DEMONSTRATION

Visiobo Demo: Augmenting Static Prints with Projection-based Visual Cueing and Concept Mapping via LLM Reasoning

JIAQI JIANG, Southern University of Science and Technology, Shenzhen, Guangdong, China

KEXIN HUANG, Southern University of Science and Technology, Shenzhen, Guangdong, China

HANQING ZHOU, Southern University of Science and Technology, Shenzhen, Guangdong, China

HUIYING LU, Harbin Institute of Technology, Harbin, Heilongjiang, China

PENGCHENG AN, Southern University of Science and Technology, Shenzhen, Guangdong, China

Open Access Support provided by:

Southern University of Science and Technology

Harbin Institute of Technology

CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (April 2025)
https://doi.org/10.1145/3706599.3721169
ISBN: 9798400713958

Visiobo Demo: Augmenting Static Prints with Projection-based Visual Cueing and Concept Mapping via LLM Reasoning

Jiaqi Jiang*
School of Design
Southern University of Science and Technology
Shenzhen, China
jiangjiaqi1206@gmail.com

Kexin Huang*
School of Design
Southern University of Science and Technology
ShenZhen, China
12333220@mail.sustech.edu.cn

Hanqing Zhou
School of Design
Southern University of Science and Technology
ShenZhen, China
12331483@mail.sustech.edu.cn

Huiying Lu
Harbin Institute of Technology
Harbin, China
24s034086@stu.hit.edu.cn

Pengcheng An†
School of Design
Southern University of Science and Technology
Shenzhen, China
anpc@sustech.edu.cn

(a) (b)

Figure 1: (a) A typical interaction scenario with Visiobo, showcasing its ability to talk with users and project visual enhancements. (b) System components, including the camera, projector, ultrasonic switch, power source, and mobile chassis.

Abstract

Posters are widely used for information dissemination, but their static nature and limited interactivity often hinder engagement and personalized learning experiences. We introduce a guide agent Visiobo designed to enhance poster reading by providing personalized support and dynamic visual augmentations. Visiobo overlays cueing (e.g., highlights, arrows) to help users quickly locate key information and employs concept mapping to address the content limitations of static posters by offering structured, relevant information. The user study with 12 participants demonstrated that by dynamically adapting to user interests and seamlessly integrating voice interaction with visual overlay cueing and concept mapping, Visiobo provides an engaging and enriched poster reading experience.

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

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3721169

CCS Concepts

• **Human-centered computing → HCI design and evaluation methods**; **User interface design**; • **Computing methodologies → Mixed / augmented reality**.

Keywords

Interactive posters, Visual enhancements, Conversational agents, Cueing

ACM Reference Format:
Jiaqi Jiang, Kexin Huang, Hanqing Zhou, Huiying Lu, and Pengcheng An. 2025. Visiobo Demo: Augmenting Static Prints with Projection-based Visual Cueing and Concept Mapping via LLM Reasoning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3706599.3721169

1 INTRODUCTION

Posters serve as a common medium for information dissemination, commonly employed in contexts such as science popularization exhibitions, educational activities, academic conferences, and related settings. They utilize text and rich visual elements to facilitate the spread and exchange of knowledge and information [4]. However, research has pointed out that posters receive limited attention due to their inherent constraints in length and lack of interactivity [7]. For example, different readers have varying points of interest in the content displayed on posters. However, the content of posters is static and unchangeable, and it is impractical for creators to consistently accompany their work and interact with the audience [9]. Some research has explored the use of conversational agents to meet the personalized requirements of users, boost interactivity, and sustain interest levels [6]. However, these solutions predominantly reside in textual or auditory modalities and have yet to exhaustively explore the potential of visual enhancements. In the context of museums, augmented reality (AR) glasses have been used to provide additional guidance or supplementary information for exhibits visually [10], though these experiences are typically limited to individual users, limiting their effectiveness in the context of reading posters — a public and social activity. Furthermore, in current reading and learning scenarios, visual enhancements usually focus on the area around physical materials, with little attention given to enhancing the texts themselves.

In response to the limitations of static posters and the need for a personalized and interactive reading experience, we propose Visiobo, a guide agent designed to engage viewers through dialogue and real-time visual enhancements. Visiobo facilitates conversational learning by enabling users to ask questions and interact based on their interests. Simultaneously, it dynamically generates personalized visual information enhancements tailored to the content being discussed through real-time reasoning. Two primary forms of visual enhancements are introduced: (1) Overlay cueing involves visual indicators such as arrows, highlights, or blinking signals that highlight critical information on the poster. Prior research suggests that cueing serves as a cognitive indicator, significantly aiding learners in focusing on essential information and reducing cognitive overload [11]. By emphasizing key elements, this method has potential to enhance attention allocation and improves learning efficiency. (2) Concept mapping is employed to expand and organize related learning materials, offering users a structured view of the expanded content. Leveraging the capabilities of large language models (LLMs) and Retrieval-Augmented Generation (RAG) technologies, the system can generate scalable and personalized responses.

2 WALKTHROUGH AND SETUP

2.1 Walkthrough

To provide a clear understanding of Visiobo, we provide a detailed description of the interaction process during a typical session.

We constructed a small exhibition featuring three posters related to the history of design, simulating the poster reading experience in a real-world scenario. When users enter the exhibition area, Visiobo directs them to a specific poster location (Figure 2-a). Upon reaching the location, Visiobo introduces the poster content and invites users to engage in a dialogue (Figure 2-b). After browsing the poster, users can initiate a question by hovering their hands over a non-contact switch. This triggers a dialogue system, allowing users to inquire about specific aspects of the poster content that capture their interest (Figure 2-c). Visiobo responds by combining user queries with poster information, while visual cues guide users to locate and confirm areas on the poster relevant to the agent's responses. It overlays cognitive visual cues, such as arrows, highlights, and blinking effects, directly onto the physical poster content with word-level precision. These cues help users quickly locate relevant key information and provide real-time feedback on the discussion (Figure 2-d). Additionally, Visiobo presents supplemental information in nearby blank areas, such as background details, profiles of related individuals and their connections. This concept mapping, retrieved from external knowledge databases, enriches the discussion by covering content that complements the poster but is not directly included in it (Figure 2-e). This approach allows users to access additional information beyond the poster that aligns with their interests.

2.2 Installation Setup

The system is composed of several primary components (Figure 1-b): a mobile chassis, a mobile power source, a projector, a camera, an ultrasonic sensor to create a non-contact switch, and a computer. The mobile base and power source enable the system to move within the exhibition space, while the projector can accurately project visual cues and supplementary materials. The camera is used to calibrate the position between the projector and the poster. The non-contact switch, made with an ultrasonic sensor, allows users to interact with the system without the need to touch the device, and its casing is produced using 3D printing technology.

During the interaction process, the projector is positioned facing the poster and the ultrasonic switch faces the user. Users stand at the side or rear of the system, which conveniently facilitates triggering dialogue through the switch and allows for easy viewing of both the poster and the projected content. Integrated into the camera is a microphone for audio input, while speakers on the projector handle audio output.

3 TECHNICAL REALIZATION

Our technical framework consists of three main modules: the voice module, the cueing and dialogue generation module, and the concept mapping generation module (Figure 3). In voice module, the speech-to-text functionality employs OpenAI's Whisper [2], while

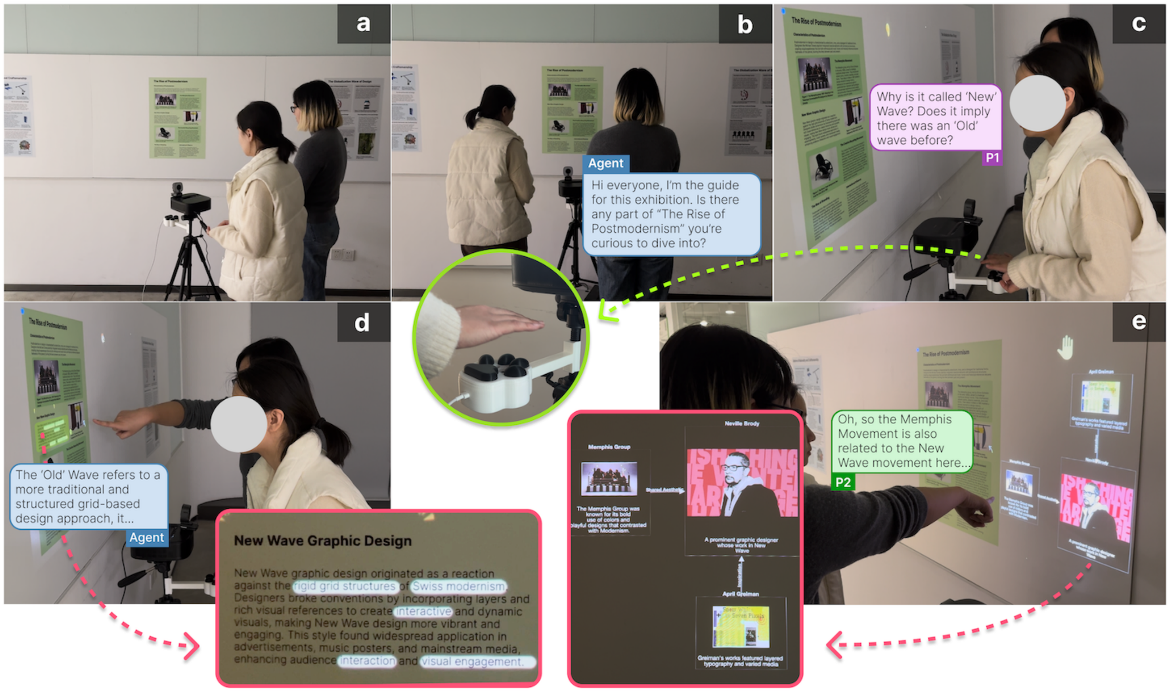

Figure 2: Overview of the interaction with Visiobo in the exhibition scenario. (a) Users are guided to a specific poster location upon entering the exhibition area. (b) Visiobo introduces the poster content and encourages users to engage in dialogue. (c) Users initiate questions using a non-contact switch, triggering a voice-based interaction. (d) Visual cues such as highlights, arrows, and blinking effects are overlaid on the poster. (e) Supplemental information, retrieved from external knowledge databases, is displayed in the form of concept mapping around the poster.

text-to-natural speech conversion utilizes the open-source model CosyVoice [1]. Throughout the process, all API requests for the large language models used in the system are powered by GPT-4o. All visual content is displayed through a web-based application and projected onto the poster and its surrounding area. The system's frontend is implemented using the React framework, while the backend is powered by Python in combination with FastAPI. Now we will focus on the functionalities and designs of the cueing generation module and the concept mapping generation module.

3.1 Cueing and Dialogue Generation

To achieve precise location targeting of poster elements, it is necessary to preprocess the digital poster to extract text content and corresponding spatial location information. Firstly, Optical Character Recognition (OCR) technology is used to extract the precise coordinates of each word in the poster. Following this, spatial clustering analysis of these words identifies different paragraphs and their positions. These paragraphs are then input into a Large Language Models (LLMs), which conducts semantic analysis to categorize the paragraph content into labels such as "title", "subtitle", "content", or "caption", grouping labels with similar semantics. This method effectively combines the precision recognition capability of traditional visual text processing technologies with the deep semantic understanding of large language models. Lastly, the extracted labels, text, and location information are integrated into a JSON file, providing data support for subsequent retrieval and interaction.

During interaction with Visiobo, the LLM infers and extracts relevant keywords based on the dialogue content. These keywords are used to search for corresponding positions and labels in the JSON file. Based on the search results, Visiobo dynamically projects visual cues through the projector, such as highlighting keywords within paragraphs, adding blinking effects to titles, or attaching direction arrows to images, thereby directing the user's attention to relevant content on the poster. In addition to these visual cues, the module also generates dialogue responses that align with the extracted keywords and the context of the ongoing conversation. The final response is converted to speech and played through the projector's speakers.

3.2 Concept Mapping Generation

To complement the poster content and provide a personalized learning experience, Visiobo integrates information related to the poster content through Retrieval-Augmented Generation (RAG) technology. In the previously mentioned poster exhibition scenario, the content of the book "History of Design" is vectorized to support efficient semantic similarity calculations. During user interactions, when the user engages with a particular topic or asks a question,

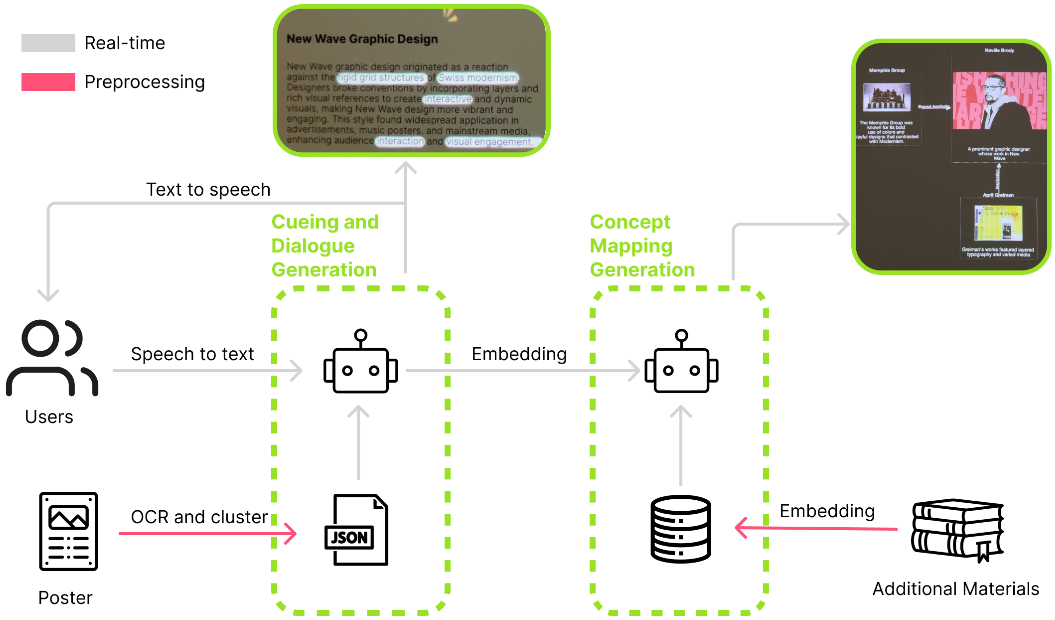

Figure 3: Software framework of the system: illustrating the integration of voice interaction, cueing generation, and concept mapping modules.

Visiobo semanticly compares the current discussion topic with the content in the vector library, to extract relevant information in real-time. It identifies texts, images, and uses large language models to uncover and establish relationships between them. Specifically, the model outputs a JSON content containing multiple nodes, each representing an information point, with attributes like title, description, image (optional), and relationships to other nodes.

The extracted content and images are processed by the fCoSE layout algorithm[3] and presented in the form of a relational graph, offering users a clear and structured supplementary information view.

4 USER STUDY

We conducted a user experiment with 12 participants, aged between 18 and 44 years, with diverse backgrounds in design, engineering, and education. Each participant experienced three experimental conditions: voice only, voice with cueing, and voice with cueing and concept mapping, in a within-group controlled design. In each condition, participants interacted with Visiobo alongside their group members to collaboratively read and comprehend a poster. The order of conditions was fully counterbalanced across participants to minimize order effects, with a minimum interval of two hours between successive conditions.

After completing each experimental condition, participants filled out the User Engagement Scale (UES) [8], the System Usability Scale (SUS) [5], and a custom-designed questionnaire comprising six items (see Appendix A). At the end of the final condition, a semi-structured interview was conducted with each participant. During the interview, participants were invited to share their overall impressions of the experiment, their perceptions of the different functional modules, and the impact of the system on their interactions with group members.

Through quantitative analysis, the data revealed significant differences in the following aspects: (1) Condition 3 (Voice + Cueing + Concept Mapping) demonstrated a significant improvement over Condition 1 (Voice) ($p = 0.034$) in user engagement. (2) Condition 3 also showed a substantial enhancement compared to Condition 1 ($p < 0.001$) in providing personalized and extended information. (3) Both Condition 2 (Voice + Cueing) and Condition 3 significantly improved collaborative understanding of the poster content compared to Condition 1, with p-values of 0.009 and < 0.001, respectively. Furthermore, the mean System Usability Scale (SUS) scores for all three conditions exceeded 70 (out of 100), indicating that the system performed well in terms of usability across all conditions. Analysis of user interviews revealed that participants noted that cueing effectively captured their attention, and concept mapping not only provided more relevant information tailored to their interests but also added novelty and enjoyment to their experience.

5 CONCLUSION

This paper presents Visiobo, a guide agent that combines conversational interaction with visual enhancements, transforming static posters into interactive and engaging experiences. We explored a method that integrates precise positional data from traditional visual recognition with semantic understanding powered by large language models, enabling the system to accurately perceive elements on physical posters and precisely overlay cueing onto the corresponding locations. Additionally, by using the discussed content as a retrieval base for querying a vectorized database, Visiobo provides supplementary knowledge displayed around the physical poster. User experiments demonstrated that combining dynamic cueing and concept mapping enables Visiobo to deliver real-time personalized experiences, richer knowledge supplementation, and enhanced engagement. Future work will explore the system's impact on learning outcomes, particularly on comprehension, and enhance the agent's mobility for autonomous navigation.

References

[1] 2025. FunAudioLLM/CosyVoice. FunAudioLLM.
[2] 2025. Openai/Whisper. OpenAI.
[3] Hasan Balci and Ugur Dogrusoz. 2022. fCoSE: A Fast Compound Graph Layout Algorithm with Constraint Support. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (Dec. 2022), 4582–4593. doi:10.1109/TVCG.2021.3095303
[4] Larissa D'Angelo et al. 2010. Creating a framework for the analysis of academic posters. *Language Studies Working Papers* 2 (2010), 38–50.
[5] James R Lewis. 2018. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590.
[6] Chen-Chung Liu, Mo-Gang Liao, Chia-Hui Chang, and Hung-Ming Lin. 2022. An analysis of children'interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education* 189 (2022), 104576.
[7] Anu MacIntosh-Murray. 2007. Poster presentations as a genre in knowledge communication: a case study of forms, norms, and values. *Science communication* 28, 3 (2007), 347–376.
[8] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.
[9] Nicholas Rowe and Dragan Ilic. 2015. Rethinking poster presentations at large-scale scientific meetings–is it time for the format to evolve? 3661–3668 pages.
[10] Wilbert Tabone. 2020. The effectiveness of an augmented reality guiding system in an art museum. *Rediscovering heritage through technology: A collection of innovative research case studies that are reworking the way we experience heritage* (2020), 197–214.
[11] Heping Xie, Fuxing Wang, Yanbin Hao, Jiaxue Chen, Jing An, Yuxin Wang, and Huashan Liu. 2017. The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: A meta-analysis and two meta-regression analyses. *PloS one* 12, 8 (2017), e0183884.

A Custom Questionnaire

A1: It is easy for me to confirm whether the voice agent understands my question.
A2: I can clearly identify which part of the poster the voice agent is referring to during the discussion.
A3: The system provides information based on the topics I am interested in.
A4: The system provides information that is not included on the poster.
A5: I had meaningful discussions with my partner about specific topics related to the poster.
A6: I could clearly understand which part of the poster my partner was focusing on.

Each statement has five response options: Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree.