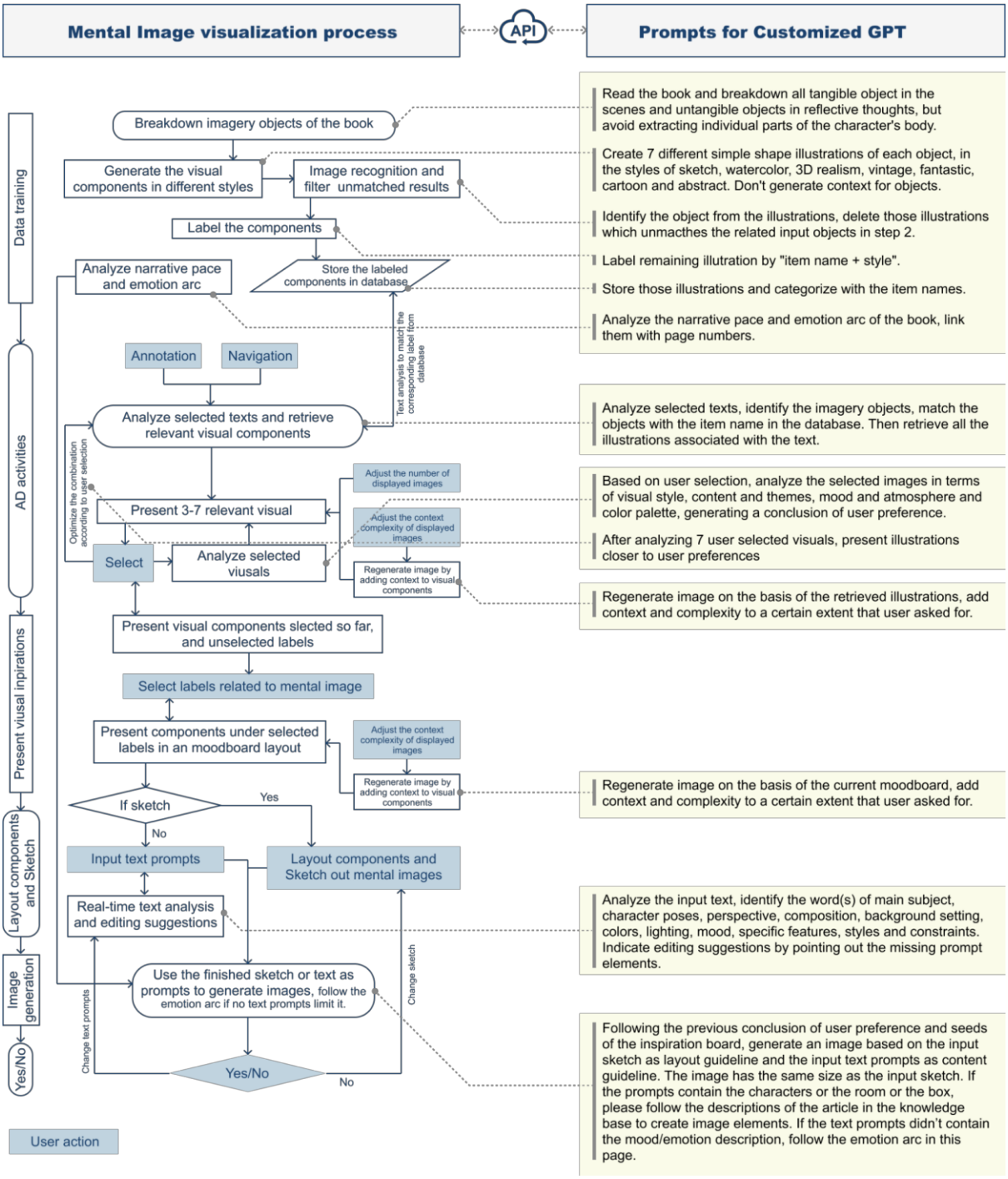


Figure 39: Prompts for customized GPT in realizing the designed system flow

7.5.4.2 Input/Output System Flow and Feature Design

The following input/output workflow explains how the user action leads to the desired output through data processing and computing.

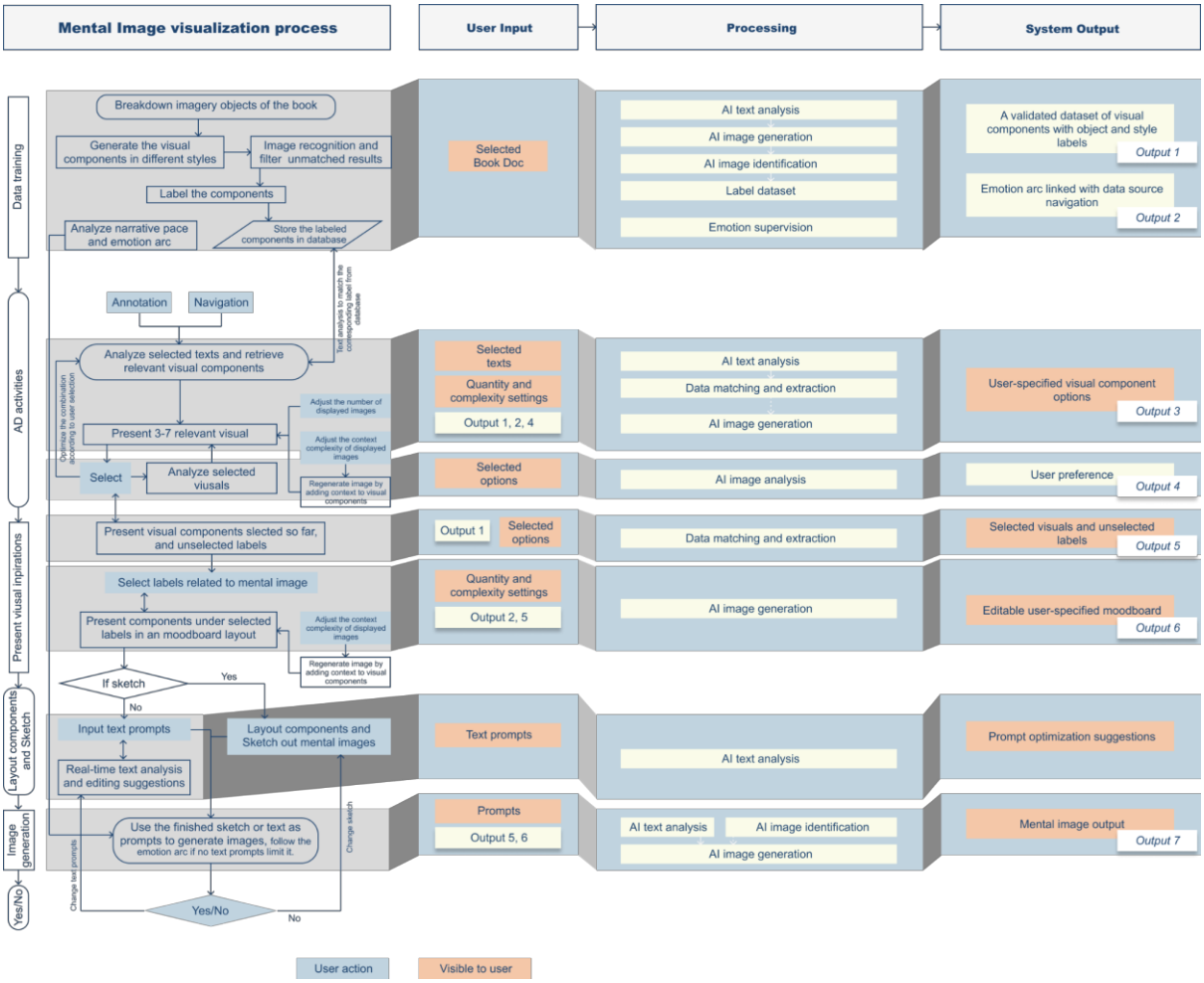


Figure 40: Input/Output System Flow of the iterated system

From the user input to system output flow chart, key technologies are listed to support the realization of each function. And the features that are visible to the user provide a guide for feature design of the user interface. The following figure shows how those visible features are designed in the final mental image visualization interface prototype.

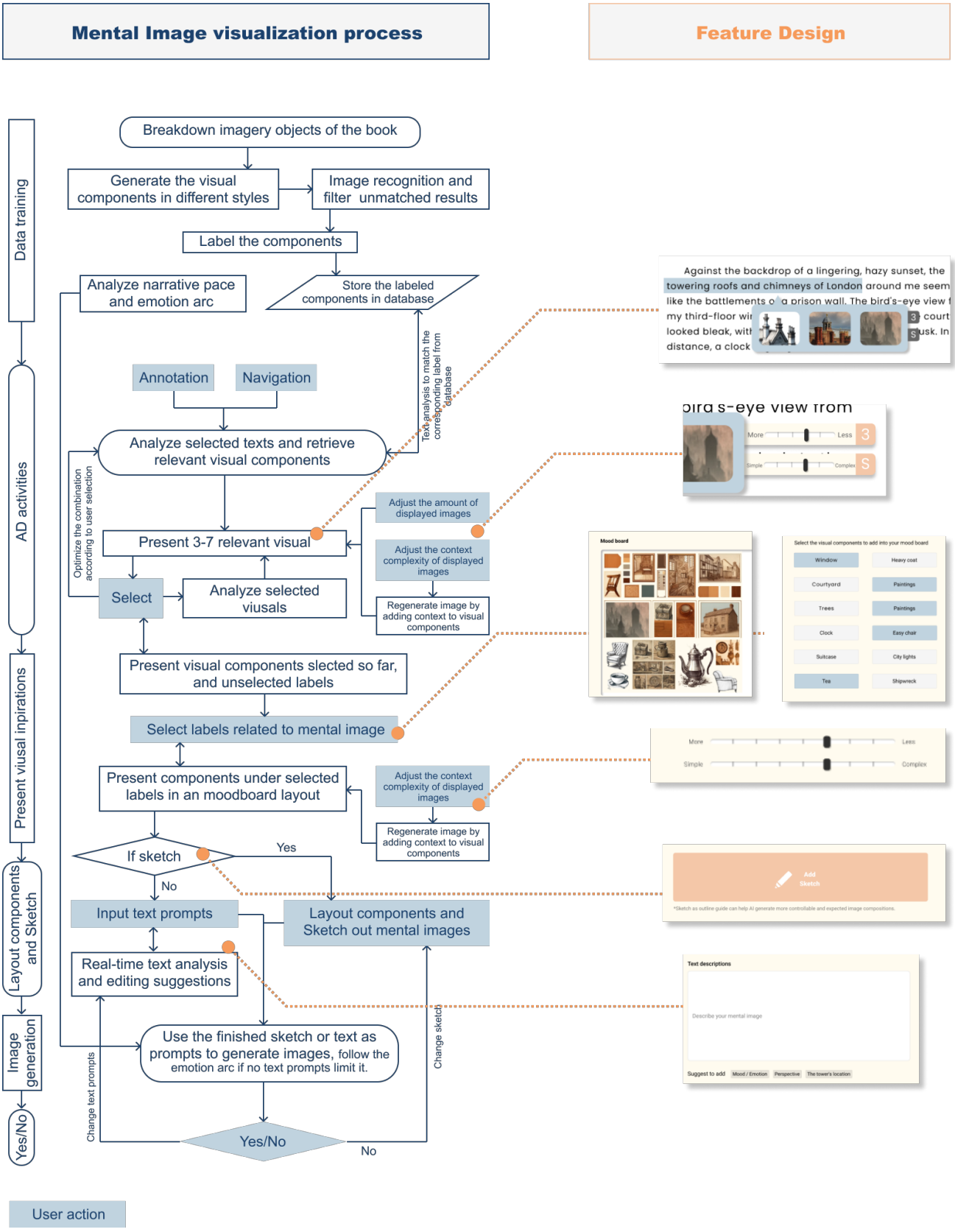


Figure 41: Feature design in the mental image visualization interface

Chapter 8

Design round 2: Service design for the whole active reading experience

In this chapter:

8.1 Design task

Design round 2 extends the design of round 1 to the personalized design of the entire active reading system, in order to create a holistic user experience for active reading in narrative books. This design round aims to build up toolkit personalization in an active reading experience. Therefore, coming back to the design vision defined before (Chapter 4), the design task is:

A human-machine collaborative system for personalized digital reading experience on knowledge construction (and reflection) in libraries for diverse readers who seek for active reading.

8.2 Concept of the product service

The existing mainstream products I researched in Chapter 5.3.7 only support basic note-taking and reading functions, none have features I identify that could improve personalized active reading. Considering the differences between different people in active reading behavior and information visualization of narrative books, the concept of this product service is to provide a human-computer collaborative system with personalized digital reading experience for library knowledge construction (and reflection) for readers with different visual imagery abilities and cognitive experiences who seek active reading.

It can be integrated with the library system to support users to customize active reading auxiliary functions that meet their reading preferences. And use AIGC technology to provide unique visualization mental model functions to enhance user engagement and personalized reading comprehension. The core of the knowledge construction system driven by AIGC is mental image visualization. It presents editable visual inspiration by analyzing narrative books and user preferences, helping users organize mental images and generate the required image output.

Software

The digital reading platform embedded in library reading devices and personal tablets allows active readers to **customize functional modules and reading modes, visualize mental models and share them**. In order to meet all behaviors of user active reading, the digital reading platform will be supported by AI text analysis, conversational AI, AI chart generation, AI image generation, etc.

The library collection pool on the platform displays visualized knowledge construction and promotes services. This is a public knowledge base where patrons can contribute their visualizations, building a collective resource that everyone can benefit from.

Hardware

A tablet with a stylus supports the actual use of the platform. Reading hardware can be a personal device or a public device provided by the library. To ensure a return on the library's investment, the use of the library's tablets will be open only to paying subscribers.

8.3 User Journey Flow

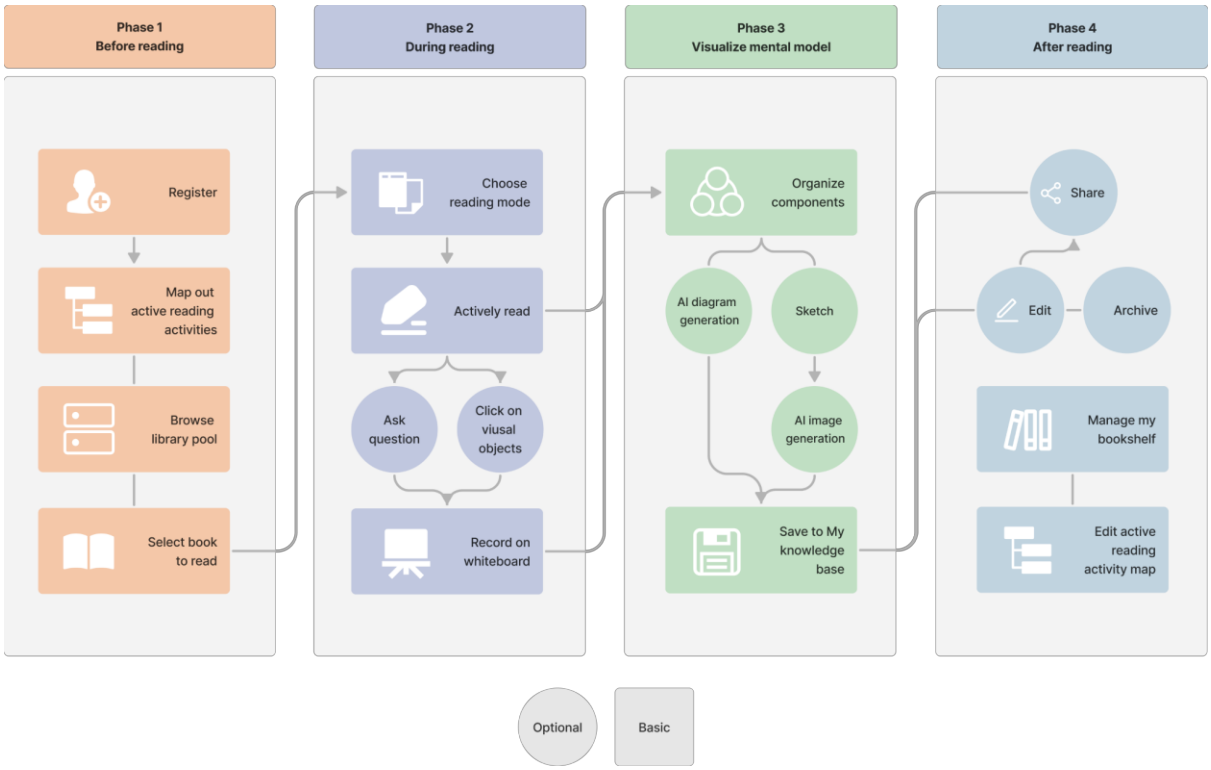


Figure 42: User journey flow of active reading experience

The figure above illustrates the user journey of using the AI powered platform to enhance active reading experiences through visualization (and sharing) of mental models. This platform is integrated into both library reading devices and personal tablets, offering a personalized and interactive reading environment. It also features a library pool to share and display visualized knowledge constructions, promoting a collaborative learning atmosphere.

Phase 1: Before Reading

1. **Register**: Users begin by registering on the platform and linking to the library.
2. **Map Out Active Reading Activities**: They then map out their planned active reading activities.
3. **Browse Library Pool**: Users can browse a library pool to discover available resources and visualized knowledge.
4. **Select Book to Read**: They select a book from a digital library or upload from device to read.

Phase 2: During Reading

1. **Choose Reading Mode**: Users choose their preferred reading mode on the device.
2. **Actively Read**: As they read, they can engage with the content actively.
    - **Annotation, navigation etc. as mapped before.**
    - **Ask Question**: Users can ask questions to deepen their understanding.

- **Click on Visual Objects**: They can click on visual objects for more information.
- **Record on Whiteboard**: The selected text and visual objects, together with thoughts and notes will be automatically recorded on a digital whiteboard alongside the reading page.

Phase 3: Visualize Mental Model

1. **Organize Components**: After gathering information, users organize the components of their mental model.
    - **AI Diagram Generation**: The software can assist in generating diagrams through AI, and allow users to edit them on the whiteboard.
    - **AI Image Generation**: AI tools help generate images that visualize complex concepts.
        - **Sketch**: Users can also create sketches to represent their understanding, as prompts for AI image generation.
2. **Save to My Knowledge Base**: All visualizations and notes can be saved to a personal knowledge base.

Phase 4: After Reading

- **Share**: Users can share their visualized knowledge to friends or to the library pool.
- **Edit**: They have the option to edit their notes and visualizations.
- **Archive**: Readings and visualizations can be archived.
- **Manage My Bookshelf**: Users can manage their reading materials and organize their bookshelf.
- **Edit Active Reading Activity Map**: They can also update their active reading activity map based on their progress and new interests.

The target user journey takes into account the various user needs found during the user ritual research. The figure below shows how the user needs of Chapter 5.3.2 are linked to each link of the system flow.

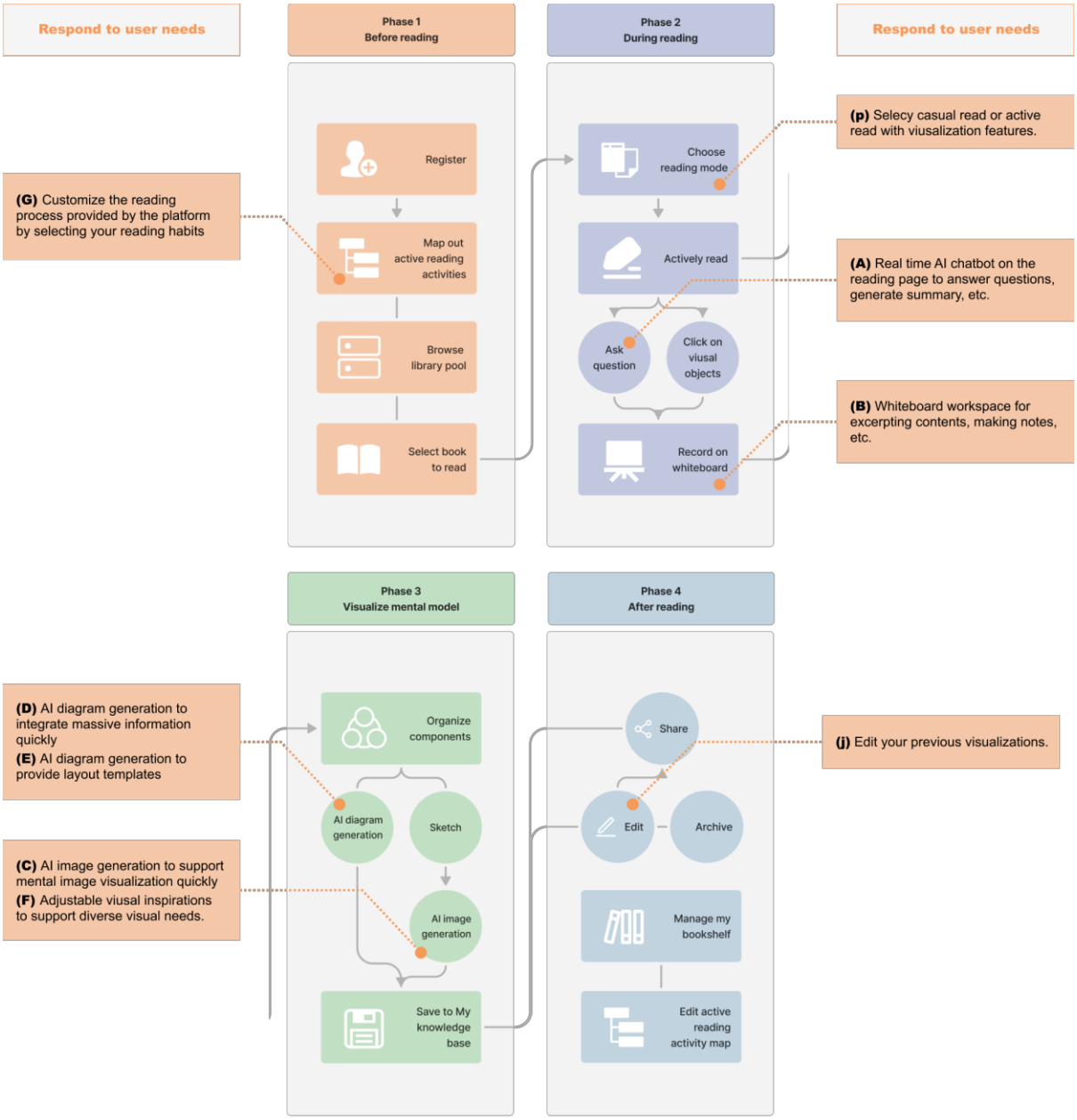


Figure 43: Explanations of how the user journey flow design responding to the user needs of active reading experience

The platform design addresses user needs as follows:

- **Real-time AI chatbot**: Enhances productivity by providing quick, clear answers and summaries without disrupting reading (A). It also helps associate relevant concepts, enriching user understanding.
- **Whiteboard workspace**: Improves efficiency by making note-taking quicker and more organized, thus maintaining engagement (B).
- **AI image generation**: Facilitates rapid mental image visualization, ensuring users get appropriate images with minimal effort (C). The personalized visual assistance: helps users with varying levels of visualization skills to picture and sketch scenarios, boosting inclusivity and active reading (F).
- **AI diagram generation**: Quickly integrates and clarifies massive

information, enhancing productivity and comprehension (D). It also allows multiple output formats for personalization (E).

- **G. Customized reading process**: Allows users to tailor their reading experience based on their habits, promoting a personalized and active reading process (G).

Phase 2 of choosing reading modes is designed for the user need (p) —- visualizing mental image is not necessary and for the recognition of the reader's different active reading process. Only users who choose visual reading mode can add images into their whiteboard.

Phase 4 of editing notes and visualizations is designed for the user need (j) —- they want to maintain the flexibility to imagine different images each time they reflect on the article. Users can always generate new images or edit old images in the saved drafts.

This user journey flow ensures a comprehensive and interactive reading experience, collaborating with AI tools for visualizing mental models, and promoting active reading empowerment and learning efficiency.

8.4 Function architecture and Wireframes

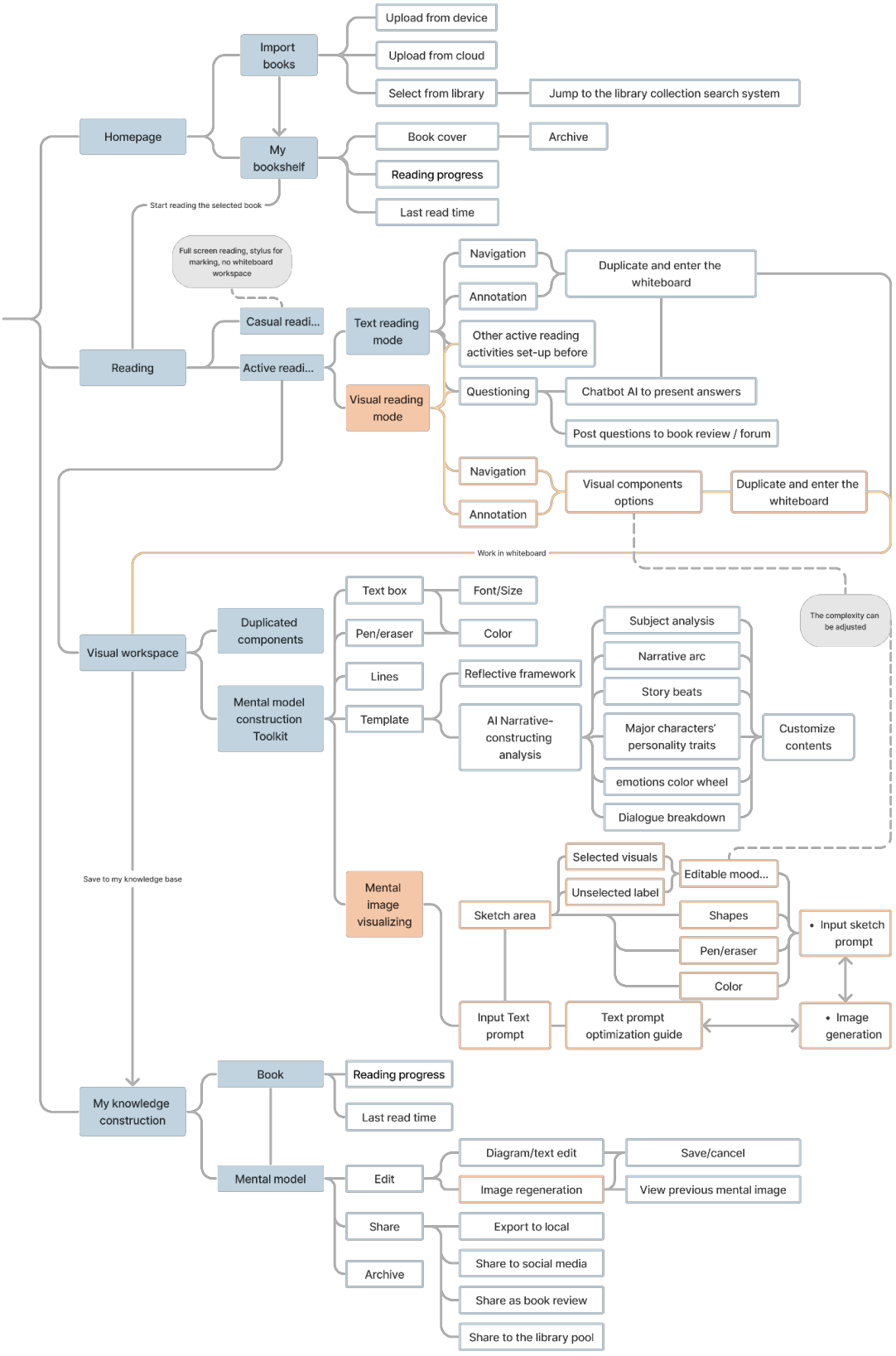


Figure 44: Function architecture of active reading mode

The digital reading platform, embedded within both library reading devices and personal tablets, is designed to enhance the reading experience by allowing active readers to visualize mental models through Artificial Intelligence Generated Content (AIGC). The platform's architecture is composed of several interconnected modules, each serving a series of functions within the reading system. The following wireframes showcase the personalized active reading function and mental image visualization function, the rest of the basic features of Scenebites are in Appendix 6.

Map active reading behaviors into stages

To empower active reading behavior personalization, the platform is designed with various reading assistance functions according to the integrated knowledge construction process from user ritual research (chapter 5.3.5.2). Users need to set up their preferred active reading process on the first log-in.

Figure 45: Wireframe of function - Map active reading behaviors into stages

Reading Modes

The platform offers two primary reading modes: casual read and active read. Casual read mode caters to leisurely reading, providing a simple and distraction-free interface. In contrast, active read mode is designed for in-depth study and engagement, featuring tools for note-taking, highlighting, and annotation. Under active reading mode is its bifurcation into text reading mode and visual reading mode. Text reading mode focuses on traditional text-based interactions, whereas visual reading mode incorporates AIGC to create dynamic visual representations of the text, which will be used as visualization assisting components afterwards.

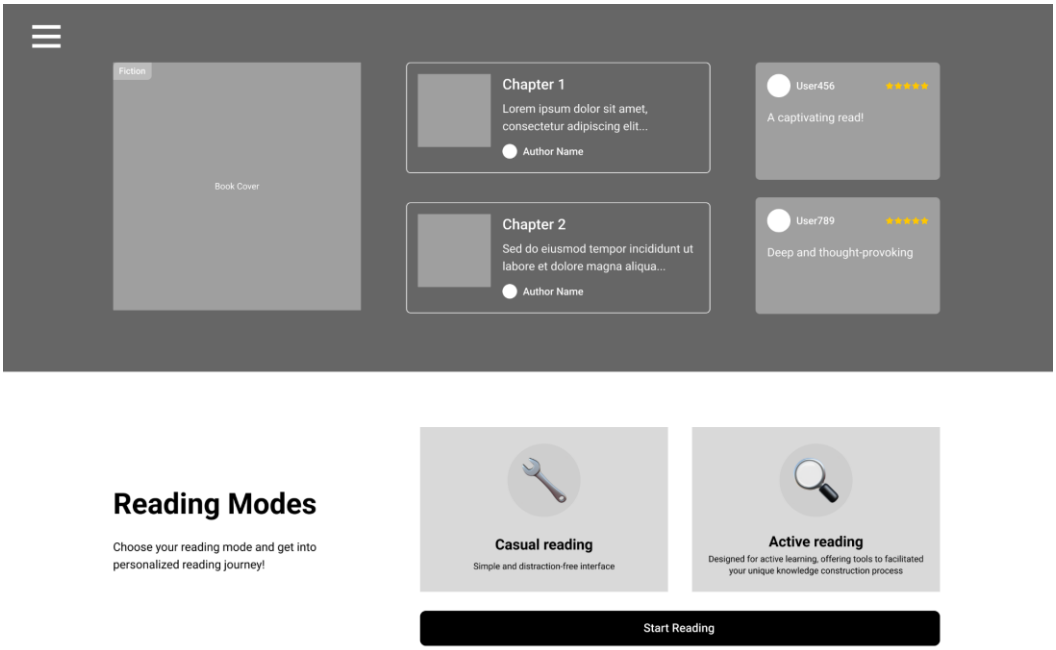


Figure 46: Wireframe of function - choose reading mode

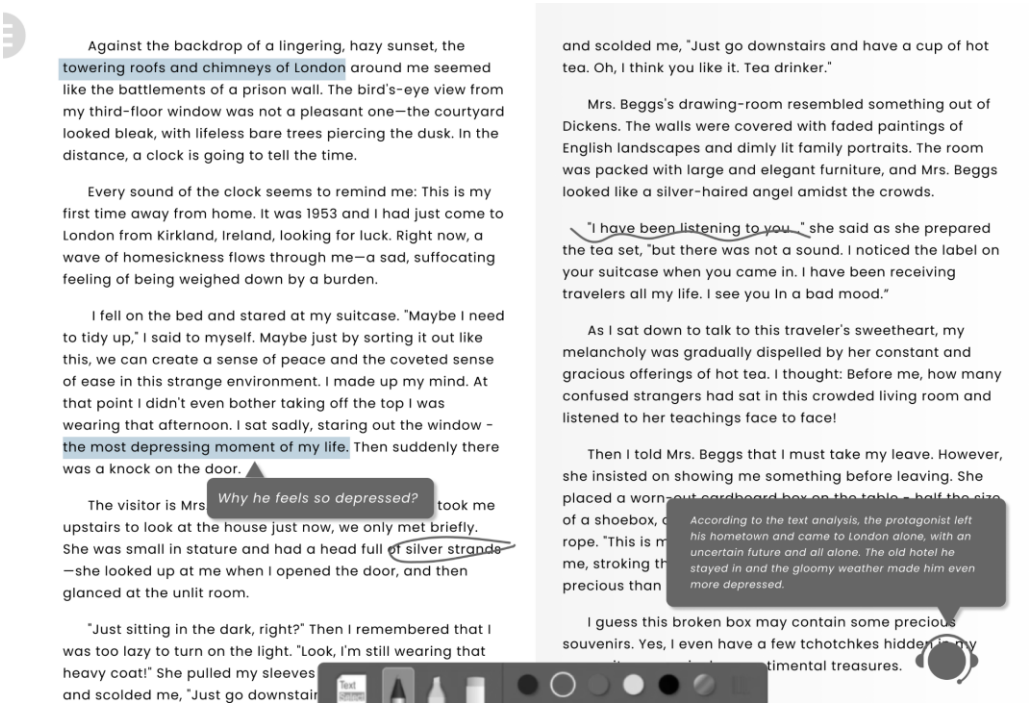


Figure 47: Wireframe of function - casual reading mode

Whiteboard Workspace and Mental Model Construction

The visual workspace is an advanced feature that allows users to manipulate and interact with visual components. This workspace includes duplicated components from user annotation and

navigation, and a mental model construction toolkit. These tools are particularly helpful for constructing detailed mental models that reflect the user's understanding of the material.

AI diagraming - Narrative-constructing analysis

The AI narrative-constructing analysis is accessible in visual reading mode. Under user request, AI will analyze the text to identify key themes, plot points, and character developments, providing readers with summaries and insights that help them construct the narrative structure. In visual reading mode, the AI also utilizes visual aids to illustrate these elements, making complex analyzing results more accessible. Users are able to edit it into a more appropriate version in their own knowledge construction.

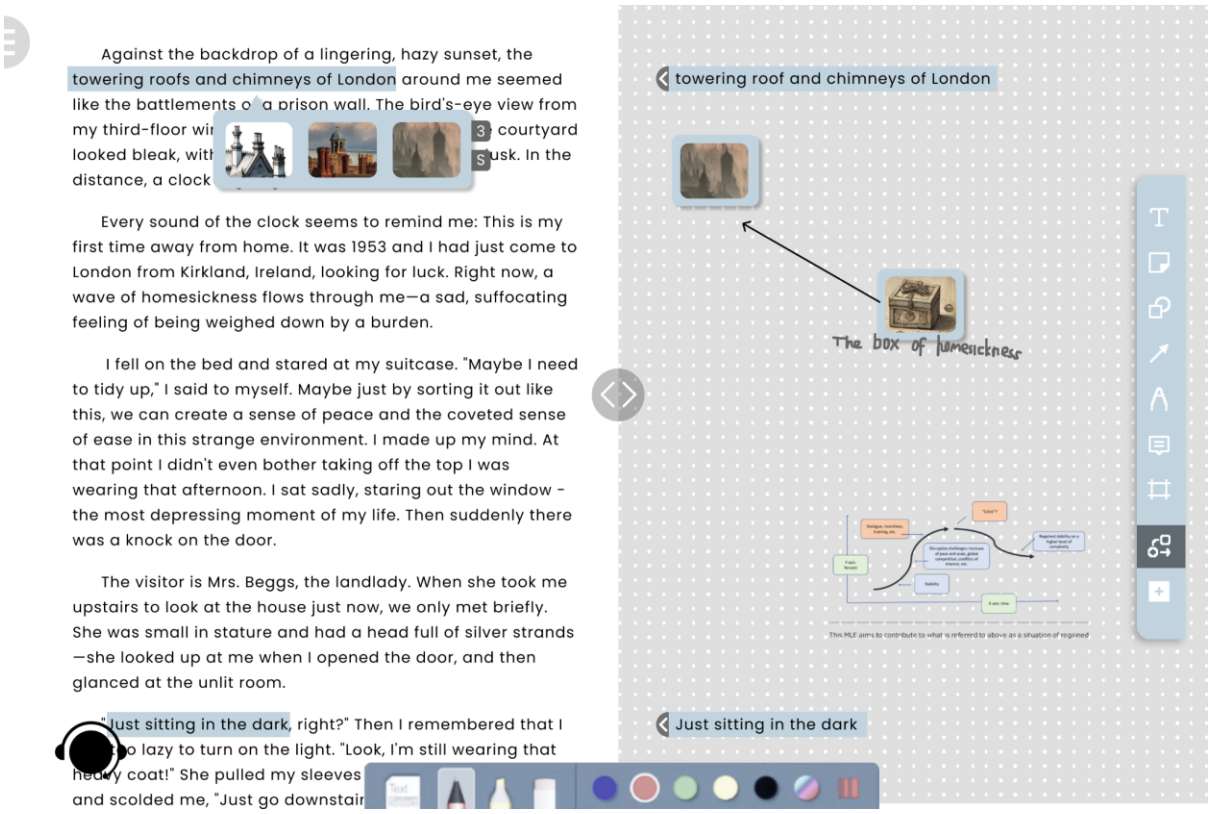


Figure 48: Wireframe of function - active reading mode

Mental Image Visualizing

A standout capability of the platform is its mental image visualizing function, where AIGC generate images based on the reader's input. If the user decides to insert an original image into the mental model, the image generation workspace will be called. Under this function, users can also choose to expand a sketching workspace, where editable visual components will be presented to assist sketching. Finally, the text input and sketch input will serve as prompts for AI to generate images.

This feature is optional, based on the feedback from user testing in round 1 of the design process. It showed that many users did not think that visualizing mental images was a necessary behavior, and for users who hate sketching, it would even greatly increase the cognitive burden. Therefore, the interface will prompt whether to generate an image and whether to draw a sketch.

Besides, text prompts will also be analyzed by AI and suggested to modify to achieve a higher control of the image generation output.

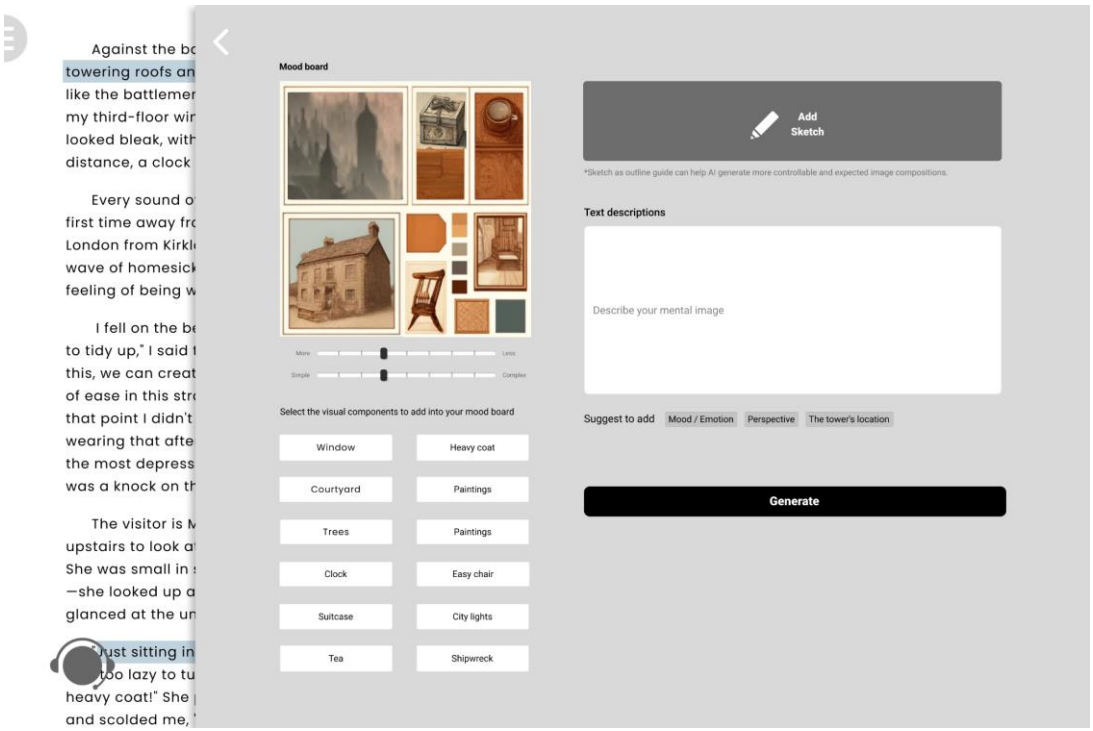


Figure 49: Wireframe of function - generate mental image with AI

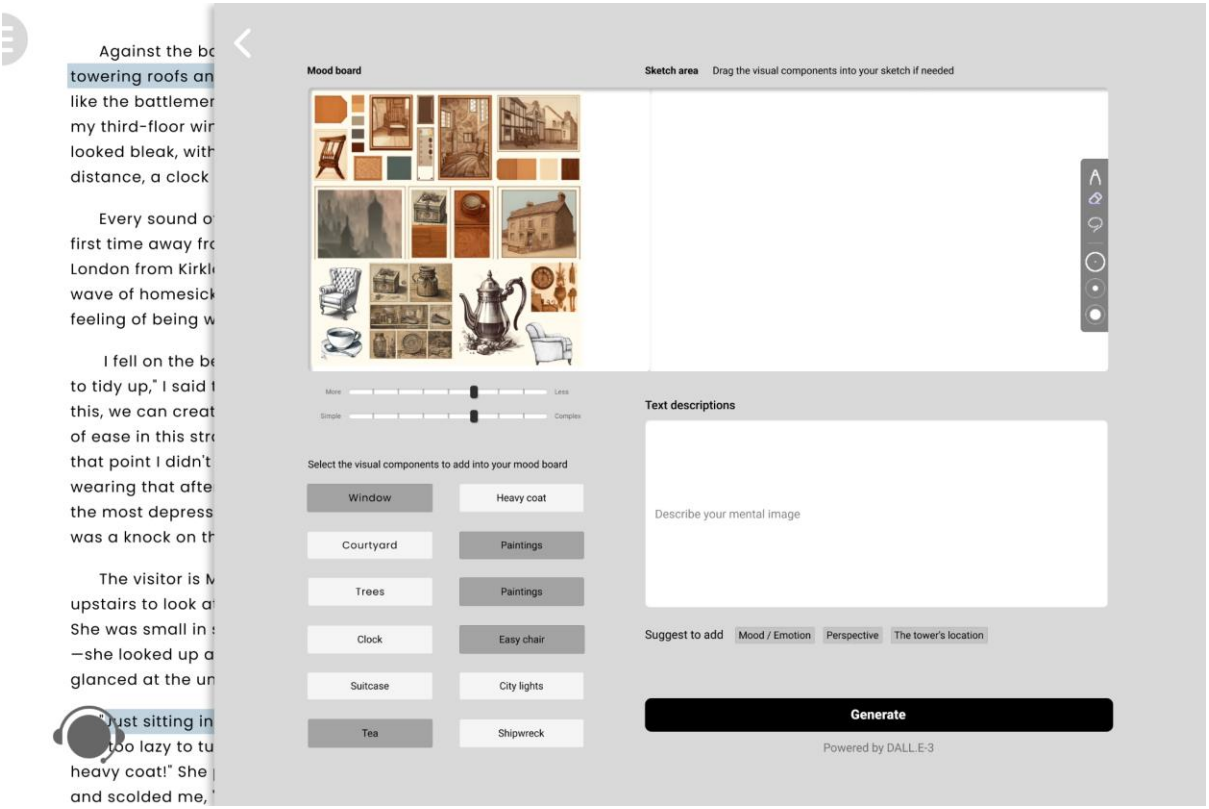


Figure 50: Wireframe of function - add sketch to generate mental image

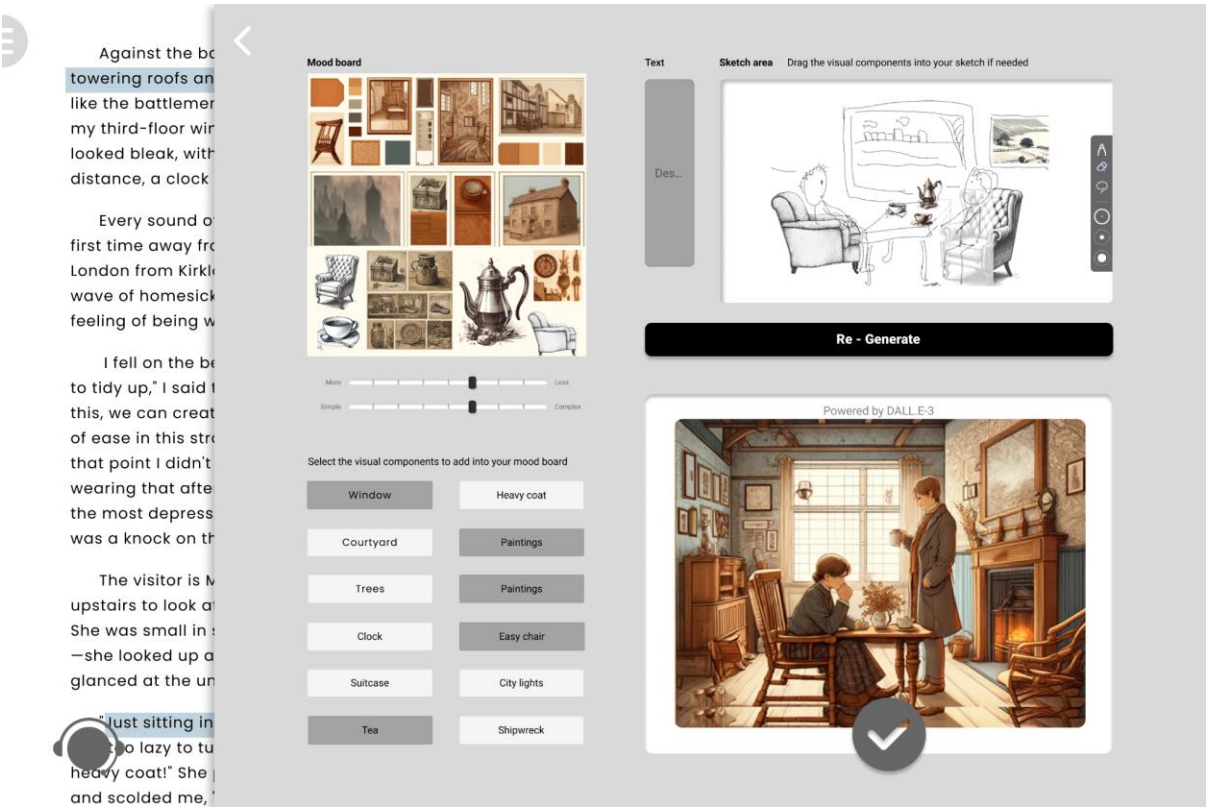


Figure 51: Wireframe of function - mental image generation output

My Knowledge Construction

Users are allowed to save the created visual mental models in their own knowledge construction library. All the contents from the whiteboard will be saved and linked to the corresponding book. Users can share the knowledge construction outcome with personal connections or display in the library pool to communicate with other users.

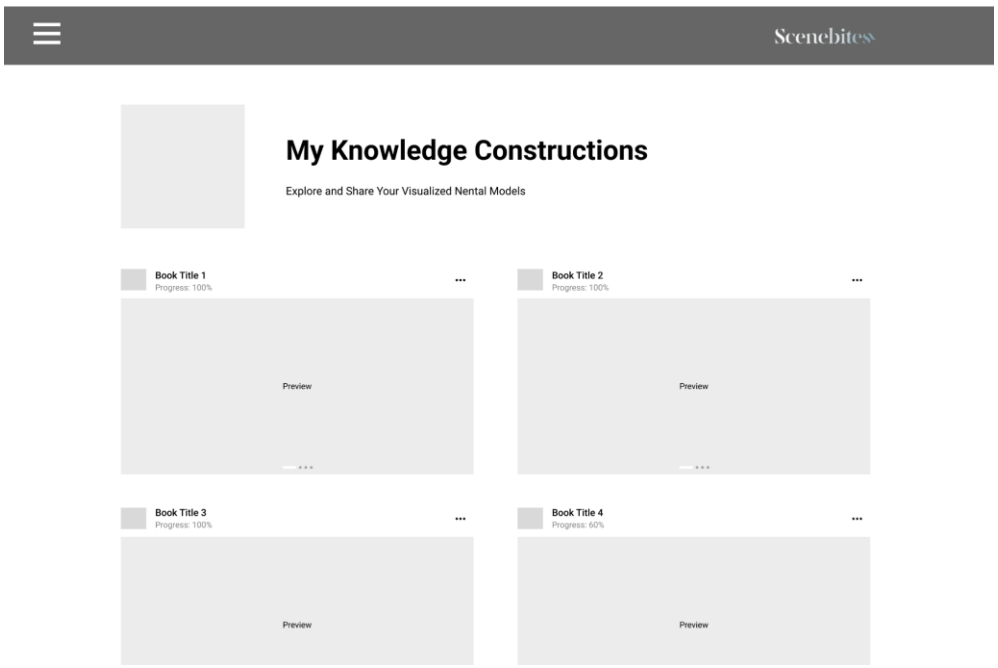


Figure 52: Wireframe of function - my knowledge constructions

Library pool

The library pool is also a key feature of the digital reading platform, to facilitate community engagement, resource sharing, and collaborative learning. It acts as a dynamic repository where users can access, contribute, and promote a wide range of mental model visualizations. The platform encourages users to contribute to the library pool by uploading their own visualizations, annotations, and reflective journals.

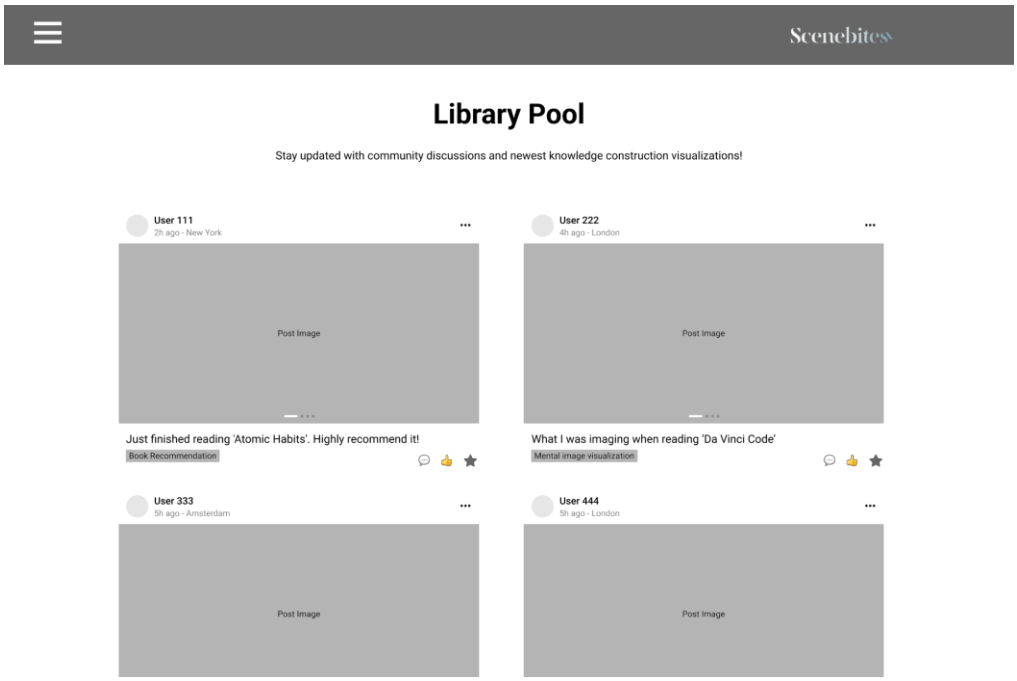


Figure 53: Wireframe of function - Library pool where active reading share visualizations

8.5 Branding

8.5.1 Product name - Scenebites

The name for the human-machine collaborative system for personalized digital reading experience in libraries is "**Scenebites**":

Scene: This part of the name implies visualization and immersion. It suggests that the system creates an active, vivid, and engaging platform for narrative readers. This is especially important for visual readers, as it emphasizes the creation of scenes or mental images that enhance understanding and knowledge construction.

Bites: This indicates the idea of digestible, manageable pieces of knowledge components. In the context of digital reading and knowledge construction, it suggests that the system breaks down complex information into smaller, more comprehensible segments, making it easier for readers to process and reflect on the material.

8.5.2 Visual identification design

Scenebites

Scenebites

Figure 54: logo design of Scenebites

Figure 55: Use scenario display of Scenebites

Figure 56: Use scenario display of Scenebites

8.6 Technological Feasibility Implication

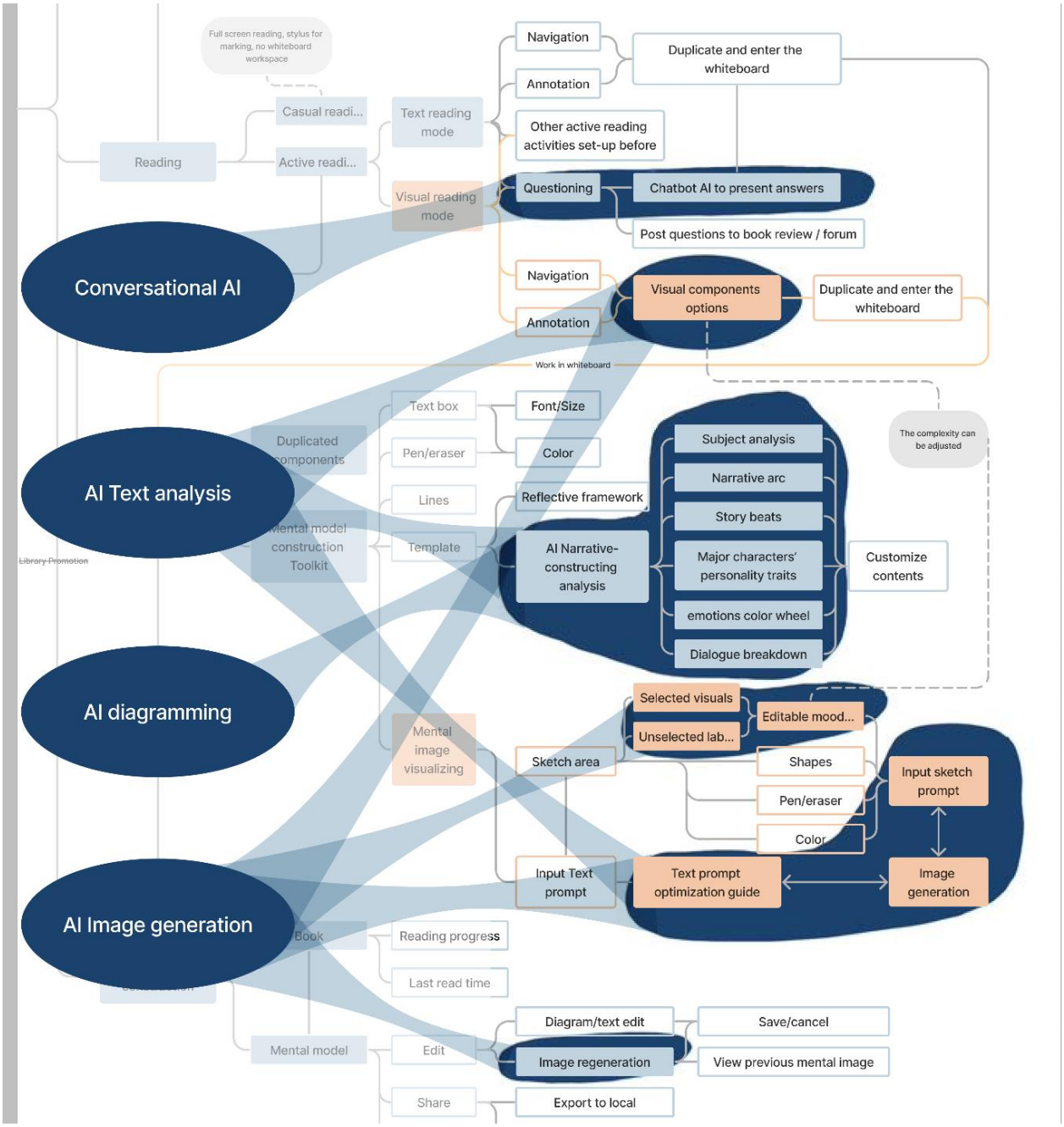


Figure 57: The technological feature in the function architecture

This figure represents the technological features and their interactions within a digital reading platform that supports advanced functionalities like AI-based mental image visualization, text analysis, diagramming, and conversational AI. The technological feasibility of integrating real-time AI conversational capabilities, whiteboard recording, AI text analysis, and diagram generation into digital reading platforms is well-supported by current advancements. Conversational AI, powered by models like GPT-4, enhances user interaction through natural language processing. Real-time whiteboard functions, as seen in platforms like LiquidText, facilitate interactive note-taking and annotations. AI text analysis and narrative construction tools, such as those used by Marlowe and

OddBooks, help in understanding and visualizing complex narratives and emotional arcs. AI-driven diagramming tools like ChatUML and Story Explorer further support the creation of visual aids. The explanation of each technology and their feasibility is in Appendix 8.

8.7 Minimum Viable Product

A minimum viable product, or MVP, is a product with enough features to attract early-adopter customers and validate a product idea early in the product development cycle(Bosch et al., 2013). The purpose of the minimum viable product in this stage is to clearly and effectively demonstrate the actual flow and functionality of the designed system. At the same time, it provides display materials for subsequent user verification to facilitate user understanding.

The minimum viable product I created is a product showcase video with high-fidelity interfaces. This video comprehensively demonstrates the innovative user experience of the Scenebites platform by showing product usage scenarios, product usage requirements, product functions and product interaction effects. The video begins by showing a narrative book reader actively reading in a library, emphasizing that different people construct different mental images of the same text. Then, the active reading support interface of Scenebites is shown, explaining how it uses AIGC and human-computer collaboration to personalize visual mental images and enrich knowledge construction. The video ends by showing the sharing function of the Scenebites platform, and a scene of a narrative book reader sharing his or her insights in a library using Scenebites as a material for self-expression.

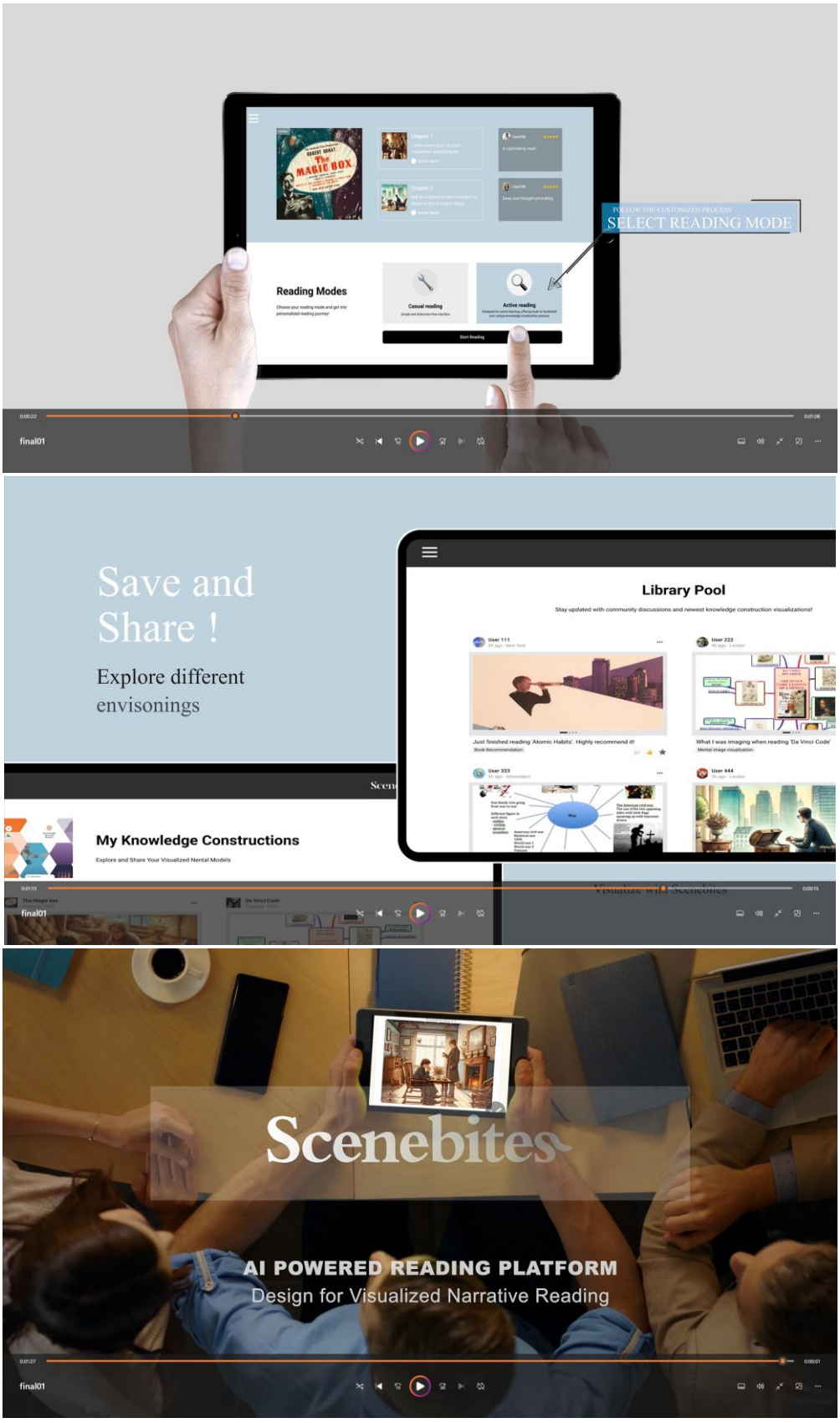


Figure 58: Key frames of the product showcase video

The original video can be viewed via [this link](#).

8.8 User validation

8.8.1 Assessment Criteria

The assessment criteria for evaluating and validating the digital service “Scenebites” powered by AI image-generating tools to visualize mental images in narrative books based on human inputs include several key factors. First, **Active Reading** is assessed through interactivity and knowledge construction. Knowledge construction is further broken down into reducing extraneous cognitive load, strengthening comprehension, enhancing recall, and making discoveries in the knowledge (Victor, 2010). Additionally, **Personalization** evaluates how well the service tailors content to individual users' needs and preferences. **User-Controlled** measures the extent to which users can manipulate and guide the AI's outputs. Lastly, **Interest to try** gauges the overall appeal and willingness of users to engage with the service.

These user validation metrics are necessary for validating the effectiveness and user desirability of Scenebites. **Active reading** in terms of visualized mental image is the key design opportunity, it ensures the achievement of key user experience design. **Knowledge construction** is the key outcome of active reading as elaborated in Chapter 3.2, Scenebites should support deep cognitive processes such as understanding, recalling, and discovering new information in the narrative contexts. **Personalization** is significantly relevant as it is the key user need (Chapter 5.3.2) and design opportunity (Chapter 3.4). It ensures the AI-generated visuals align with different users' mental imagery abilities and preferences, enhancing satisfaction and individuality. **User-Controlled** highlights the importance of user agency, allowing users to dominate the outcome and ensuring the tool complements their creative processes. Lastly, **Interest to Try** is essential for assessing the initial and sustained user desirability of the service, indicating its potential adoption and success among users.

The completed user validation plan is in Appendix 6.

8.8.2 Data collection

User recruitment target is 4 participants with high visual imagery ability and low visual imagery ability (same people I recruited in previous user test -Chapter 7.5.2). The user test method is a focus group, which helps the different types of users communicate and share their experience feelings. Based on the assessment guideline, the data is collected through a mixed-methods approach, combining quantitative rating scales with qualitative follow-up interviews (Schensul & LeCompte, 2012). Quantitative Rating (0-5 Scale) provides a numerical measure of their responses.

Figure 59: User validation focus group for user experience of Scenebites

8.8.3 Results

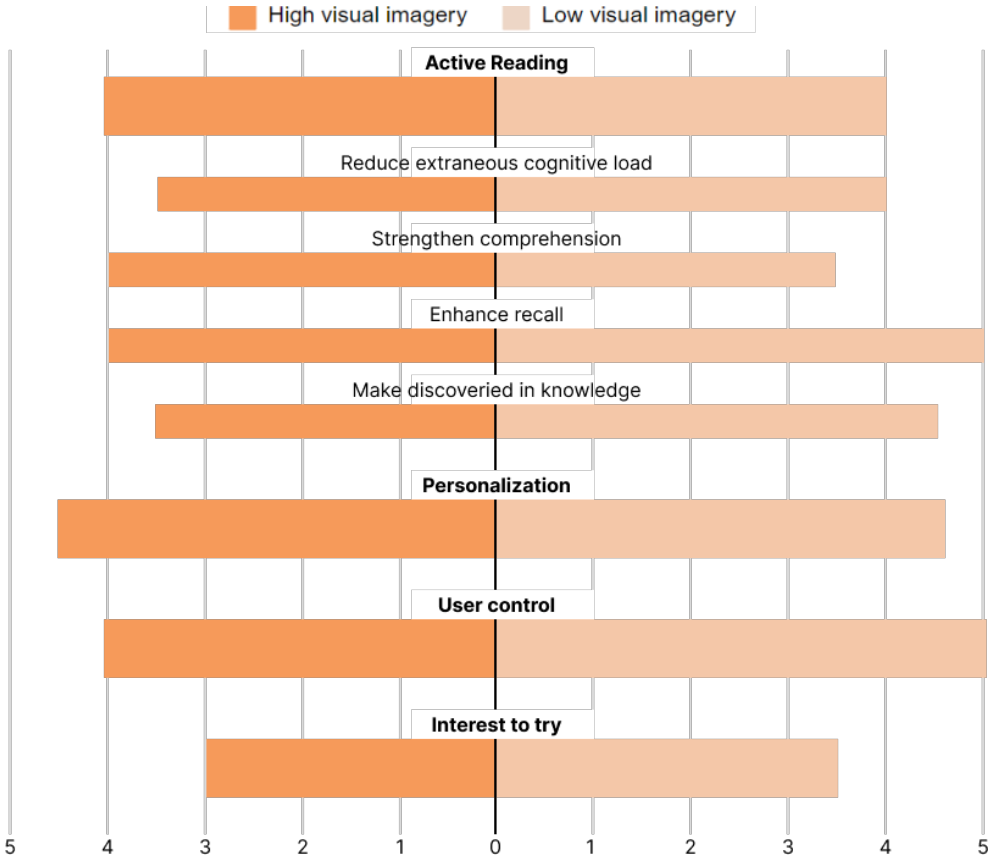


Figure 60: Results of quantitative rating of user satisfaction for Scenebites

From the quantitative rating scales result, the overall user satisfaction is relatively high compared to the user test of Design round 1.

All the assessment criteria are scored over 3.5 besides the interest to try from high visual imagery participants.

In terms of **active reading process facilitation**, all participants think the overall interactivity and utility of this system is smooth and reasonable. The interactive interface is clear and easy to understand, and friendly to novices.

In terms of **reducing extraneous cognitive load**, the optional sketch function and text prompt optimization suggestion function can significantly improve the operational efficiency and prevent thoughts from being pulled away from the narrative itself. And the adjustable mood board helps them to create desired visual assistance. However, for participants with less sketching skills, if they want to add a sketch, it is inevitable to think about how certain elements of the mental image should be represented by lines.

In terms of **strengthening comprehension**, all participants agree that the image generated can help them understand the context in detail, and the process of generating a mental image also pushes them to think deeply into the narrative contents. Besides, the chatbot in the reading interface can help comprehension through answering questions instantly. Moreover, the AI diagramming function can help clarify the deeper logic of the article/character relationships/value transfer, etc. However, some participants mentioned that they were used to thinking about their own experiences while reading to understand the text, and this systematic process did not support this level of understanding.

In terms of **enhancing recall**, all participants think the visualized mental models can help them better recall what they read and what they were thinking about when they read. However, some high-visual imagery users believe that they will have different thoughts about the book when they recall it, thereby updating their mental model, and the preserved knowledge construction may limit this updating behavior.

In terms of **making discoveries in knowledge**, low visual imagery users can feel the discovery of information through choosing visuals and generating vivid images, especially when they can see something surprising by the collaboration with AI. For some high visual imagery users, thinking while reading is more important for knowledge exploration, and the visualization of mental images only plays a supporting role.

In terms of **personalization**, all participants mentioned the design of the entire process has been iteratively personalized, including a plannable active reading process, selectable reading modes, and an adjustable inspiration board. At the same time, the visual options when reading analyze the psychological image preferences of different users in advance, and the movable visual elements assist those with poor drawing skills. People with different visual imagery abilities can easily change the variables of some key features, to make the whole process more comfortable.

In terms of **user control**, almost all aspects of image generation are controlled by the user's subjective input; they largely feel that they are in charge of the creation of this image.

In terms of **Interest to try**, most participants believed that if they were satisfied with the trial experience, they would be willing to subscribe and use the platform to create and share their own mental models. Some users believed that they would only want to use the platform when reading some relatively abstract, unrealistic narrative books, such as science fiction/fantasy novels.

In summary, the system is very effective in promoting an active reading process, and the features are designed to reduce extra cognitive load and enhance comprehension recall. Personalization is a prominent feature, and most features are designed to adapt to various user preferences and abilities, thereby improving inclusivity and usability. Users have

a high degree of control over image generation and are willing to try it out in future narrative book reading experiences.

Therefore, the value proposition and user desirability of Scenebites as a human-AI collaboration platform for active reading are validated.

8.9 Business model

According to Osterwalder's book 'Business Model Generation', the business model is analyzed and elaborated through the following business model canvas, comprising nine business blocks (Osterwalder & Pigneur, 2010):

Figure 61: Business model design of Scenebites

8.9.1 Value delivery

By delivering a platform that allows users to customize their reading experience through functional modules, visualized mental models, and shared knowledge construction, the system enhances user engagement and comprehension of narrative books. To promote and maintain this system, regular updates and improvements to the AI algorithms are needed to enhance the accuracy and utility of text analysis, diagram generation, and image creation. Marketing specialists should advertise the new product service in both library and social media to stimulate the first trial. Besides, regular

exhibitions and events should be organized for ongoing exposure and user feedback collection.

8.9.2 Value creation

The value creation of this system lies in its ability to foster a personalized active reading environment. The adoption of AI text analysis and content generation further enriches the narrative reading experience by providing real-time interactions and mental model visualizations. The system's design ensures that user behavior, preferences, and cognitive abilities are central to the experience, allowing for adaptive visual interpretations and customized features. This approach not only aids in better understanding and retention of information but also encourages active participation and co-creation among readers by library pool and community activities, making the narrative reading more inclusive and interactive.

8.9.3 Value capturing

Considering the large amount of investing in early-stage platform development, hardware configuration, personnel training and service promotion, it's important to design a business strategy to recycle funds for libraries. To ensure the sustainability and profitability of the platform, several revenue streams have been identified. These diverse streams not only ensure financial viability but also support continuous improvement and expansion of the platform

Licensing fees from libraries: integrating the platform into future library systems provide another significant revenue source. Additionally, revenue can be captured through co-creation events where readers and authors collaborate, as well as potential advertising within the platform., ultimately delivering a high-value service to users and stakeholders.

Paid Subscriptions for Library Devices: Access to the library's tablets and the visualization features of the platform is available through a paid subscription. Paid subscriptions generate a steady income, offsetting the costs associated with hardware and software development.

Reader-author co-creation activities: With enough data sharing of knowledge construction visualizations, revenue can be captured through co-creation events where readers and authors collaborate on new illustration creation, film creation, exhibition creation, etc.

Library exhibitions: Regular exhibitions of mental images for designated narrative writers can generate revenue for the library.

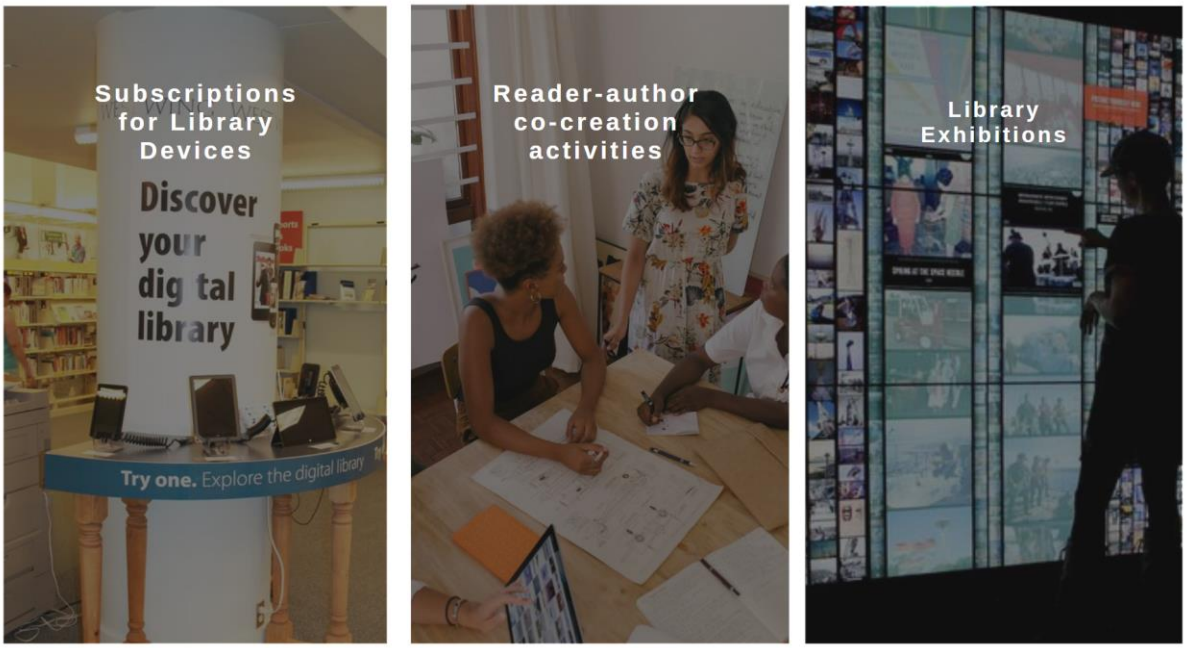


Figure 62: Potential funding activities that libraries can take

More Implementation suggestions for future libraries are in Appendix 9.

Chapter 9

Discussion

In this chapter:

The active reading system design realizes the design vision of enhancing the active reading experience in the library of the future by creating a collaborative human-computer interface that prioritizes personalized digital reading. The system is structured to promote active reading by enabling the unique knowledge-building process of narrative books and readers. It does this through personalized reading modes and image generation processes that work with each user's visual image capabilities. The system inherits from the future library environment through rentable facilities such as tablets and styluses, supporting digital transformation through enhanced accessibility and interactivity. Its human-centered approach ensures that the design process focuses on human needs and rituals, translating user needs into design points while enabling readers to maintain control of the mental model visualization process. Personalization is key, with AI-generated inspiration boards and text prompts optimizing suggestions that can be adjusted based on personal input, preferences, and knowledge level. This flexibility helps users actively and effectively build and reflect on knowledge. In addition, the system is future-proof, using cutting-edge technological advances to provide innovative and productive reading experiences.

User interaction with AI-powered mental model visualization tools revealed mixed feedback regarding user experience and satisfaction. After design and design iteration, and integrating two rounds of user survey feedback, the Scenebites system process finally achieved very positive results in user validation. Scenebites as an advanced active reading system fills the gap in mental image visualization in current market products and brings a more personalized and efficient active reading experience. However, not all the user needs are responded to in the design iteration. User need (b) — the final output will automatically be added with details not mentioned in the prompts, and the user need (g) — the final output cannot always follow every detail of the prompts, are not answered in the design iteration. This is due to the limitations of the DALL.E system itself in the prototype design. It usually automatically adds or modifies scene elements in the picture based on big data analysis. This is currently difficult to iterate through user experience process design. Hope that more technical experts will optimize the capability of AI image generation in the future.

Moreover, the integration of AI-generated imagery into reading experiences raises important ethical and creative considerations. It is crucial to establish guidelines that ensure AI tools complement human creativity rather than overshadow it. Setting creative constraints and ethical guidelines for AI outputs can help align the technology with human values and objectives. Additionally, fostering collaboration between humans and AI in creative processes can lead to more innovative outcomes, as AI tools can enhance human creativity by providing new perspectives and possibilities

9.1 limitations

First, there are some conceptual limitations in my study, which may cause conflicts between literature review and practical applications.

The literature research identified that mental images are highly subjective and vary significantly among individuals (Brosch, 2018). This ambiguity presents a challenge when using AI tools to generate visuals that are meant to align with a reader's mental model. While AIGC models can produce high-quality images, they cannot fully capture the unique and personal nature of each reader's imagination, sometimes even the user himself is not very sure about every element in his

mind. The literature review emphasized the importance of personalized visualization in narrative comprehension, yet AI-generated images may inadvertently standardize or homogenize this process, potentially stifling the individuality of readers' interpretations.

The concept of enhancing reading with AI-generated images aims to optimize cognitive load and facilitate comprehension. However, the introduction of external visualizations can sometimes increase unnecessary cognitive load, especially if the visuals are perceived as complex or inconsistent with the reader's mental images. The literature research identified that active reading involves cognitive strategies that are highly individualistic (McCarthy et al., 2021; Schnotz & Rasch, 2005; Verhoeven & Arthur, 2008). Therefore, adding AI-generated visuals might conflict with existing cognitive processes, creating an additional layer of complexity that not all readers can easily navigate.

Second, it’s still under consideration that AI can provide accurate responses to indicate it understands human input and requests. Especially for the prompts I create in customized GPT as a key tool to integrate AI in my design prototype, AI only executes my instructions through data retrieval and large language models. There was no analytical code architecture behind it for specific instruction functions. AI, particularly large language models like GPT-4, fundamentally operate by generating responses based on patterns and information from vast datasets. They don't inherently possess analytical code architecture but can simulate understanding and execute instructions through a combination of pattern recognition and probabilistic reasoning.

Third, given the stylistic impositions of tools (each having its own distinct aesthetic and style), choosing a single AI image generation tool may limit the ways in which readers may imagine the visuals. While tools like DALL-E, Stable Diffusion, and Midjourney are powerful in generating diverse and high-quality visuals, each comes with distinct stylistic biases that can influence the final output. As I researched before in Chapter 7.2.1, DALL-E often provides surreal and imaginative elements, generating images that may prioritize novelty over realism. Stable Diffusion, on the other hand, provides more photorealistic and detailed results, which might align with specific narrative genres but could limit interpretative flexibility. Midjourney often creates artistic and abstract visuals, providing stylistic flair that may appeal to certain audiences while potentially constraining those seeking more literal representations (Jie et al., 2023). These inherent aesthetic tendencies of AI tools can subtly guide readers' interpretations, potentially constraining their imagination by imposing pre-defined stylistic choices that may not align with their personal mental models. Consequently, while DALL-E is selected as the user testing tool with better image generation quality and speed, there is a risk of narrowing readers' creative engagement by favoring certain visual styles over others. Addressing this challenge involves incorporating mechanisms for diverse stylistic options, allowing readers to tailor visuals that best match their imaginative preferences.

Fourth, the limitation regarding the sample size in user test and user validation causes constraint on the reliability and generalizability of the findings. This limited sample size affects both qualitative and quantitative aspects of the research. In qualitative research, a larger and more diverse group of participants allows for a richer and more nuanced understanding of user experiences, preferences, and challenges. Meanwhile, in quantitative research, a small sample size reduces the statistical power of the study, making it difficult to draw definitive conclusions or identify significant patterns and correlations. Moreover, the scarcity of participants with low visual imagery abilities, and the absence of individuals with extremely

low visual imagery, means that the study may not fully capture the range of user interactions with AI-generated images. This oversight could lead to skewed results that do not accurately reflect the needs and preferences of all user groups, particularly those who may benefit the most from enhanced visual aids.

Fifth, the digital prototype is created via researcher's manually connecting GPT with the user webpage. The product prototype does not have actual usability because the API is not embedded in the interactive interface through code construction and data exchange. Since Scenebites is a first-generation design idea and has not been verified by a sufficient number of users and market feasibility, it is unnecessary to invest a lot of energy in developing a fully automated product at this stage. In the two rounds of user testing of system and service design, the tester, as the operator of the background AI, controlled the user interface output and was able to simulate the product effect and convey the product intention to the participants.

Sixth, due to the current limitations of AIGC computing power and technological development, the loading time of image generation in this design is relatively long in actual applications, usually between 8-15 seconds. Most of the subjects in a round of user testing felt that they waited for a long time, and this waiting time caused some interruption to their mental image memory and continuous reading experience. Efficiency, as a key user requirement of this design, is difficult to be audited at this point because it is difficult for users to ignore the loading time of AI responses. However, improving AIGC computing power requires the joint efforts of designers and artificial intelligence technology researchers, and it is difficult to rely on the user experience design in this project to solve this technical problem.

9.2 Future implications

Future research should focus on improving the alignment of AI-generated images with user expectations and mental models. Technological advancements in AI algorithms could enhance this alignment by enabling more nuanced interpretations of textual descriptions. Embedding analytical functions in platform building is recommended to ensure AI actually executes the specific instructions, not only just pretend to understand human inputs. Software engineers can combine AI with traditional programming to handle analytical or computational tasks. For example, use AI for natural language understanding and a separate codebase for executing specific functions. Software engineers can also create a pipeline where the AI processes input, extracts relevant information, and passes it to a backend system for further processing. Besides, they can utilize APIs to connect the AI model with library systems or digital book databases for dynamic data retrieval and processing.

With the continuous development of technology and methods, the efficiency of AI image generation will undoubtedly become higher and higher (Hernandez & Brown, 2020). This means not only the improvement of generation speed, but also the improvement of generation quality and the optimization of resource utilization. With the development of hardware technology and the application of cloud computing and distributed computing, especially the performance improvement of dedicated processors such as GPU and TPU, the training and reasoning speed of AI models will be significantly improved (Hodak & Dholakia, 2021). At the same time, more advanced deep learning models and algorithms will continue to be developed. These new models will greatly reduce the demand for computing resources

while maintaining or improving the quality of generated images (Geetha et al., 2023).

The stylistic biases inherent in AI image generation tools suggest a need for greater flexibility and customization in future iterations of these technologies. To mitigate the potential constraint on user imagination, the future Scenebites system should try to integrate different AI tools in the user interface and offer a wider range of stylistic options to allow users to choose or customize the aesthetic of generated images.

For this design project, Scenebites as a new software tool and platform will simplify the deployment and use of AI models and reduce the time and resources required to generate images. For example, advanced analysis and visual extraction of narrative books will provide a large amount of learning material for subsequent image generation, and provide constraints on the controllability of the output.

Additionally, expanding the scope of user testing to include a diverse range of participants will provide deeper insights into the effectiveness of AI-generated imagery across different cultural backgrounds and cognitive styles. At the same time, it’s recommended to gradually explore the market feasibility and the potential to collaborate with pilot libraries. Finally, exploring the integration of AI-generated imagery with other media forms, such as audio and video, could create more immersive reading experiences and further enrich the user's engagement with the narrative

References

Adar, E., Gearig, C., Balasubramanian, A., & Hullman, J. (2017, May). PersaLog: Personalization of news article content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3188-3200).

Ali, S., Ravi, P., Williams, R., DiPaola, D., & Breazeal, C. (2024, March). Constructing dreams using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 21, pp. 23268-23275).

Askwall, S. (1985). Computer supported reading vs reading text on paper: a comparison of two reading situations. International Journal of Man-Machine Studies, 22(4), 425-439.

Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press, USA.

Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Bosch, J., Holmström Olsson, H., Björk, J., & Ljungblad, J. (2013). The early stage software startup development model: a framework for operationalizing lean principles in software startups. In *Lean Enterprise Software and Systems: 4th International Conference, LESS 2013, Galway, Ireland, December 1-4, 2013, Proceedings* (pp. 1-15). Springer Berlin Heidelberg.

Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. arXiv preprint arXiv:1906.05317.

Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. Science advances, 6(32), eaba2196.

Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. Science advances, 6(32), eaba2196.

Brahman, F., & Chaturvedi, S. (2020). Modeling protagonist emotions for emotion-aware storytelling. arXiv preprint arXiv:2010.06822.

Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters. *Library & information science research*, *22*(4), 371-392.

Brosch, R. (2018). What we 'see'when we read: Visualization and vividness in reading fictional narratives. Cortex, 105, 135-143.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of educational psychology*, *98*(1), 182.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. Journal of economic behavior & organization, 81(1), 1-8.

Chen, X. (2023). ChatGPT and its possible impact on library reference services. Internet Reference Services Quarterly, 27(2), 121-129.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182.

Conlen, M., Kale, A., & Heer, J. (2019, June). Capture & analysis of active reading behaviors for interactive articles on the web. In *Computer Graphics Forum* (Vol. 38, No. 3, pp. 687-698).

Cox, G. C., Smith, D. L., & Rakes, T. A. (1994). Enhancing comprehension through the use of visual elaboration strategies. Literacy Research and Instruction, 33(3), 159-174.

Crane, G., Bamman, D., Cerrato, L., Jones, A., Mimno, D., Packel, A., ... & Weaver, G. (2006, September). Beyond digital incunabula: Modeling the next generation of digital libraries. In International Conference on Theory and Practice of Digital Libraries (pp. 353-366). Berlin, Heidelberg: Springer Berlin Heidelberg.

D'Angiulli, A., Runge, M., Faulkner, A., Zakizadeh, J., Chan, A., & Morcos, S. (2013). Vividness of visual imagery and incidental recall of verbal cues,

when phenomenological availability reflects long-term memory accessibility. Frontiers in psychology, 4, 1.

Daniel, D. B., & Woody, W. D. (2013). E-textbooks at what cost? Performance and use of electronic v. print texts. *Computers & education*, *62*, 18-23.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of educational research*, *61*(2), 239-264.

Eppler, M. J. (2006). A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information visualization*, *5*(3), 202-210.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Ferrante, M., Ozcelik, F., Boccato, T., VanRullen, R., & Toschi, N. (2023). Brain Captioning: Decoding human brain activity into images and text. *arXiv preprint arXiv:2305.11560*.

Garnham, A., & Oakhill, J. (2014). The mental models theory of language comprehension. In *Models of understanding text* (pp. 313-339). Psychology Press.

Geetha, P., Shashikala, K. S., & Kumar, N. A. (2023, July). AI-Enabled Edge Computing Models: Trends, Developments, and Future Implications. In 2023 2nd International Conference on Edge Computing and Applications (ICECAA) (pp. 63-67). IEEE.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological review, 101(3), 371.

Haiken, M., & Furman, L. R. (2022). *Personalized Reading: Digital strategies and tools to support all learners*. International Society for Technology in Education.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-Holland.

Hermann, E. (2022). Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective. *New media & society*, *24*(5), 1258-1277.

Hernandez, D., & Brown, T. B. (2020). Measuring the algorithmic efficiency of neural networks. *arXiv preprint arXiv:2005.04305*.

Hertog, P.D. (2000), “Knowledge intensive business services as co-producers of innovation”, International Journal of Innovation Management, Vol. 04 No. 04, pp. 491-528, doi: 10.1142/s136391960000024x

Hervieux, S., & Wheatley, A. (2021). Perceptions of artificial intelligence: A survey of academic librarians in Canada and the United States. The Journal of Academic Librarianship, 47(1), 102270.

Hodak, M., & Dholakia, A. (2021, December). Recent efficiency gains in deep learning: Performance, power, and sustainability. In *2021 IEEE international conference on big data (big data)* (pp. 2040-2045). IEEE.

Hohman, F., Conlen, M., Heer, J., & Chau, D. H. P. (2020). Communicating with interactive articles. *Distill*, *5*(9), e28.

Holyoak, K. J. (1982). An analogical framework for literary interpretation. *Poetics*, *11*(2), 105-126.

Homel, L. I. N. D. L. E. Y. (2017). Data visualization. *ACRL’s Instruction Section, Tips and Trends*.

Imran, A. S., & Kowalski, S. J. (2014). HIP–A technology-rich and interactive multimedia pedagogical platform. In *Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences: First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 1* (pp. 151-160). Springer International Publishing.

Jayavadivel, R., Arunachalam, M., Nagarajan, G., Shankar, B. P., Viji, C., Rajkumar, N., & Senthilkumar, K. R. (2024). Historical Overview of AI Adoption in Libraries. In AI-Assisted Library Reconstruction (pp. 267-289). IGI Global.

Jayawardana, C., Hewagamage, K. P., & Hirakawa, M. (2001). Personalization tools for

active learning in digital libraries. *MC Journal: The Journal of Academic Media Librarianship*, *8*(1), 1-19.

Jie, P., Shan, X., & Chung, J. (2023). Comparative Analysis of AI Painting Using [Midjourney] and [Stable Diffusion]-A Case Study on Character Drawing. *International Journal of Advanced Culture Technology*, *11*(2), 403-408.

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: an interdisciplinary synthesis of theory and methods. Ecology and society, 16(1).

Kant, I. (1787). On a supposed right to lie from benevolent motives. *The critique of practical reason and other writings in moral philosophy*, *1949*, 346-350.

Kim, E., & Klinger, R. (2018). A survey on sentiment and emotion analysis for computational literary studies. arXiv preprint arXiv:1808.03137.

Kim, N. W., Bach, B., Im, H., Schriber, S., Gross, M., & Pfister, H. (2017). Visualizing nonlinear narratives with story curves. *IEEE transactions on visualization and computer graphics*, *24*(1), 595-604.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

Kolkman, M. J., Kok, M., & Van der Veen, A. (2005). Mental model mapping as a new tool to analyse the use of information in decision-making in integrated water management. *Physics and chemistry of the earth, Parts A/B/C*, *30*(4-5), 317-332.

Kowsalya, V., Sundari, A. S., & Divya, C. (2023). An analysis of the development of pictorial storytelling. In *Artificial Intelligence, Blockchain, Computing and Security Volume 2* (pp. 277-281). CRC Press.

Leonhard, G. (2016). *Technology vs. Humanity: The coming clash between man and machine*. Lodge, UK: FutureScapes.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684*.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. Computers in Human Behavior, 25(2), 284-289.

Mackey, M. (2019). Visualization and the Vivid Reading Experience. Jeunesse: Young People, Texts, Cultures, 11, 38 - 58.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251-283.

Masson, D. (2023). Transforming the Reading Experience of Scientific Documents with Polymorphism.

Mayer, R. E. (1984). Aids to text comprehension. *Educational psychologist*, *19*(1), 30-42.

Mayer, R. E. (2014). Introduction to multimedia learning.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

McCarthy, K. S., Magliano, J. P., Levine, S. R., Elfenbein, A., & Horton, W. S. (2021). Constructing Mental Models in Literary Reading: The Role of Interpretive Inferences. *Handbook of Empirical Literary Studies*, 85-118.

Merrill, M. D. (2000, December). Knowledge objects and mental models. In *Proceedings International Workshop on Advanced Learning Technologies. IWALT 2000. Advanced Learning Technology: Design and Development Issues* (pp. 244-246). IEEE.

Noh, Y. (2015). Imagining library 4.0: Creating a model for future libraries. The Journal of Academic Librarianship, 41(6), 786-797.

O'Hara, K. (1996). Towards a typology of reading goals. Tech. Rep. EPC–1996–107, Rank Xerox Research Centre, Cambridge UK.

Okunlaya, R. O., Syed Abdullah, N., & Alias, R. A. (2022). Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*, *40*(6), 1869-1892.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers* (Vol. 1). John Wiley & Sons.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford university press.

Persig, R. M. (1974) Zen and the Art of Motorcycle Maintenance: An Inquiry into Values, New York, USA: William Morrow & Company.

Phelan, J., PhD. (2022). Cognitive Load isn't all bad! - Julia Phelan, Ph.D - Medium. *Medium*. https://juliaphelan.medium.com/cognitive-load-isnt-all-bad-bdc7473a94fb

Rahimi, M., Beyzavi, G., & Saffari, F. (2021). Effects of a personalized reading software program on the development of language learners' reading comprehension. *Quarterly of Iranian Distance Education Journal*, *3*(1), 9-22.

Rodak, M., & Storey, B. (2022). Prose Fiction: An Introduction to the Semiotics of Narrative.

Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and instruction*, *14*(3), 373-408.

Saux, G., Britt, M. A., Vibert, N., & Rouet, J. F. (2021). Building mental models from multiple texts: How readers construct coherence from inconsistent sources. *Language and Linguistics Compass*, *15*(3), e12409.

Schensul, J. J., & LeCompte, M. D. (2012). Essential ethnographic methods: A mixed methods approach (Vol. 3). Rowman Altamira.

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, *53*(3), 47-58.

Schwartz, J. (2018). Visual literacy: academic libraries address 21st century challenges. *Reference Services Review*, *46*(4), 479-499.

Story structure: 7 types All writers should know. (z.d.-b). Reedsy. https://blog.reedsy.com/guide/story-structure/#3__three_act_structure

Subramonyam, H., Seifert, C., Shah, P., & Adar, E. (2020, April). Texsketch: Active diagramming through pen-and-ink annotations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14453-14463).

Tashman, C. S., & Edwards, W. K. (2011, May). LiquidText: a flexible, multitouch environment to support active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3285-3294).

Technology, V. (2023, 25 september). Principles of Generative AI Design - Vidhema Technology - Medium. Medium. https://medium.com/@viddeveloper0016/principles-of-generative-ai-design-688eb2396224

Van Boeijen, A., Daalhuizen, J., Van Der Schoor, R., & Zijlstra, J. (2014). Delft design guide: Design strategies and methods.

Van Erkel, P. F., & Thijssen, P. (2016). The first one wins: Distilling the primacy effect. *Electoral Studies*, *44*, 245-254.

Verhoeven, L., & Graesser, A. (2008). Cognitive and linguistic factors in interactive knowledge construction. *Discourse Processes*, *45*(4-5), 289-297.

Verma, M. (2023). Novel study on AI-based chatbot (ChatGPT) impacts on the traditional library management. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, *7*(1), 961-964.

Victor, B. (2011). Explorable explanations. *Online. http://worrydream. com/ExplorableExplanations*, *17*.

Wang, L., Chen, W., Yang, W., Bi, F., & Yu, F. R. (2020). A state-of-the-art review on image synthesis with generative adversarial networks. Ieee Access, 8, 63514-63537.

Williams, R. (2023, 20 november). Impact of Aphantasia on the reading experience. Aphantasia Network. https://aphantasia.com/article/science/reading-experience/

Winn, W. (1991). Learning from maps and diagrams. Educational Psychology Review, 3, 211-247.

Woolley, G. (2014). Using visualisation and imagery to enhance reading comprehension. Literacy in the Arts: Retheorising learning and teaching, 215-234.

Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). Ai-generated content (aigc): A survey. *arXiv preprint arXiv:2304.06632*.

Yıkılmaz, I. (2020). The future of knowledge management: Web 4.0. Data, Information and Knowledge Management; Mert, G., Sen, E., Yılmaz, O., Eds, 519-539.

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3836-3847).

Zhuo, F. (2021, July). Human-machine co-creation on artistic paintings. In *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)* (pp. 316-319). IEEE.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010, August). An analysis and critique of Research through Design: towards a formalization of a research approach. In *proceedings of the 8th ACM conference on designing interactive systems* (pp. 310-319).

Appendix

Appendix 1: Narrative analysis formats

Narrative structure

Narrative structure is the order in which plot events are told to the reader or audience. While stories can be told in a wide variety of ways, most story structures commonly share certain elements: exposition, rising action, climax, falling action, and resolution (Rodak).

- **Exposition**. This first part establishes a protagonist's normal life and greater desires, and usually culminates in the inciting incident.
- **Rising action**. The protagonist pursues their new goal and is tested along the way.
- **Climax**. Our hero achieves their goal — or so they think!
- **Falling action**. The hero now must deal with the consequences of achieving their goal.
- **Resolution**. The conclusion tying together the plot, character arcs, and themes.

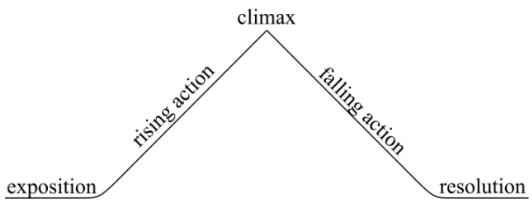


Figure 2: The project approach

A tightly controlled structure will answer a reader's questions, provide a climax followed by resolution and information at the end of the story, further the characters' development, and unravel any central conflicts.

Narrative-constructing factors

Reading narrative books for constructing knowledge involves deep engagement with the text to understand and appreciate the complexities of storytelling, character development, and thematic depth. Reflecting a narrative requires dissecting various elements that contribute to the overall knowledge construction of the story.

- **Subject analysis:**

  Identifying and examining the themes and topics explored in the narrative.

- **Narrative arc and plot structure analysis:**

  This involves studying the narrative's structure—the exposition, rising action, climax, falling action, and resolution—to understand how the story unfolds. Analyzing the narrative arc and plot structure helps in understanding how the author constructs the story to build tension, develop characters, and ultimately resolve the narrative in a satisfying way.

- **Story beats placement:**

  Story beats are specific, significant events that push the plot forward, often changing the story's direction or revealing important character information. Analyzing story beats' placement involves examining how these events are distributed throughout the narrative to maintain engagement, develop the plot, and contribute to the thematic depth of the story.

- **Pacing analysis**

  Pacing refers to the speed at which the story progresses. Pacing analysis looks at how the author controls the flow of the story—when the narrative speeds up to create excitement or tension and when it slows down to allow for character development or thematic exploration. Effective pacing is crucial for maintaining reader interest and ensuring the story's emotional and dramatic beats have their intended impact.

- **Major characters' personality traits:**

  This involves a detailed study of the main characters' behaviors, motivations, desires, and changes they undergo throughout the story. Understanding the characters' personality traits helps readers empathize with them, understand their decisions, and see how they contribute to the narrative's thematic concerns.

- **Primary emotions color wheel:**

  It is a tool used to analyze the emotional spectrum in narrative books. It can help categorize and visualize the emotions portrayed in the story, examining how these emotions are expressed, evolve, and interact.

- **Dialogue vs. narrative breakdown**

  This analysis contrasts the spoken interactions between characters (dialogue) with the descriptive or expository parts of the text (narrative). Examining the balance between dialogue and narrative can reveal how the author uses different elements to develop characters, advance the plot, and build the story's world.

Appendix 2: User ritual research plan

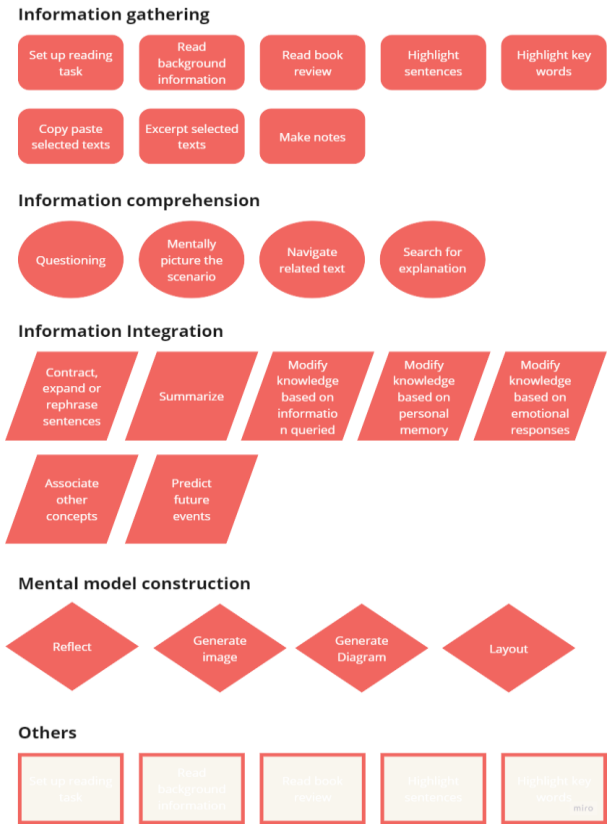
Workshop materials

Activities list is integrated from literature review on Active Reading (chapter 3.2)

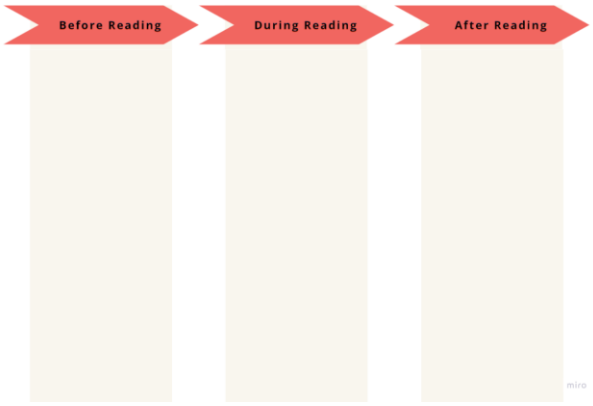
1. Information gathering
    a. **Set up reading task**: Set up goals, expectations, assignment etc
    b. **Activate background information**: Research the story context, author, book reviews, etc.
    c. **Highlight**: Highlight, underline or circle important information (sentences and key words).
    d. **Annotate**: Code or write important notes in the margins or use sticky notes to record ideas, questions, predictions, and responses.
    e. **Content extraction**: copy and paste selected texts, excerpt sentences.
2. Information comprehension
    a. **Navigate related text**: searching for text, turning a page, or flipping between locations to compare parts of a text.
    b. **Picture the scenario mentally**: Imagine the plot, visualize the characters, and mentally construct storylines and scenes based on the information provided.
    c. **Questioning**: Have questions about allusions, history, and plot settings in literary works.
    d. **Search for explanation**: ask experts or other documents for explanations.
3. Information integration
    a. **Coding**: Contract, expand or rephrase sentences to accurately express personal understanding.
    b. **Modify knowledge based on personal memory**: Make connections and adjustments with what I already know or have experienced.
    c. **Modify knowledge based on information queried**: Modify understandings of texts with complementary information acquired from other people or documents.
    d. **Modify knowledge based on emotional responses**: Modify understanding of texts with personal emotion, expectation and atmosphere, etc.
    e. **Associate other concepts:** Associate the correlations to people, events, and concepts in memories.
    f. **Summarize**: Provide a summary, general description of the narrative excerpt.
    g. **Predict future events**: Make predictions about plot development and story endings based on personal reasoning or expectations
4. Mental model construction
    a. **Reflect:** Recall the story, think about what I learn, how it is important.
    b. **Diagraming**: Creating graphical representations by diagramming to visualize mental models.
    c. **Generate image**: Visualize characters or scenarios to enhance comprehension and memory.

d. **Layout**: Typesetting layout of different pieces of texts, visuals, lines, etc.

5. Other activity _____



Process map is designed to let participants map out their active reading activities in stages of before, during and after reading.



Equipment

- 5 One-page narrative materials (Literary, fantasy, scientific, news, history) - Appendix 1.1
- Tablet and pen
- Paper cards about active reading activities - Appendix 1.2
- Map of three reading stages - Appendix 1.3
- Consent form
- Recording device
- Inspirations - Appendix 1.4

Research plan

Key themes

1. What are the contexts in which knowledge construction is a desired or sought-after outcome when reading books?
2. How do they construct the knowledge model?
3. What does the mental model look like?
4. What part of active reading do they find difficult?

Introduction:

“Hi, welcome to this research, I am Meng Cai and thank you for participating in the test. The test and interview afterwards will be recorded, the audio is only used for me and the project team analysis and will be deleted after the project is finished. By signing the consent form you will agree with this. The research consists of three parts: First I will ask you some questions to know your basic information. Then I will provide you with materials and introduction on concepts used in the test and ask you to use paper cards to complete activities. If possible, it is better for you to think out loud while performing. And if there is any question, you could ask me for help. I would help you if the question doesn’t affect the research. At the end of the test I will ask you to reflect on your knowledge construction process and have a short interview with you. Do you have any questions?”

Before experiment

Ask the following questions to know the basic information of the participant. This will decide which type of narratives is provided in the experiment, and contribute to defining the design scope.

1. For what kind of narrative book, do you want to actively read and understand text information, and mentally think and construct the knowledge framework regarding symbolic meaning, theme, world view, humanity and social implications of the text, etc? Not just for fun. And why?
2. What tool do you use now to read narratives and construct knowledge?
3. What language do you usually read books in?

During experiment

Provide a tablet with selected reading materials and pen, suggest thinking out loud while reading to describe the activities. The participant will take time to read and I will observe their activities at the same time.

"I want you to actively read it with a purpose to get some knowledge out of it. You are encouraged to write anything with the pen."

The researcher draws the mental model based on participants' description / action, comparing the sketch with the participant's sketch after the experiment.

After experiment

4. If not many active reading actions,
    a. Do you find this reading material useful to construct knowledge on? I see you didn't mark anything.
        i. Can you reflect on the most recent book that you read and …

Ask the participant to recall the reading material, and map out the mental model.

5. What does the mental model about this narrative look like?
    a. What information are you recalling in your mind? Can you describe them to me? (drawing done by researcher)
    b. Can you draw me your mind map?
    c. Repeat the model map, "does it correctly convey your mental model?"
    d. So, your mental model of this reading material is (E.g. scenario, storyline, character interaction, cause and effect, value concept…), Why?

Ask the participant to reflect on reading experiences, and use the provided paper cards to visualize the knowledge construction process on the provided map of stages to answer the following questions. Show the explanations of the concepts on the paper cards (see 3.2.1).

6. How do you gather knowledge and construct the knowledge model when you actively read a book? Can you use these paper cards to facilitate the process on this map?
    a. If you picture stories, which part do you picture? What did you see?

    b. Can you connect the components if they have a linear or circular relationship?

7. What part of this mental model constructing process do they find difficult? Why?

Pilot feedback

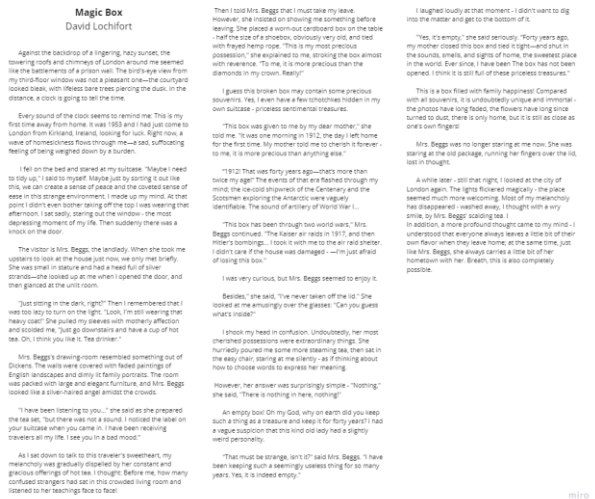
The interview plan has been modified based on the following feedback (Yellow texts).

1. Mental modeling is a thinking process, not a physical action. Participants find it hard to visualize mental models on paper. They tend to draw something randomly.
    a. It's better to draw the mind map by myself based on the participant's description, and double check with them.
    b. In this case, I will show them the inspiration list, and ask them to say and draw the mental model. And I will draw it as well as a double check.
    c. Give some icons to guide the participants.
2. When I ask participants to tell me their mental model, they tend to only focus on the storyline.
    a. Emphasize that it's your selection and extension of information that is useful for you to store and reflect on the story.
3. Sometimes the provided reading material cannot stimulate active reading because of personal preference, so they tend to take no action while reading, and not much reflection afterwards.
    a. Ask them to reflect on a book that they actively read most recently, and draw the mental model and reading activities.
4. Non-native language may cause difficulties in reading, and language difference in reading is out of scope for this study.
    a. Provide materials in the language that users find most comfortable to read.

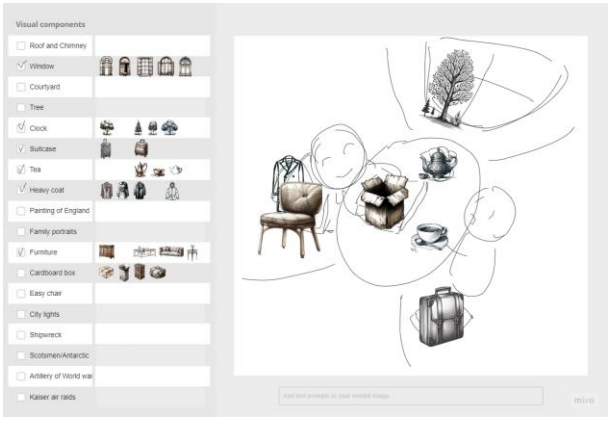
Appendix 3: Prototype of visualizing mental image platform

To reduce the interference of external factors on user testing, the carriers of this prototype design are selected as laptop readers and tablet readers commonly mentioned in previous user research. The size of the interface will try to fit into both laptop and tablet readers. The reading material is selected as the most chosen article in the previous user research — *The magic box*.

Solution 1



Interface 1 - Narrative Reading



Interface 2 - Sketching workspace

Interface 3 - Image generation output

The interface 3 is assisted by customized GPT with the following prompts:

The user is reading the narrative article in the knowledge base. The purpose of the mental image generation session is to help active readers visualize mental images that are predicting user preference with knowledge base and well aligned with expectation.

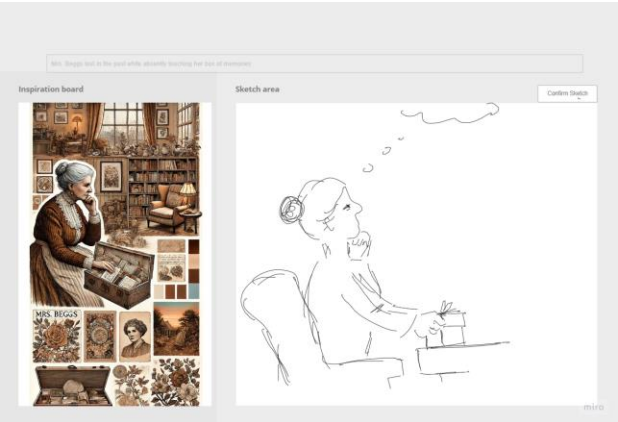
1. *Generate an image based on the input sketch as layout guideline and the input text prompts as content guideline. The sketch will include some figurative sketch objects, make sure the generated image has those objects in the same size and location. Predict the image style and color it by analyzing the knowledge base and text prompts in terms of mood, themes, atmosphere. The picture should reflect a touch of sadness and homesickness (emotional arc analysis output). If the text and sketch prompts include characters of Mrs. Beggs and the man, please create their images based on the physical descriptions of the*

characters in the original text uploaded.

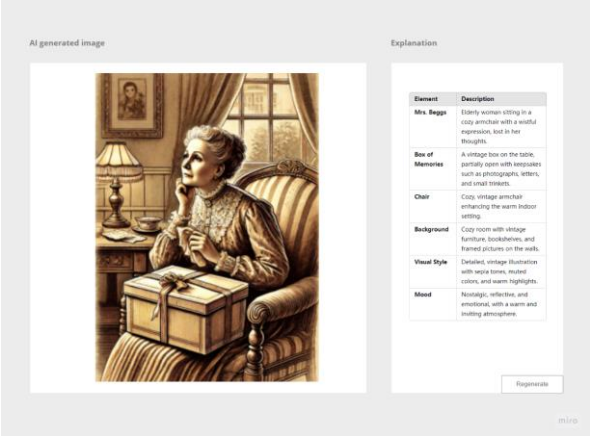
2. *Ask users for feedback of further modification.*
3. *If the user input the table with modified explanations again, then regenerate an image based on the table.*

Solution 2

Interface 1 - Narrative reading and visual options selecting



Interface 2 - Sketching workspace



Interface 3 - Image generation output and explanation table

The prototype is assisted by customized GPT with the following prompts:

The purpose of the mental image generation session is to help active readers visualize mental images that are well aligned with their preference and expectation.

Step by step:

1. *Welcome the user and tell them the purpose, ask them to input their selected images.*
2. *Analyze the images in terms of visual style, content and themes, mood and atmosphere and color palette, generating a conclusion of user preference.*
3. *Then, ask the user to input text prompts of their mental images.*
4. *Use the conclusion of user preference to generate an inspiration board based on the input text prompts. The inspiration board should layout specific visual elements in a vertical A4.*
5. *Then, ask the user to input their sketch prompts of their mental images.*
6. *Following the previous conclusion of user preference and seeds of the inspiration board, generate an image based on the input sketch as layout guideline and the input text prompts as content guideline. The image has the same size as the input sketch. If the*

prompts contain the characters or the room or the box, please follow the descriptions of the article in the knowledge base to create image elements.

7. *Explain the reasoning to this output from the input text prompts, sketch and inspiration board. Create a table and display all the explanations in short. Ask for the following modification.*
8. *If the user input the table with modified explanations again, then regenerate an image based on the table.*

Always ask for the user's response and keep the dialogue going. If I say "back" then start from the very beginning.

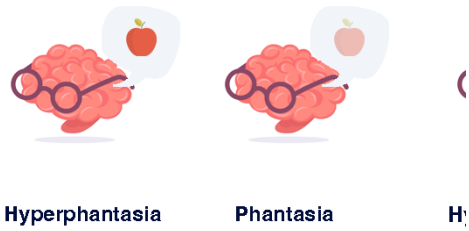
Appendix 4: User Test plan of Design round 1

Introduction:

“Hi, welcome to this research, I am Meng Cai and thank you for participating in the test. The test and interview afterwards will be recorded, the audio is only used for me and the project team analysis and will be deleted after the project is finished. By signing the consent form you will agree with this. The research consists of three parts: First I will ask you some questions to know your basic information. Then I will provide you with materials and instructions to help you complete activities. And if there is any question, you could ask me for help. I would help you if the question doesn’t affect the research. At the end of the test I will ask you to reflect on the whole workflow and have a short interview with you. Do you have any questions?”

Before experiment

1. VVIQ (Marks, 1973) test to understand participants’ mental imagery abilities. https://aphantasia.com/study/vviq/



2. Ask the participant to use AI to visualize the image for one VVIQ question.

During experiment

For Solution 1

1. Ask the participant to actively read the given narrative materials in a comfortable way. They can stop anywhere during reading if they have a strong will to visualize their mental image.
2. Ask their reflection or think about the given material, think about an image showing up in mind.
3. Provide a list of labels for the visual components in the reading material, ask the participant to select relevant labels existing or expecting to exist in their mental image.
4. Based on the selected labels, display 3-5 visual components in easy sketch style, ask participants to drag and resize the components if applicable to sketch out their mental image. Inform participants that the sketch will be used by AI to create an image.
5. Ask them to add text instructions to further control image generation.
6. Display the output to the participant, ask if they are satisfied with it. If not, go back to the Sketch area and modify inputs. Restart from step 4 until the participant generates the desired output.

For Solution 2

1. Ask the participant to actively read the given narrative materials in a comfortable way. When they picture some scenarios mentally according to a specific paragraph, ask them to click on the paragraph, and choose one that is most close to their mental image out of the four images that appear. They can stop anywhere during reading if they have a strong will to visualize their mental image.
2. The researcher puts the images selected during reading into the customized GPT (chapter 5.5.2.1) and follows its instruction.

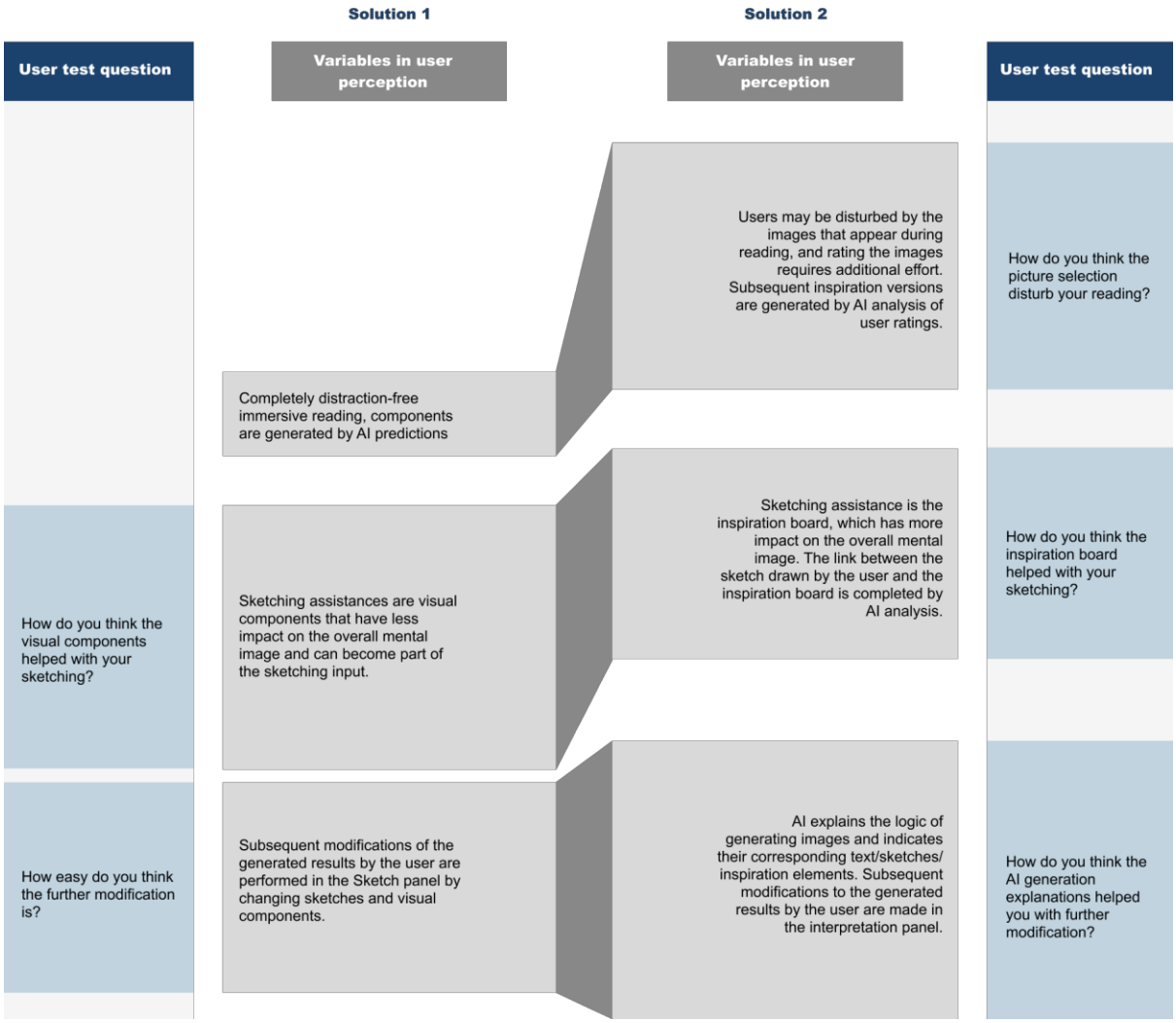
3. Ask the participant to input a text prompt for his mental image.
4. Display the inspiration board, ask the participants to freely sketch out their mental image. Inform participants that the sketch will be used by AI to create an image.
5. Display the output to the participant, together with AI explanations. Ask if they are satisfied with it. If not, ask them to modify the AI explanations. Repeat step 5 until the participant generates the desired output.

After experiment

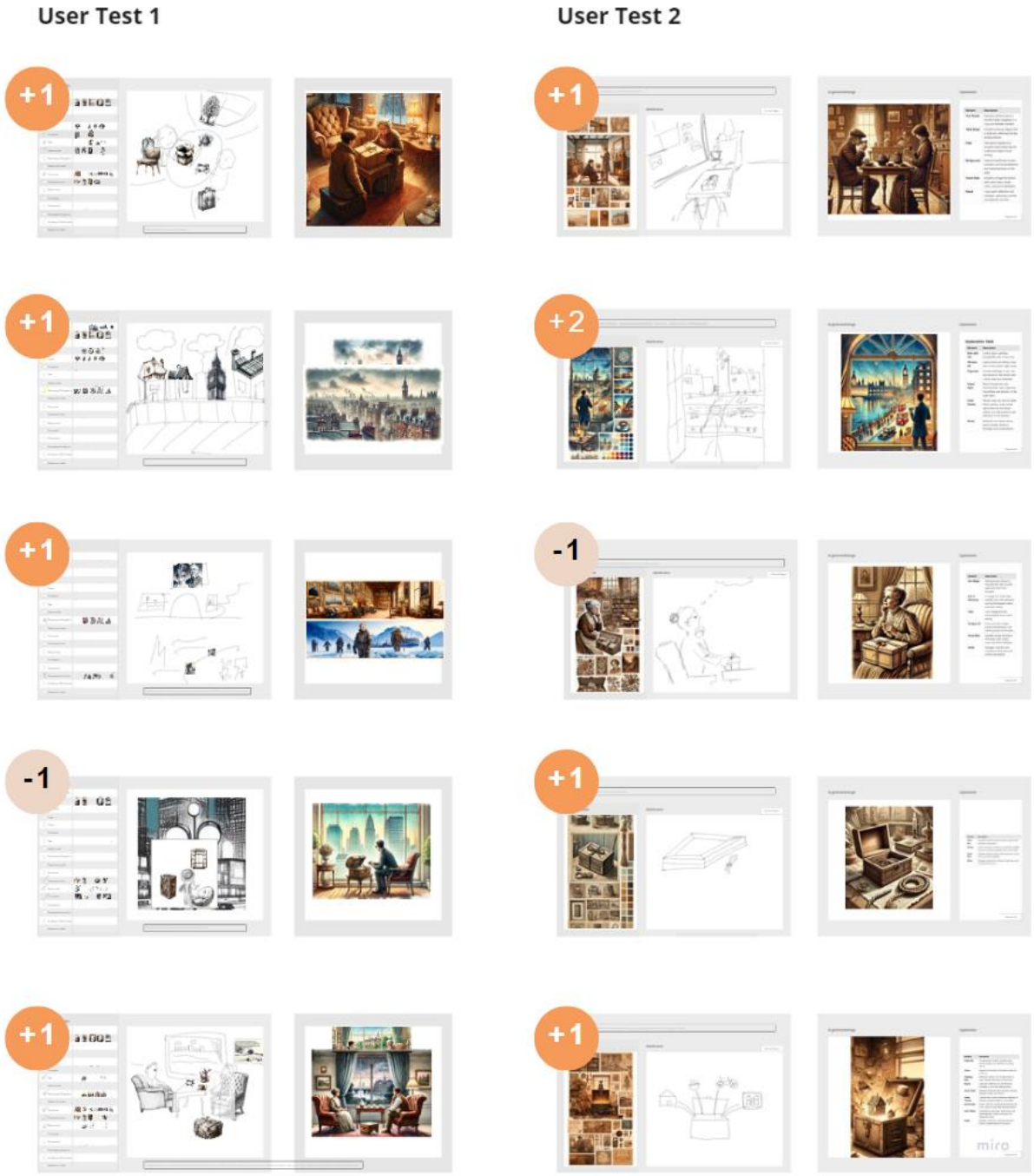
An interview will be conducted to understand the user experience of the given solution flow. The feedback will be collected for data analysis and product iteration.

Rank from 0-5, 3 is neutral.

6. To what extent do you feel satisfied with the final output?
    a. To what extent is the output aligned with your mental image?
    b. To what extent do you feel the output is beyond your expectation?
7. To what extent do you feel satisfied with the image generation process?
    a. Solution 2: To what extent do you think it causes disturbances for your reading?
    b. How much extraneous mental effort that gets you out of this story do you need to complete this process? (Phelan, 2020) And which part causes that?
    c. How confused, discouraged, irritated, stressed, and annoyed were you? (Hart and Staveland's NASA Task Load Index (TLX) method, 1988)
    d. To what extent do you think it's efficient to generate the desired output?
8. Personalization: How do the process support you visualizing mental pictures compared with the direct text to image process? Can you describe in which aspects?
    a. Solution 1:
        i. How do you think the visual components helped with your sketching?
        ii. How easy do you think the further modification is?
    b. Solution 2:
        i. How do you think the inspiration board helped with your sketching?
        ii. How do you think the sketch helped with the image generation?
        iii. How do you think the AI generation explanations helped you with further modification?
9. Knowledge construction: What is the effect of this visualization for your reading experience?
10. User-controlled: To what extent do you feel in control of the reading and visualization process?
11. Future-oriented: To what extent do you think this service will adapt to future reading trends?
12. Any comments on the user test itself?

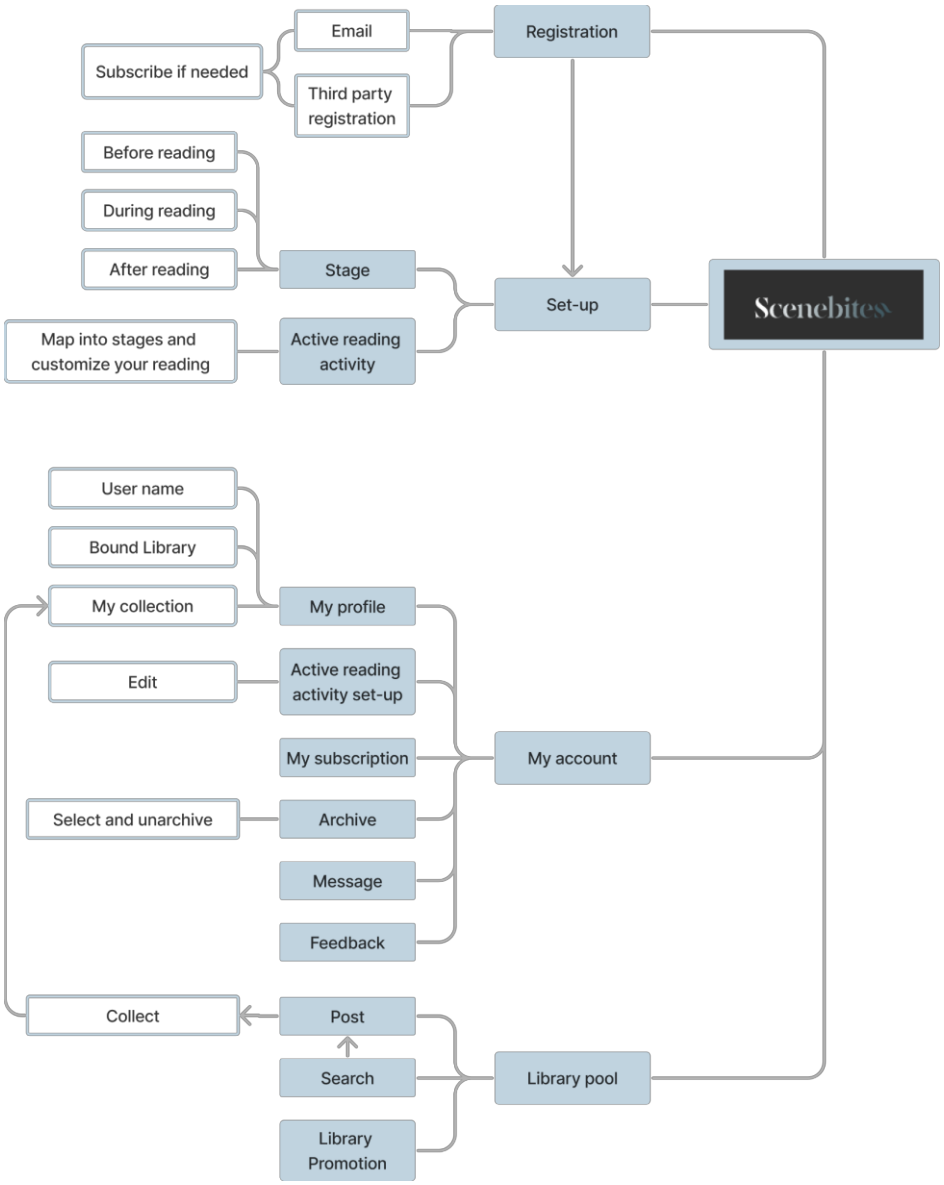

Solution 1
Solution 2
User test question
Variables in user perception
Variables in user perception
User test question
How do you think the visual components helped with your sketching?
How easy do you think the further modification is?
Completely distraction-free immersive reading, components are generated by AI predictions
Sketching assistances are visual components that have less impact on the overall mental image and can become part of the sketching input.
Subsequent modifications of the generated results by the user are performed in the Sketch panel by changing sketches and visual components.
Users may be disturbed by the images that appear during reading, and rating the images requires additional effort. Subsequent inspiration versions are generated by AI analysis of user ratings.
Sketching assistance is the inspiration board, which has more impact on the overall mental image. The link between the sketch drawn by the user and the inspiration board is completed by AI analysis.
AI explains the logic of generating images and indicates their corresponding text/sketches/ inspiration elements. Subsequent modifications to the generated results by the user are made in the interpretation panel.
How do you think the picture selection disturb your reading?
How do you think the inspiration board helped with your sketching?
How do you think the AI generation explanations helped you with further modification?

Appendix 5: Results of User Test in design round 1



* +2 means extreme high visual imagery ability, +1 means high visual imagery, -1 means low visual imagery ability, -2 means extreme low visual imagery ability.

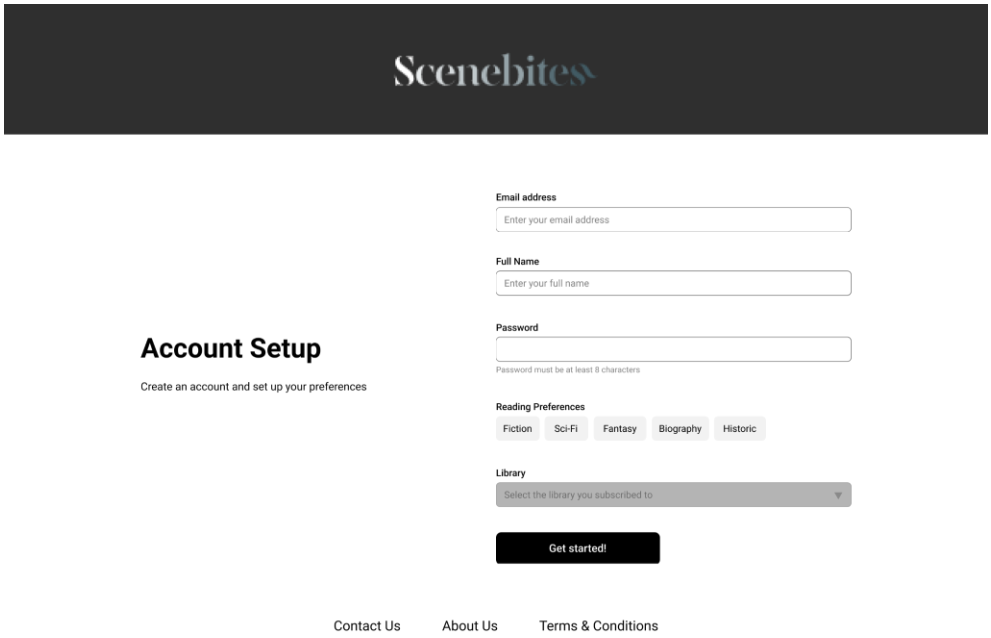
Appendix 6: Other basic features of Scenebites



Registration and Set-Up

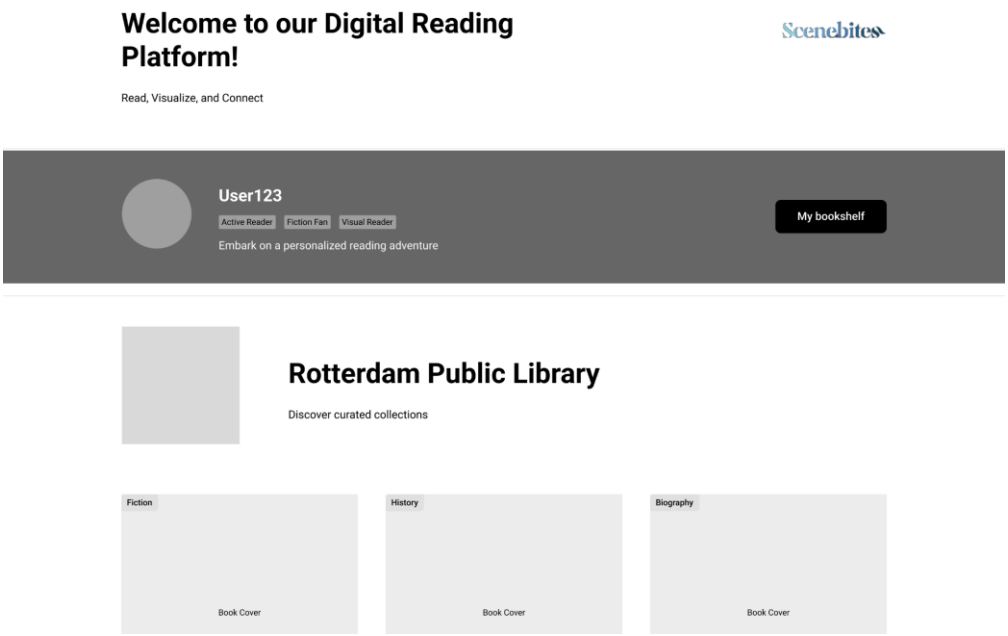
The entry point for users is the registration and set-up phase, where new users create accounts and configure their reading preferences. This initial stage ensures personalized experiences catering to individual reading habits and requirements.

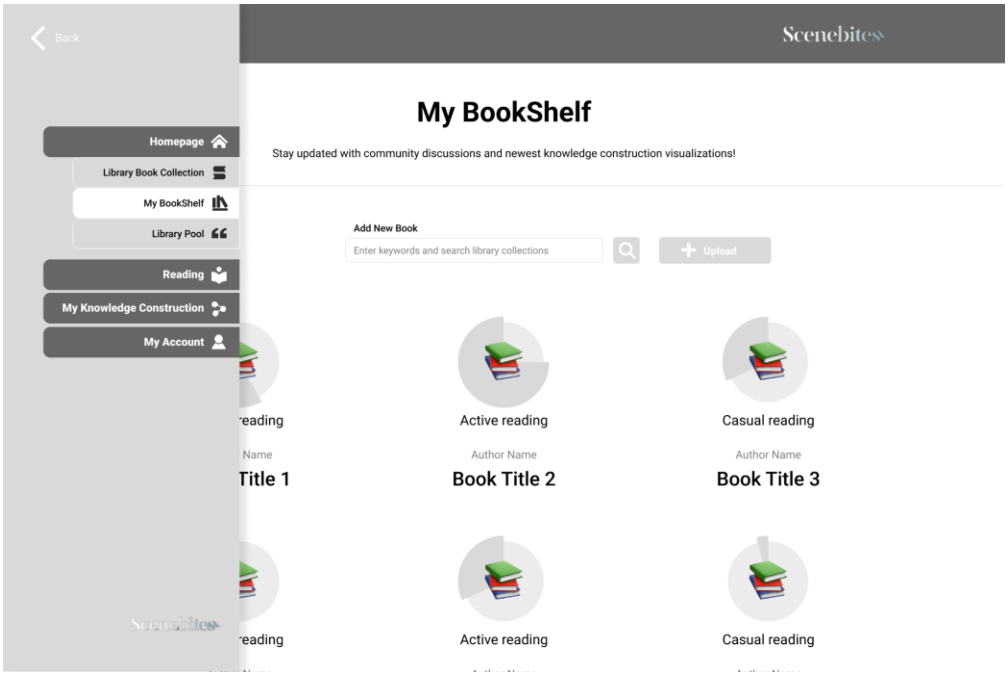
For users who log in to read using library devices, they need to subscribe to use the visualizing mental model function enabled by AIGC.



Main Interface and Navigation

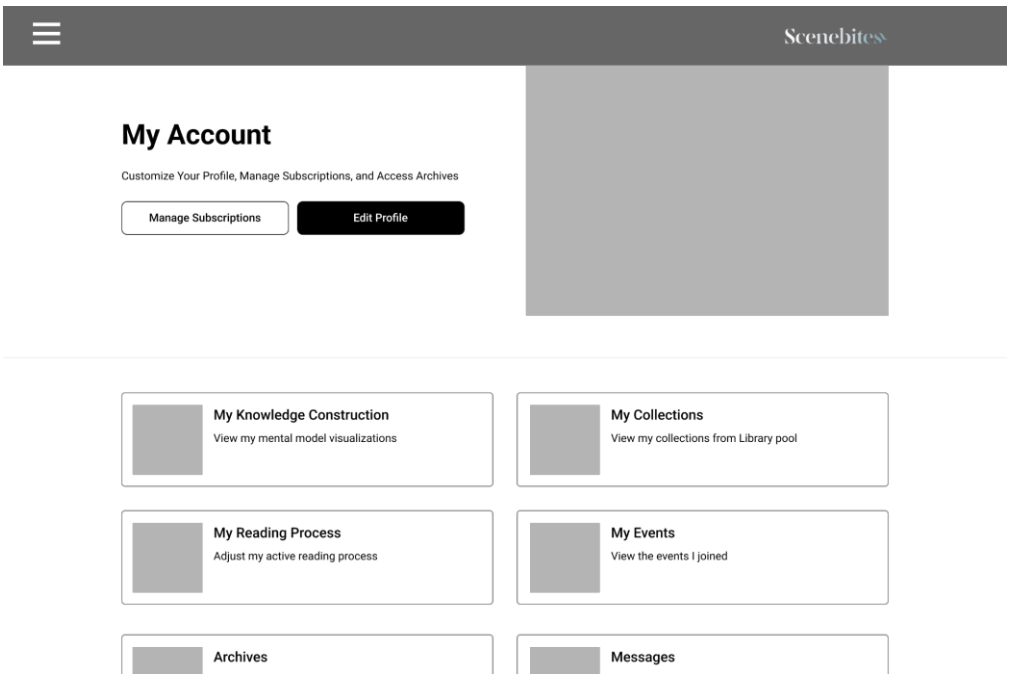
Upon successful set-up, users are directed to the homepage, which serves as the central hub for accessing various functionalities. The homepage links to essential sections such as 'My Bookshelf,' where users can import and manage their book collections.





User Account Management

User account management feature is where users manage their personal information and preference settings , including profile customization, subscription management, knowledge construction base, post collections, reading process management, and access to archives and messages.



Appendix 7: User Validation Plan

Prototype display:

The video as the minimum viable product was shown and introduced. The researcher explained the function flow of the designed platform and answered all the relevant questions from the participants.

User validation interview:

An interview was conducted to understand the user experience of the given minimum viable product. During the interview, the participants were in a focus group where they can share and compare each other's opinion.

The feedback will be collected for data analysis and product validation.

Rank from 0-5, 0 is not satisfied, 5 is very satisfied.

1. How do you feel about the interaction with the digital reading tool?
2. Does this process support knowledge construction?
    a. Reduce extraneous cognitive load - ratings of perceived mental effort
    b. strengthen comprehension
    c. enhance recall
    d. make discoveries in the knowledge (Victor, 2010)
3. How do you think the process caters to your personal visualization ability/preference?
4. To what extent do you feel in control of the reading and visualization process?
5. To what extent are you **interested in subscribing** to this platform in your future narrative reading context?

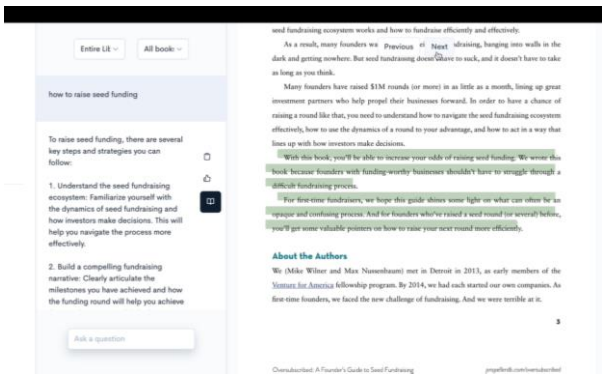
Appendix 8: AI technological feasibility analysis of Scenebites

Real-time AI conversation

Conversational AI facilitates user interaction by providing answers to questions, assisting with navigation, and helping with annotations. Current Natural Language Processing (NLP) technologies, such as those provided by models like GPT-4, are highly capable of understanding and generating human-like responses, making this feature feasible. Integration of conversational AI into the platform can be achieved through APIs that connect to pre-trained models.

Chen (2023), Verma (2023) and others have conducted research on conversational artificial intelligence, proving that embedding real-time AI conversational capabilities into library services and digital reading experiences can greatly enhance user experience, productivity and the efficiency of library services. Chatbots can interpret and respond to user queries and answer questions in an accurate and humane way, and help them easily browse the library's collections and resources.

As I researched and introduced in Chapter 5.3.7.2 AI Powered Book Analyzers, many platforms like MyReader AI and iChatBook enable readers to have real-time conversation with AI chatbot when reading books.



Interface of MyReader AI

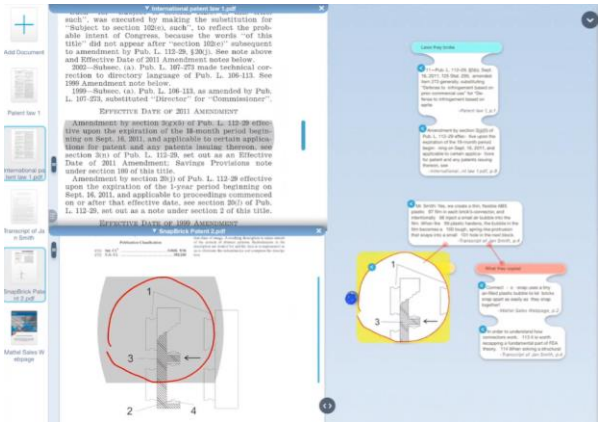
Real-time whiteboard recording

Implementing a real-time whiteboard recording and writing area linked with annotations in books involves integrating several technologies to create an interactive and seamless user experience. The necessary technologies are mature and well-supported, from web technologies for the interface to real-time collaboration tools, annotation systems, and cloud storage. Besides, to support user hand-writing in the whiteboard, hardware devices like tablets/ touch screens together with stylus are necessary. iPads (Apple Pencil), Surface tablets, and other touchscreen devices can easily be applied into future libraries to meet the feasibility.

As I researched and introduced in Chapter 5.3.7.1 AI powered book readers, LiquidText as a professional and interactive digital reader can already achieve the whiteboard function. It offering a flexible, fluid document representation built on multitouch input, with a range of interaction techniques designed to facilitate the activities of active reading. The whiteboard workspace can support 1) extensive, flexible, direct manipulation control of the visual

arrangement of content, including both original material as well as annotations, and 2) flexibility in navigating through the content, with a rich array of navigational affordances that

could be tailored and organized as required (Tashman & Edwards, 2011). Therefore, it's highly feasible to develop a real-time whiteboard recording function in Scenebites platform.



Interface of LiquidText

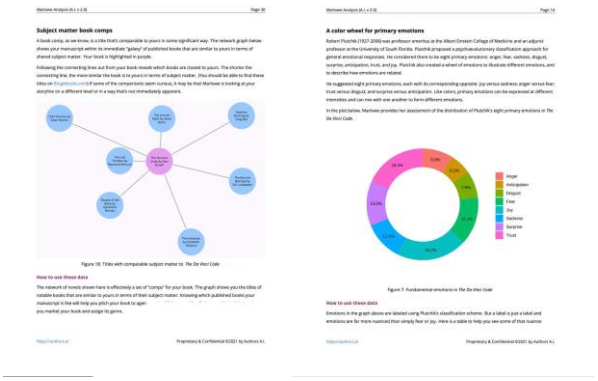
AI Text analysis

AI text analysis uses machine learning and natural language processing (NLP) techniques to extract meaningful information from text. When applied to narrative books, it involves analyzing the text to extract meaningful insights, perform subject analysis, and support various visualization tasks. First it processes the input texts and divides the text into meaningful segments, such as sentences, paragraphs, or chapters. Steps such as tokenization, lemmatization, and removal of stop words are applied to prepare the text for analysis. Then the Entity Recognition (NER) will identify key entities like characters, places, and objects within the text. With the key entities, the established algorithms could calculate the relationships between different entities and how they evolve over the narrative.

Narrative constructing analysis

AI text analysis helps users in understanding and generating the underlying structure of narratives, identifying key themes, characters, and plot points. Using the extracted data from given narrative materials, the system identifies the classic narrative arc elements (exposition, rising action, climax, falling action, resolution). It can also track the evolution of characters, their motivations, and their roles in the plot. Together with AI diagramming, it's feasible to visualize narrative constructing analysis and provide narrative readers with accurate and editable knowledge construction assistance.

As I researched and introduced in Chapter 5.3.7.2 AI Powered Book Analyzers, many platforms like Marlowe and OddBooks are able to analyze selected books and generate narrative constructing analysis in a visualized diagram.



The diagram generated by [Marlowe from authors.ai]

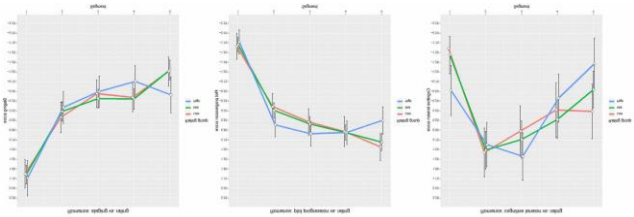
Emotion arc analysis

Emotion arc analysis in narrative books involves mapping the emotional trajectory of a story's characters or the overall narrative, identifying shifts and patterns in sentiment as the plot progresses. This process is rooted in the interdisciplinary application of natural language processing (NLP), sentiment analysis, and machine learning. The first step is parsing the text and breaking it down into manageable units such as sentences or words. Named Entity Recognition (NER) is used to identify characters and important entities within the text is crucial for tracking emotional states. Then, advanced machine learning models, such as those based on transformers (e.g., BERT, GPT-3), provide more accurate and context-sensitive sentiment analysis (Brahman & Chaturvedi, 2020). However, Literary narratives often contain complex emotional states that go beyond basic emotions. Developing models that can detect subtle emotional movements is crucial. This can be achieved through

custom datasets and deep learning techniques (Kim & Klinger, 2018). Finally, pre-setted diagram generation algorithms will convert textual information into readable and interactive diagrams to show the emotion arc.

Brahman and Chaturvedi's (2020) model constructs emotion arcs from narratives by defining sequences of 5 basic emotions (anger, fear, joy, sadness, and neutral). They use the Commonsense Transformer (Bosselut et al.,2019) to infer implicit emotional states and a fine-tuned BERT-based emotion classifier to track the protagonist's emotions.

Boyd et al. (2020) Uses computer-based language analysis methods to visualize the narrative arc that revealed three primary processes: staging, plot progression, and cognitive tension. Their model constructs emotion arcs using tools like the Commonsense Transformer (COMET) for implicit emotional states and a BERT-based emotion classifier to track protagonists' emotions, they identify how emotional peaks and valleys unfold throughout the narratives. These arcs are constructed by examining the frequency and context of emotion-laden words throughout the texts, aligning them with narrative structures and psychological theories of storytelling.



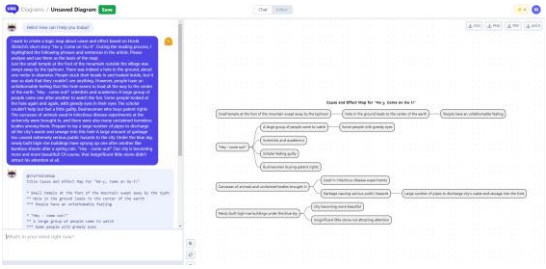
Means of each narrative dimension for romance novel (Boyd et al., 2020)

AI diagramming

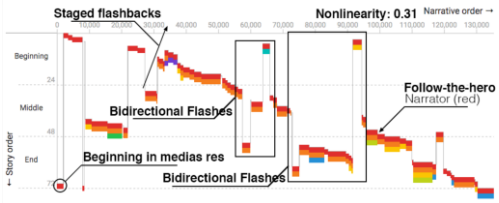
AI can automatically generate diagrams from data sets, code, or textual descriptions. First, NLP algorithms parse and understand text input from users. These models can comprehend the structure and meaning of the text to determine what kind of diagram is needed and what elements it should include. Then, specific Diagram Generation Algorithms can convert the parsed information into visual elements and arrange them in a diagram. ML models might be used to predict and place components correctly based on learned patterns. With AI provided output, users can use editing tools to edit diagrams easily. The AI system validates the logical consistency of the diagram.

Currently, there are AI-powered tools and platforms that can help generate diagrams, mind maps, and visualize information based on input human instruction. These tools often utilize natural language processing (NLP) algorithms to understand the input text and then generate structured diagrams or mind maps based on the inferred relationships between the concepts. I researched several AI-powered platforms which generate diagrams and mind maps based on user requirements and data retrieval, such as ChartAI, Diagramming, DGM, Dreamspace,Whimsical, Diagrams: Show Me, SlateBox etc. The following platforms are representative cases in generating diagrams via AI.

1. **ChatUML** is an AI-driven diagram generator utilizing OpenAI GPT-4 for creating high-quality diagrams effortlessly. It supports multiple formats like SVG, PNG, PDF, and ASCII, enabling easy collaboration and visualization of various diagram types.

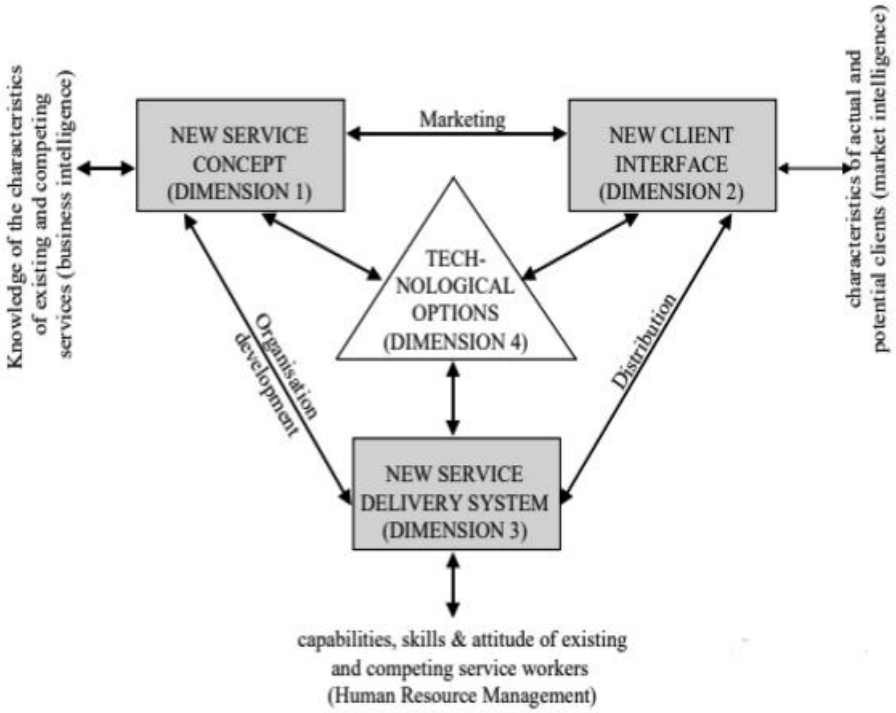


2. **Story Explorer** is an interactive tool that pairs story curves with detailed information on characters and settings, communicating the complex structures of nonlinear narratives found in films (Kim et al., 2017).

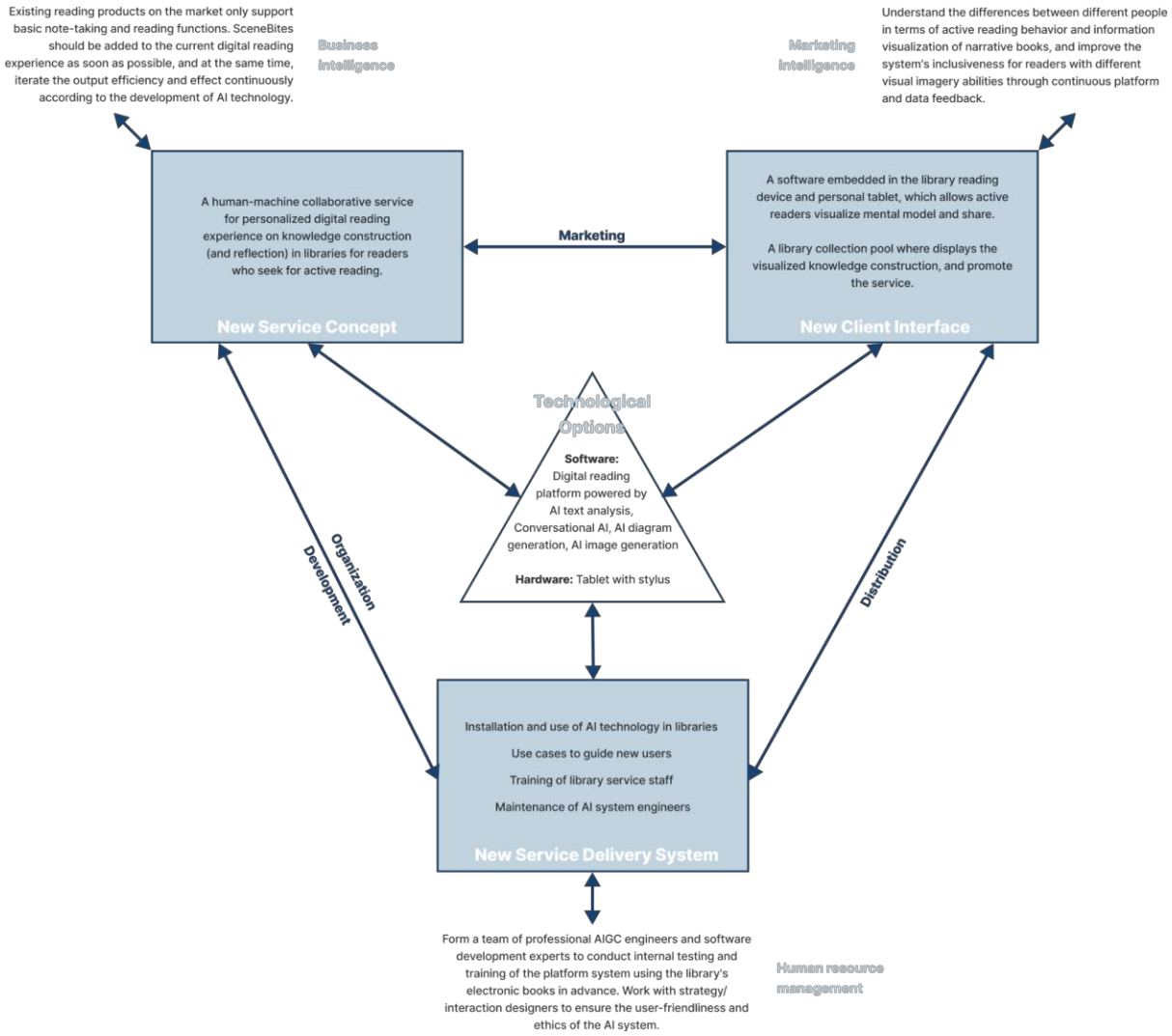


Appendix 9: Implementation suggestions for future library

“Service innovation framework” of Hertog (2000) was integrated to form a framework that guides the development and implementation of an “AIGC powered active reading and mental model visualization service” in future libraries.



Source(s): Hertog (2000)



This service innovation framework introduces a strategic approach for future libraries to adopt advanced technologies and deliver personalized, active reading experiences, by integrating human-machine collaboration in the narrative readings. The framework is structured around three core components: New Service Concept, New Client Interface, and New Service Delivery System, all interconnected through technological options and supported by business intelligence, marketing intelligence, and human resource management.

New Service Concept

At the heart of the framework is the new service concept, which emphasizes a human-machine collaborative service designed to provide personalized digital reading experiences. This service focuses on knowledge construction and reflection for narrative readers who engage in active reading. With AI analyzing and content generation, the service aims to adapt to the individual needs and preferences of readers, enhancing their comprehension and engagement with the narratives.

New Client Interface

The new client interface is a software embedded in library reading devices and personal tablets. This software facilitates active reading by allowing readers to visualize their mental models and share them with others. The interface supports a library collection pool that displays these visualized

knowledge constructions, promoting the service and encouraging collaborative learning.

New Service Delivery System

The implementation of the new service concept and client interface is supported by a robust service delivery channel. This includes the installation and use of AI technology within libraries, ensuring that the software is compatible with the library system and tablet hardware facilities. It also involves training library service staff and maintaining a team of AI system engineers to ensure the smooth operation and continuous improvement of the service.

Technological Options

Technological options form the cornerstone of this framework, linking the service concept, client interface, and delivery system. The software embeds AI and machine learning technologies to analyze and visualize readers' mental models, providing personalized and inclusive knowledge construction processes for narrative readers with diverse visual imagery abilities and cognitive experiences. To achieve particle use, libraries are required to provide tablet and stylus devices for use by subscribed users.

The framework is further supported by key business functions:

- **Business Intelligence:** Provides data-driven insights to inform the development and refinement of the service concept and interface.
- **Marketing Intelligence:** Focuses on promoting the service, attracting new users, and understanding user needs and preferences. For better user engagement and service promotion, regular community activities will be held by the marketing team.
- **Human Resource Management:** Ensures that library staff and system engineers are well-equipped and trained to support the new service, fostering a culture of continuous learning and adaptation.

IDE Master Graduation Project

Project team, procedural checks and Personal Project Brief

In this document the agreements made between student and supervisory team about the student's IDE Master Graduation Project are set out. This document may also include involvement of an external client, however does not cover any legal matters student and client (might) agree upon. Next to that, this document facilitates the required procedural checks:

- Student defines the team, what the student is going to do/deliver and how that will come about
- Chair of the supervisory team signs, to formally approve the project's setup / Project brief
- SSC E&SA (Shared Service Centre, Education & Student Affairs) report on the student's registration and study progress
- IDE's Board of Examiners confirms the proposed supervisory team on their eligibility, and whether the student is allowed to start the Graduation Project

STUDENT DATA & MASTER PROGRAMME

Complete all fields and indicate which master(s) you are in

Family name	
Initials	
Given name	
Student number	

IDE master(s)	IPD DfI SPD
2nd non-IDE master	
Individual programme *(date of approval)*	
Medisign	
HPM	

SUPERVISORY TEAM

Fill in he required information of supervisory team members. If applicable, company mentor is added as 2nd mentor

Chair		dept./section	
mentor		dept./section	
2nd mentor			
client:			
city:		country:	
optional comments			

! Ensure a heterogeneous team. In case you wish to include team members from the same section, explain why.

! Chair should request the IDE Board of Examiners for approval when a non-IDE mentor is proposed. Include CV and motivation letter.

! 2nd mentor only applies when a client is involved.

APPROVAL OF CHAIR on PROJECT PROPOSAL / PROJECT BRIEF -> to be filled in **by the Chair** of the supervisory team

Sign for approval (Chair)

Name _____ Date _____ Signature _____

CHECK ON STUDY PROGRESS

To be filled in **by SSC E&SA** (Shared Service Centre, Education & Student Affairs), after approval of the project brief by the chair. The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total _____ EC

Of which, taking conditional requirements into account, can be part of the exam programme _____ EC

	YES	all 1st year master courses passed
	NO	missing 1st year courses

Comments:

Sign for approval (SSC E&SA)

Name _____ Date _____ Signature _____

APPROVAL OF BOARD OF EXAMINERS IDE on SUPERVISORY TEAM -> to be checked and filled in by IDE's Board of Examiners

Does the composition of the Supervisory Team comply with regulations?

YES		Supervisory Team approved
NO		Supervisory Team not approved

Comments:

Based on study progress, students is …

	ALLOWED to start the graduation project
	NOT allowed to start the graduation project

Comments:

Sign for approval (BoEx)

Name _____ Date _____ Signature _____

Personal Project Brief – IDE Master Graduation Project

Name student _____ **Student number** _____

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT

Complete all fields, keep information clear, specific and concise

Project title _____

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

➔ *space available for images / figures on next page*

introduction (continued): space for images

image / figure 1

image / figure 2

Personal Project Brief – IDE Master Graduation Project

Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.
(max 200 words)

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence) As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

Project planning and key moments

*To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include a **kick-off meeting**, **mid-term evaluation meeting**, **green light meeting** and **graduation ceremony**. Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities).*

Make sure to attach the full plan to this project brief.
The four key moment dates must be filled in below

Kick off meeting _____

Mid-term evaluation _____

Green light meeting _____

Graduation ceremony _____

In exceptional cases (part of) the Graduation Project may need to be scheduled part-time. Indicate here if such applies to your project

Part of project scheduled part-time	
For how many project weeks	
Number of project days per week	

Comments:

Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.
(200 words max)