



香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

信息枢纽
INFORMATION HUB

Collaborative
Interactive
Visualisation
Analysis
Laboratory

多模态文化遗产数据 可视化与交互设计

Zeng Wei, Assistant Professor

The Hong Kong University of Science and Technology (Guangzhou)

Jun. 28, 2024



协同交互可视分析实验室

CIVAL @ 香港科技大学 (广州)



目标: 研究和开发**可视化与交互工具**, 提高分析效率, 促进信息交流。

 Collaborative
Interactive
Visualisation
Analysis
Laboratory



**PI: 曾伟博士, 助理教授 @ 信息枢纽, 计算媒体与艺术学域 (CMA)
& 数据科学与分析学域 (DSA)**

博士生:

2021级: 叶依林 (CMA)

2022级: 黄镛 (CMA), 王湛 (CMA), 侯伊涵 (CMA), 郝佳凝 (DSA)

2023级: 于健 (CMA), 杨旨窃 (CMA), 曾星辰 (DSA), 王梁炜 (DSA)
李玲 (CMA), 王钰淞 (CMA)

硕士生:

2022级: 肖诗诗 (CMA), 崔昊 (CMA), 陈奕涵 (CMA),
陈钊滢 (DSA), 方文婧 (DSA), 林海川 (CMA)

2023级: 杨曼玲 (CMA), 张议文 (CMA), 李春婷 (DSA)

研究助理/访问学生: 高子尧 (2023.06 -), 樊秋辰 (2024.05 -),
梁卓文 (2023.09 -), 刘元邦 (2024.05 -), 阮晨曦 (2024.05 -)

可视化与可视分析

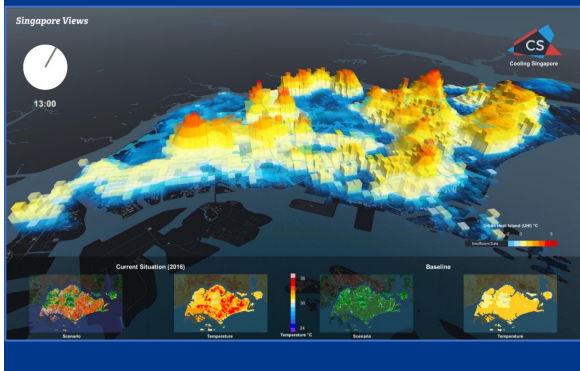
Visualization and Visual Analytics



- 通过对大数据进行分析与可视化，可以有效地展示数据模式，方便用户发现问题、探索方案影响，并做出最优决策，推动大数据研究。
- 通过信息可视化形式，可以让公众更好地参与到科技创新相关信息的交流中，进一步拓展公众在政策制定、教育普及中的参与和创新实践的作用，促进制造与服务业的透明度和公正性。

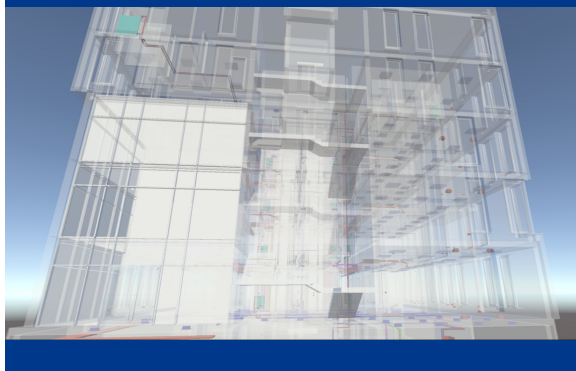
多维度

支持3D热力图、2D街景图、动态交互、三维建筑材料重利用等多维度任务分析



多尺度

支持区域、街道、人本等多尺度分析，从宏观到微观支持数据可视化



跨设备

支持不同类型的设备（电脑、手机、大屏幕、VR/AR头盔等）无缝展示与交互



协同式

支持政府、社区、企业、居民等之间开展合作和协调，共同参与政策的制定和实施



生成式模型在设计中应用

Generative AI for Design

研究背景

生成式人工智能（Generative AI）有能力生成新的数据、文本、图像以及其他各种类型的内容，适用于多个应用领域。

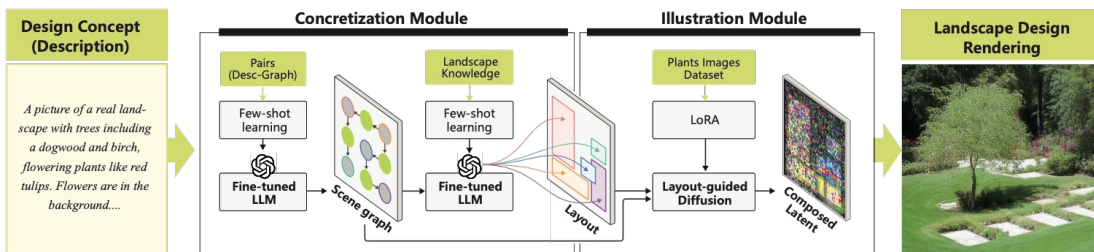
- 生成式语言模型（如GPT）可用于自动文本摘要、对话生成等。
- 文生图模型（如Midjourney）可用于生成图像、动画等创意作品。

挑战

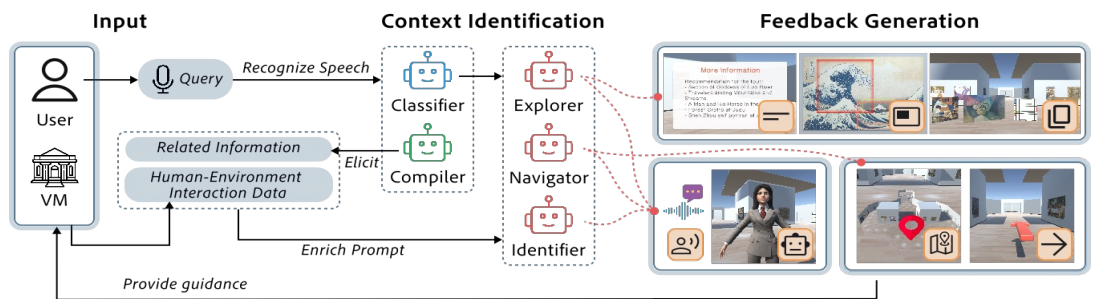
生成内容质量低、缺乏知识产权保护、以及伦理和道德问题。

- 生成内容与设计师意图对齐困难。
- 易侵害知识产权，难以保护数字资产免受未经授权的复制或分发。

将迭代设计过程融入到用于景观渲染的生成人工智能中



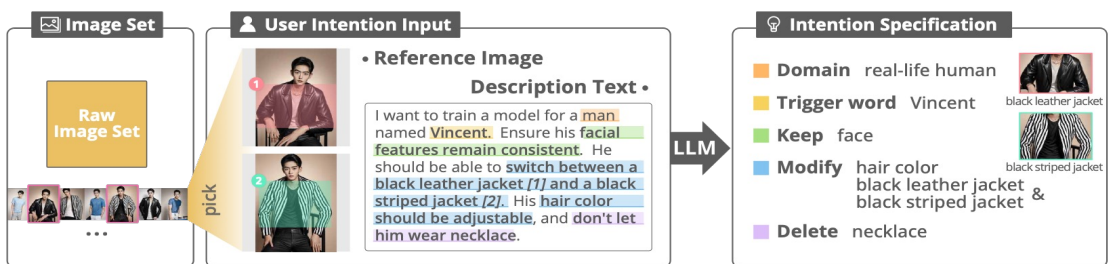
通过大型语言模型增强虚拟导览的多模态交互



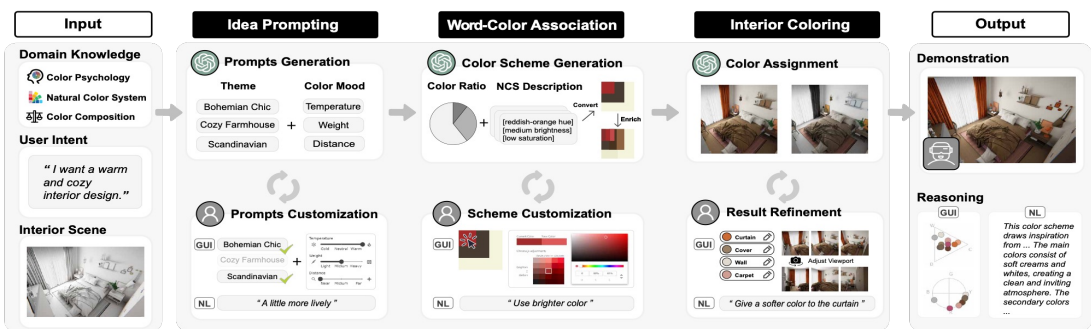
研究贡献

- 通过将传统设计流程与LLM等基座模型相结合，实现了在景观设计、室内色彩设计、虚拟导览等领域对大模型的应用。
- 增强了设计过程中模型与用户之间的交互性，通过更精准地捕捉和应用用户的意图，大型模型在设计中的角色得以进一步强化。
- 促进了设计领域中人机协作的进步，为创新性设计和跨领域合作提供了新的可能性。

在微调文本到图像生成模型中整合人类意图



利用大型语言模型支持创意室内色彩设计构思



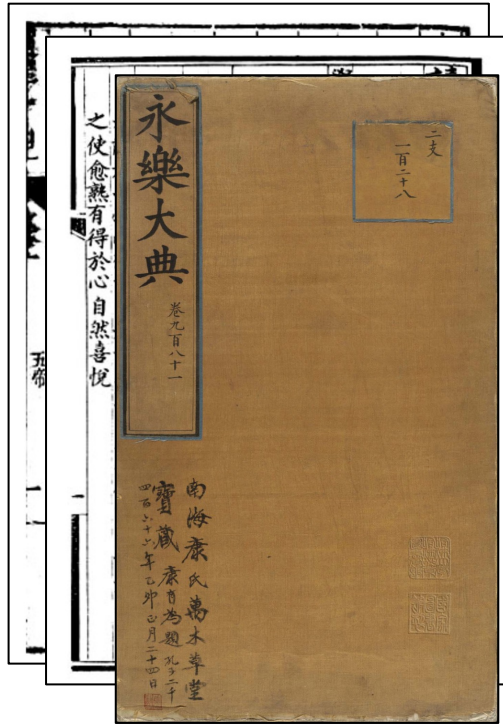


香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

信息枢纽
INFORMATION HUB

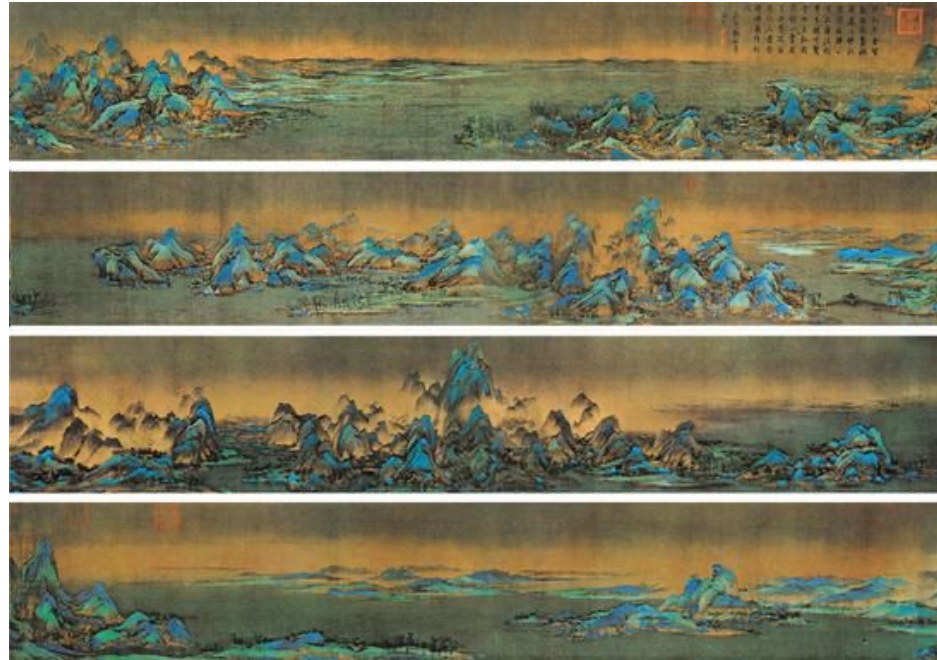
Multi-modal Cultural Heritage Data

- Text, image, 3D model, and other forms are essential components of cultural heritage (CH) data.



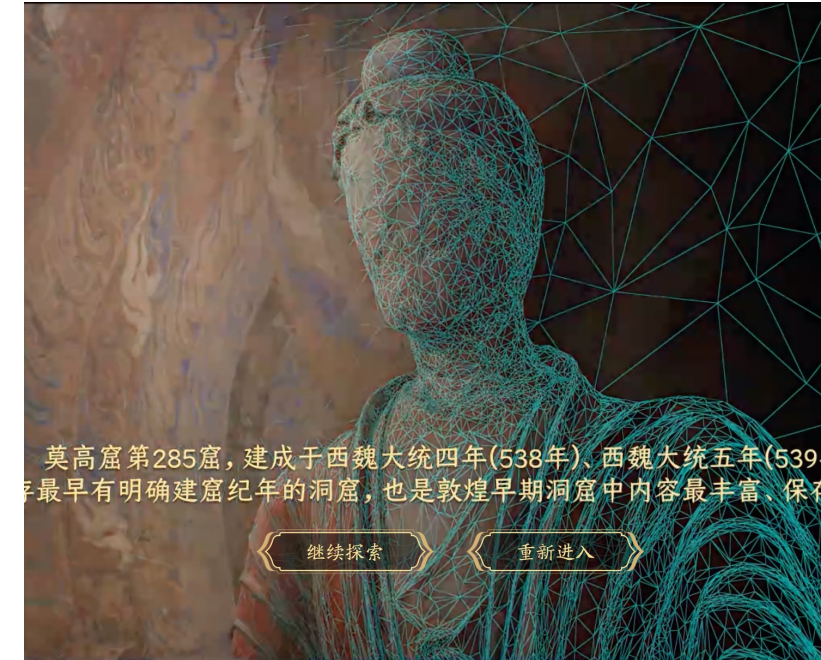
Text

cr: [中国国家数字图书馆](#)



Image

cr: [故宫博物院数字文物库](#)



3D model


cr: [数字敦煌](#)

Heritage Interpretation

- **Heritage interpretation** seeks to engage visitors and deepen their understanding of CH.
- Some key methods and strategies in heritage interpretation:
 - Multisensory experiences: Visual, auditory, tactile
 - Interaction and participation: Interactive exhibits
 - Storytelling: Guided tours, multimedia presentations
 - Virtual and augmented reality: virtual tours
- Multi-modal heritage data presents several challenges:
 - Data integration: compatibility issues, data management
 - User experience design: overloading visitors, engagement

Cultural Heritage Exploration

- Existing exploration system for CH data
 - Faceted exploration: hierarchical faceted metadata
 - Dynamically generated query previews
 - Keyword-based search




Digital Collections


[Library of Congress » Digital Collections](#)
[Share](#)

Digital Collections


Featured Content



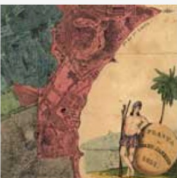
Historic American Buildings Survey/Historic American Engineering...




Chronicle America: Historic American Newspapers



Farm Security Administration/Office of War Information ...



Cities and Towns



Civil War Maps

Refine your results

Topic

American History

216

Government, Law & Politics

174

World Cultures & History

154

Performing Arts

108

War & Military

88

Local History & Folklife

66

Art & Architecture

57

Social & Business History

45

Geography & Places

33

Science & Technology

26

More Topics »

Part of

Digital Collections

120

Manuscript Division

75

Prints and Photographs Division

68

Music Division

46

American Folklife Center

39

Asian Division

29

African and Middle Eastern Division

29

Law Library of Congress

29

Digital Collections

View

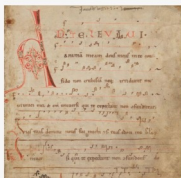
Gallery

Go


Sort By

Select


Go




COLLECTION
10th-16th Century Liturgical Chants
 The acquisition of medieval liturgical chant manuscripts that trace the history of music notation as it evolved over half a millennium, became a major collection priority in the Music Division beginning in...
Collection Items: [View 55 Items](#)



COLLECTION
A.P. Schmidt Company archives, 1869-1958
 Arthur Paul Schmidt (1846-1921) was a German-born music publisher who pioneered the development and dissemination of American music. The A.P. Schmidt Company Archives documents his firm's publishing activities in Boston, Leipzig and...
Collection Items: [View 6,687 Items](#)



COLLECTION
Aaron Copland Collection
 The first release of the online collection contains approximately 1,000 items that yield a total of about 5,000 images. These items date from 1899 to 1981, with most from the 1920s through...
Collection Items: [View 982 Items](#)




COLLECTION
Abdul Hamid II Collection
 This collection contains 1,819 photographs in 51 large-format albums date from about 1880 to 1893. They portray the Ottoman Empire during the reign of one of its last sultans, Abdul-Hamid II and...
Collection Items: [View 1,825 Items](#)

Results: 1-40 of 526 | Refined by:

Part of: Digital Collections

Available Online



Library of Congress (<https://www.loc.gov/collections/>)

7/50

VR/AR/XR

- Traditional exploration on the desktop.
 - WIMP (windows, icons, menus, pointer) metaphor.
 - Keyboard and mouse centric interaction.
- VR/AR/XR brings new opportunities for interacting with multi-modal data.
 - 3D visualization, mid-air interface
 - More natural interactions with hand gestures & voices



Minority Report, 2002

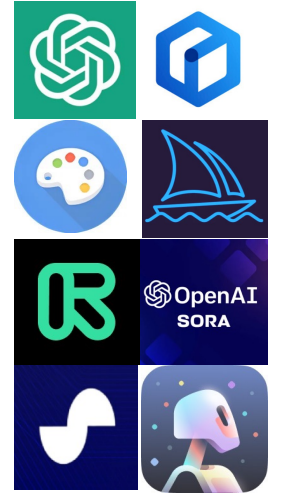


The Avengers, 2012



Generative AI

- Generative AI is a type of AI technology that can produce various types of new content based on patterns it learns from existing data.
 - Examples: GPT, 文心一言, Stable Diffusion, Midjourney, Runway, Sora, Suno, Wonder ...
- All the generated materials such as text, imagery, audio, and synthetic data, are considered AI Generated Content (AIGC).



Content produce methods



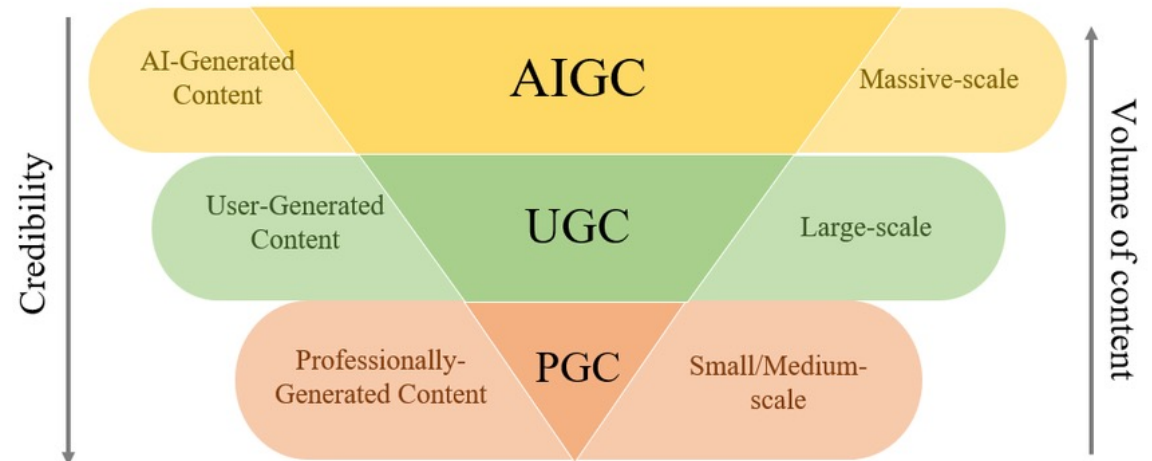
**Professional
Generated
Content**



**User
Generated
Content**

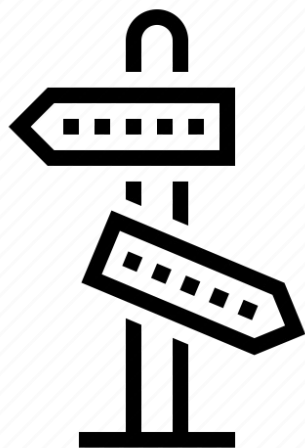


**AI
Generated
Content**



Tools for the Future

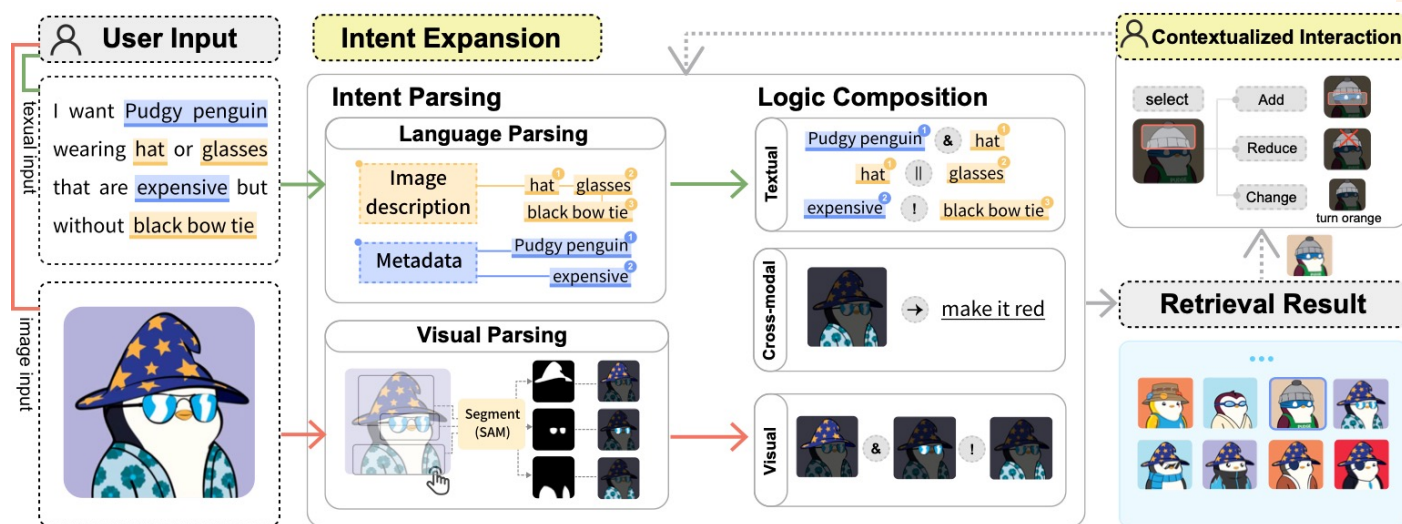
- The integration of VR/AR/XR, and generative AI has the potential to significantly reshape how we interact with multi-modal CH data.
 - Enhanced realism and immersion
 - Personalized experiences
 - Interactive and collaborative tools
- HKUST-CIVAL group dedicates to develop innovative visualization and interaction tools, aiming to create richer and more dynamic cultural heritage experiences.
 - The Contemporary Art of Image Search
 - Centennial Drama Reimagined
 - VirtuWander



The Contemporary Art of Image Search

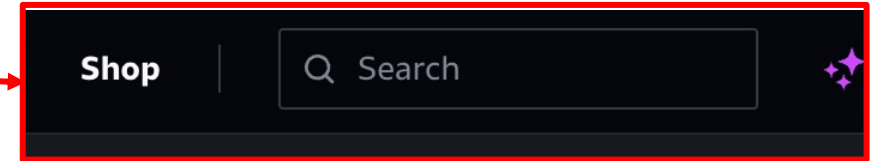
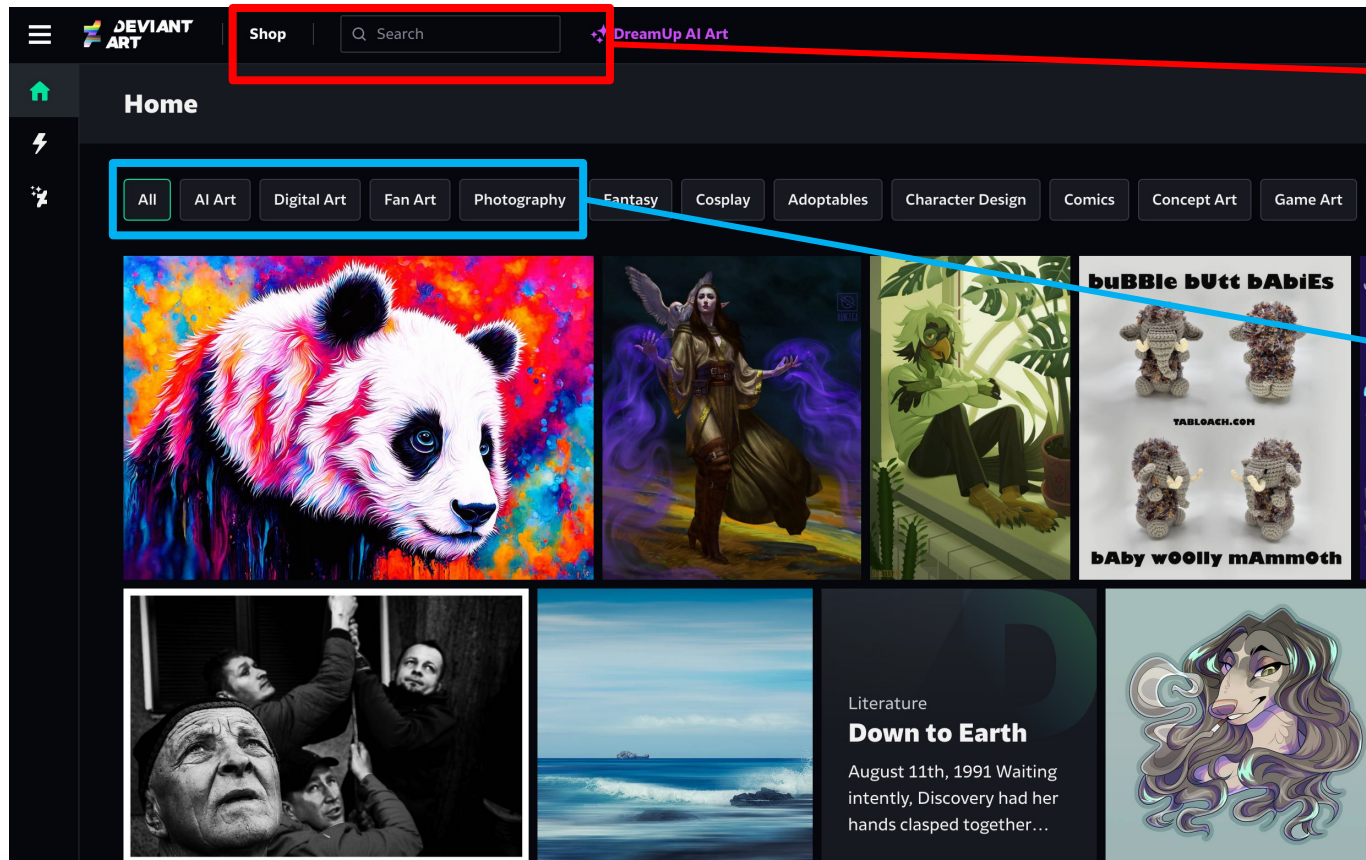
Iterative User Intent Expansion
via Vision-Language Model

CSCW 2024

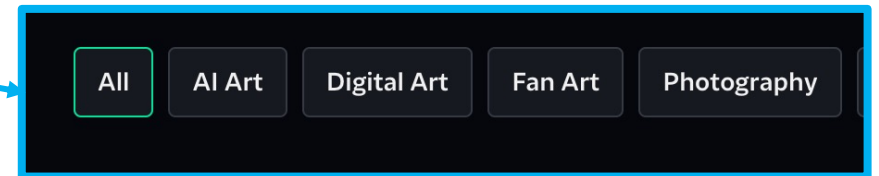


Background

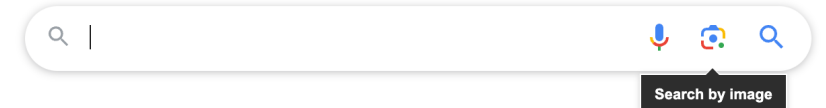
- Image search enables users to efficiently access paintings or designs of interest.



Search by 'text'



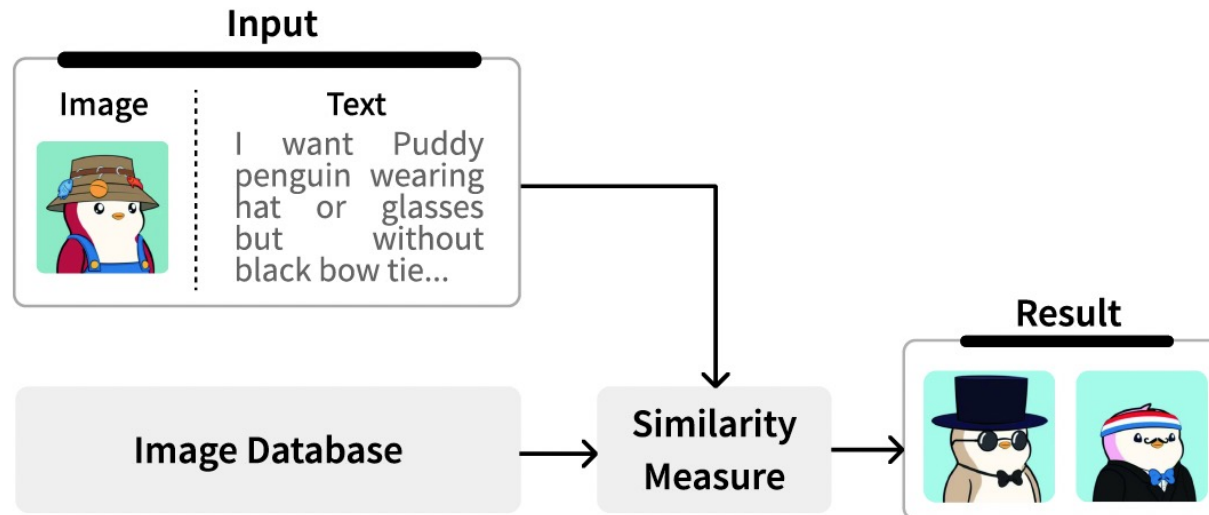
Search by 'category'



Search by 'image'

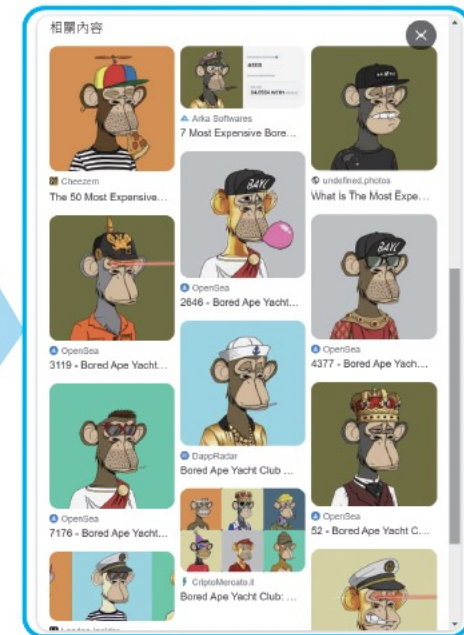
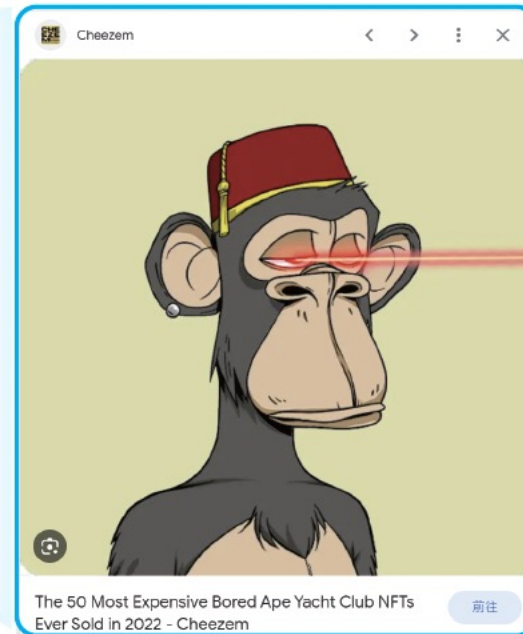
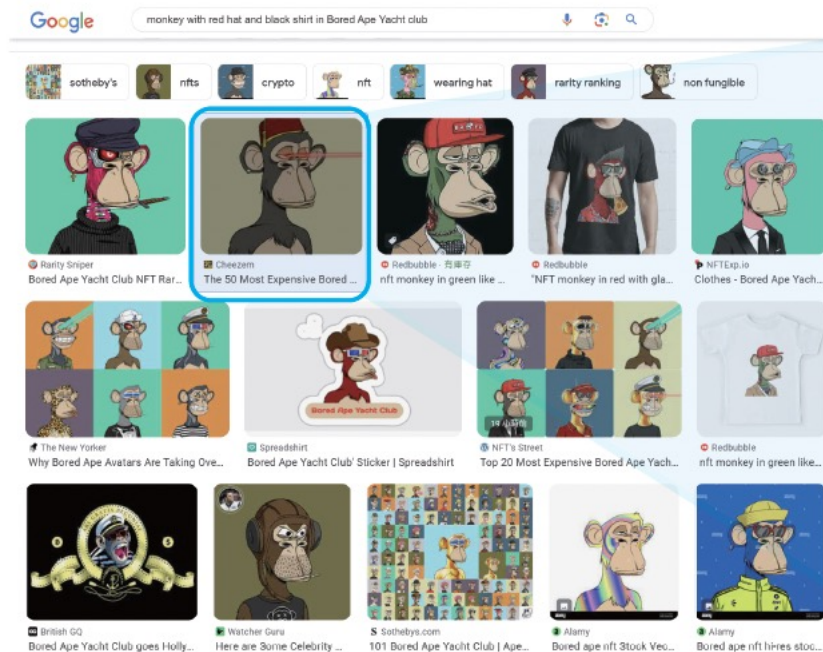
Background

- Conventional image search relies on ***proximity measurement*** between user inputs and image items in a gallery.
 - Text-to-image search
 - Rely on image labels
 - *Contrastive language image pretraining (CLIP)* for recent text-to-image models.
 - Image-to-image search



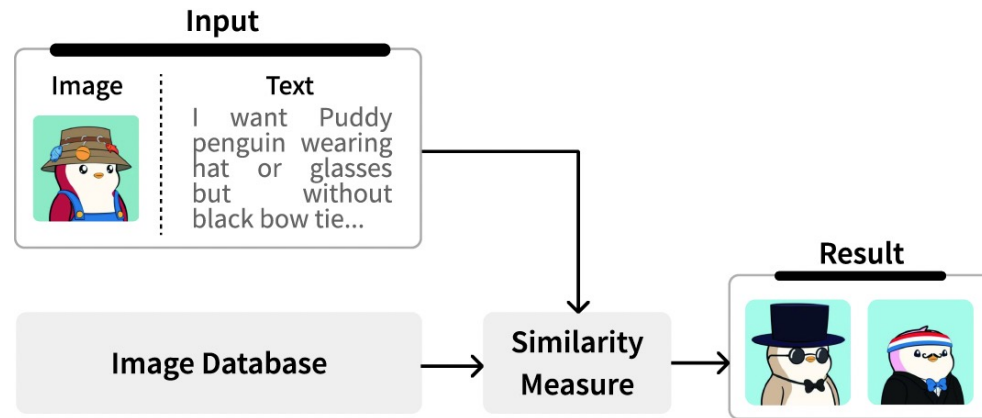
Background

- However, all methods have certain limitations
 - Lack sufficient *comprehension of user intent*.
 - Hardly support *contextualized interaction* to iteratively adjust or specify search intents.

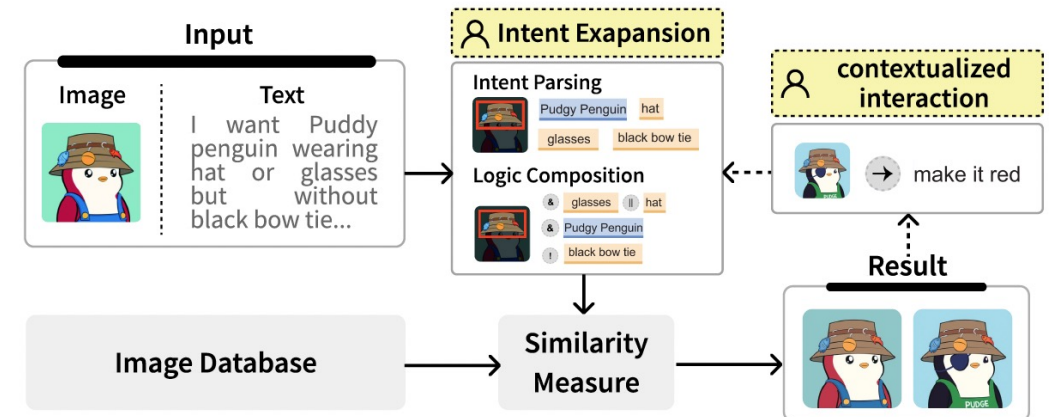


Motivation

- From traditional to contemporary image search
 - **G1: Accurate parsing of cross-modal user intents**
 - **G2: Logic expressions of user intents**
 - **G3: Contextualized cross-modal interaction**



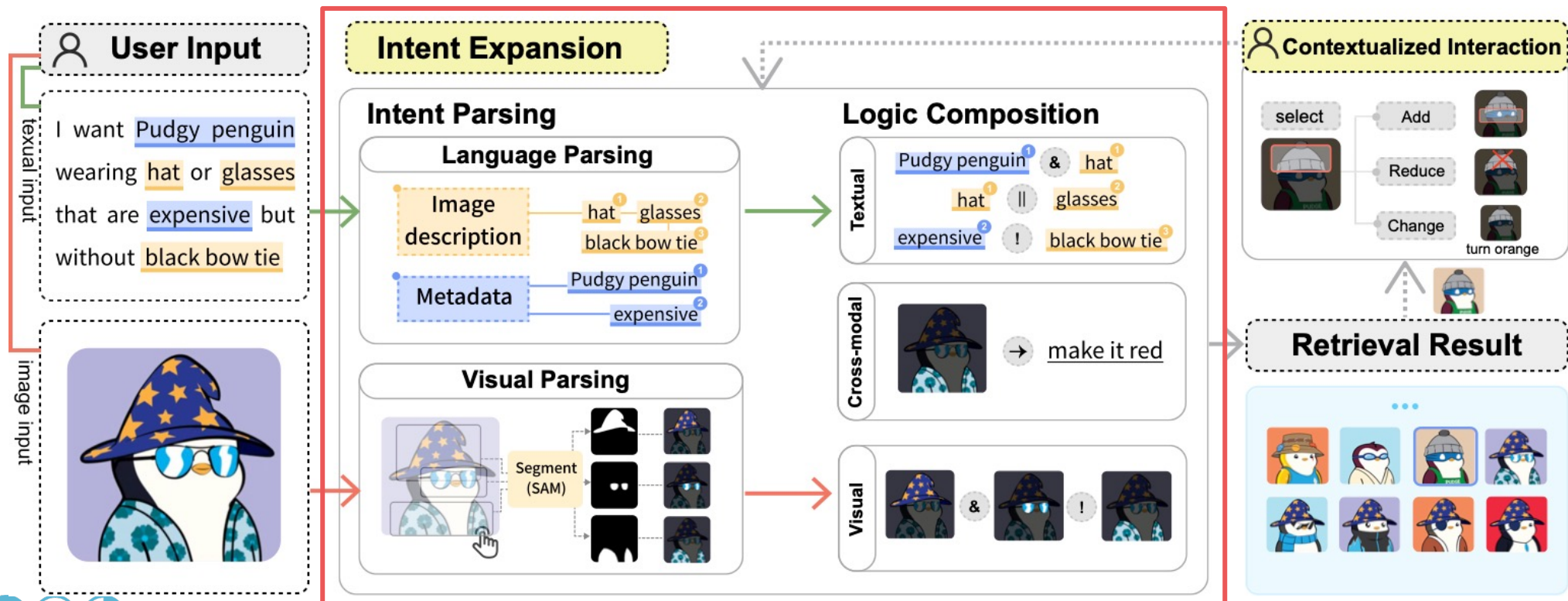
Traditional



Contemporary

Framework

1. **Intent expansion** improves comprehension of users' search intents
 - **Intent parsing** captures detailed search intents of different modalities
 - **Logic composition** grasps the logical relations to compose complex multi-element intention



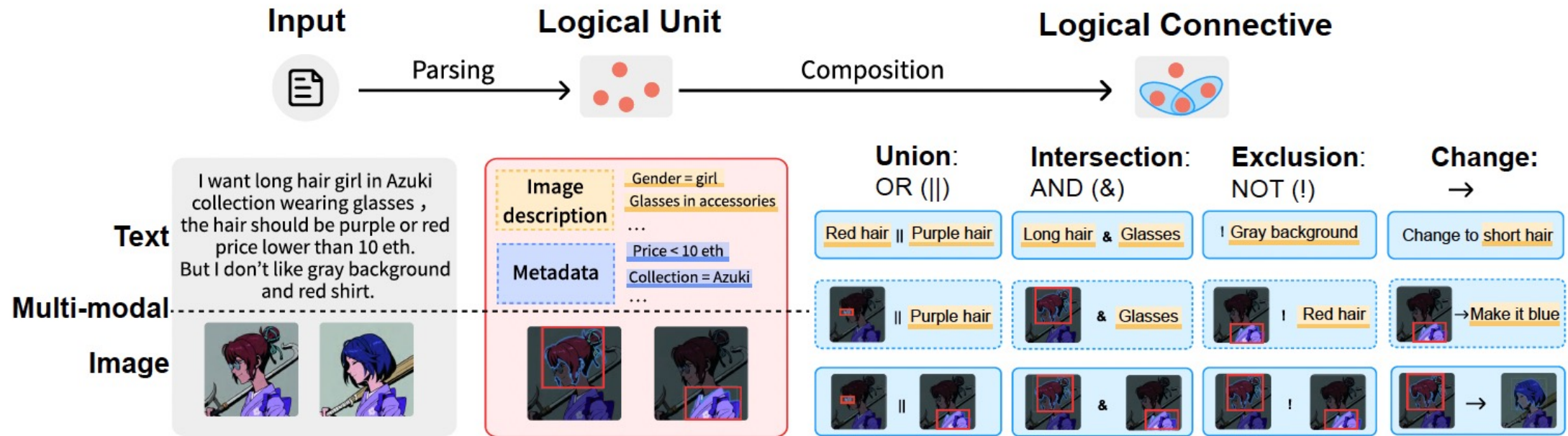
Framework

- **Visual Parsing** module leverages *Segment Anything* Model, allowing users to focus on particular visual elements by simple brushing on input image.
- (optional) Modification step allows users to express retrieval logic on the selected elements.
- The query is fed to database to retrieve images according to CLIP similarity.



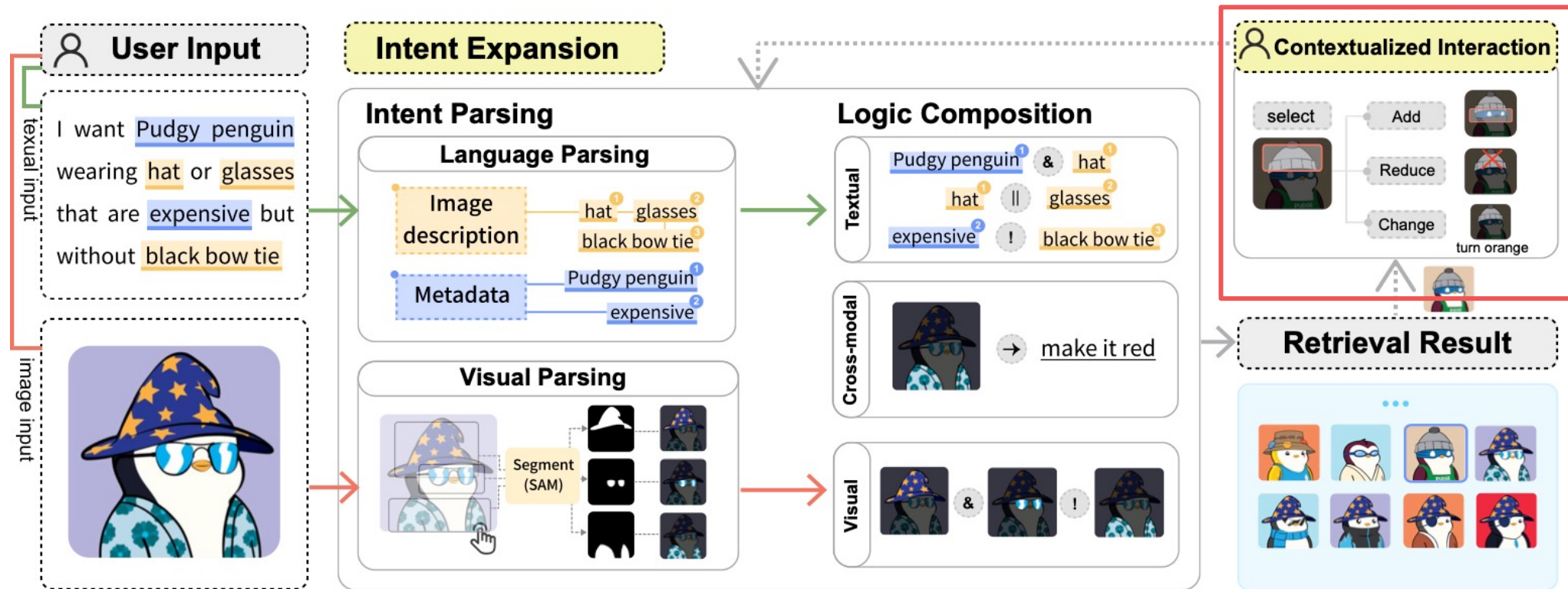
Framework

- **Logical composition** in our framework includes four major logical connectives:
 - union, intersection, exclusion and change



Framework

2. **Contextualized interactions** allow users to directly operate on result images to iteratively refine their search.



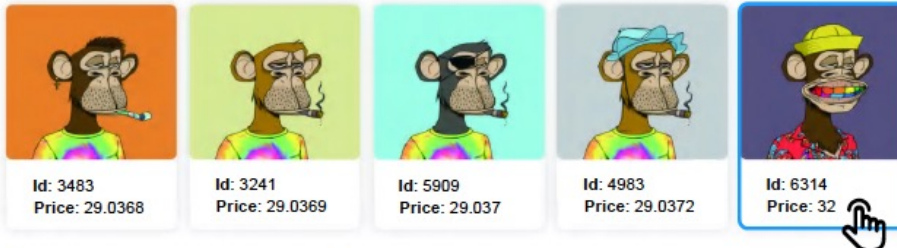
Examples

- Text search on NFT images

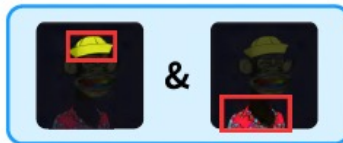
Text Input

Smoking monkey wearing colorful shirt and eye mask in Bored Ape Yacht Club

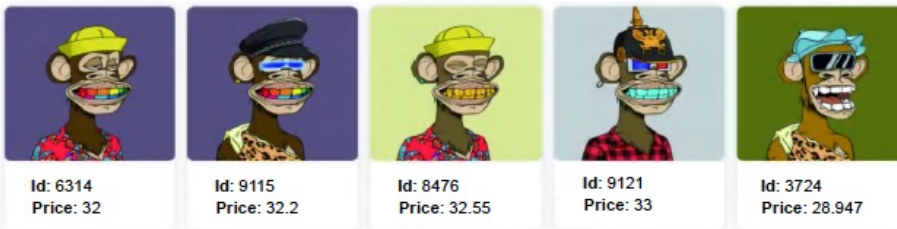
First-round
Retrieval
result



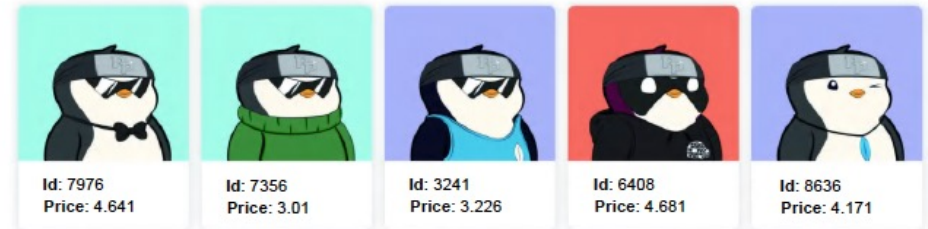
Visual
parsing



Second-round
Retrieval
result



Give me some expensive penguins wearing glasses



Examples

- Logic composition on NFT images

Text Input

Woman in pixel style.

Language parsing

woman & pixel style

Retrieval result



Woman in pixel style,
but no black hair.

woman & pixel style &! black hair



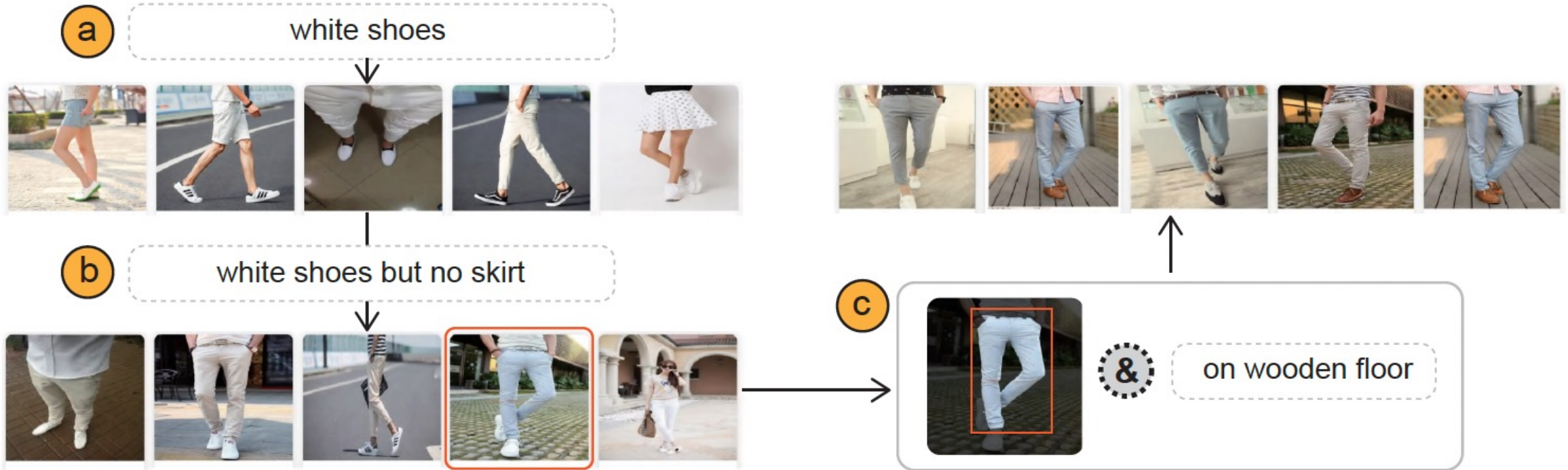
Woman in pixel style,
but no black hair or smoking.

woman & pixel style
&!(black hair || smoking)



Examples

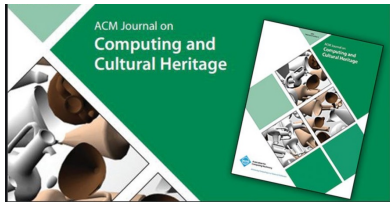
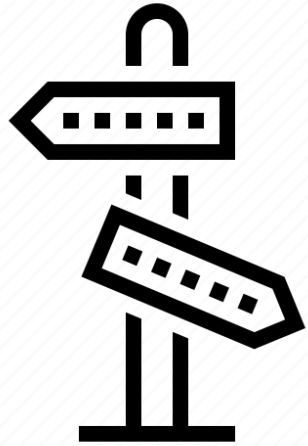
- More examples on fashion images.



Use Cases Demo

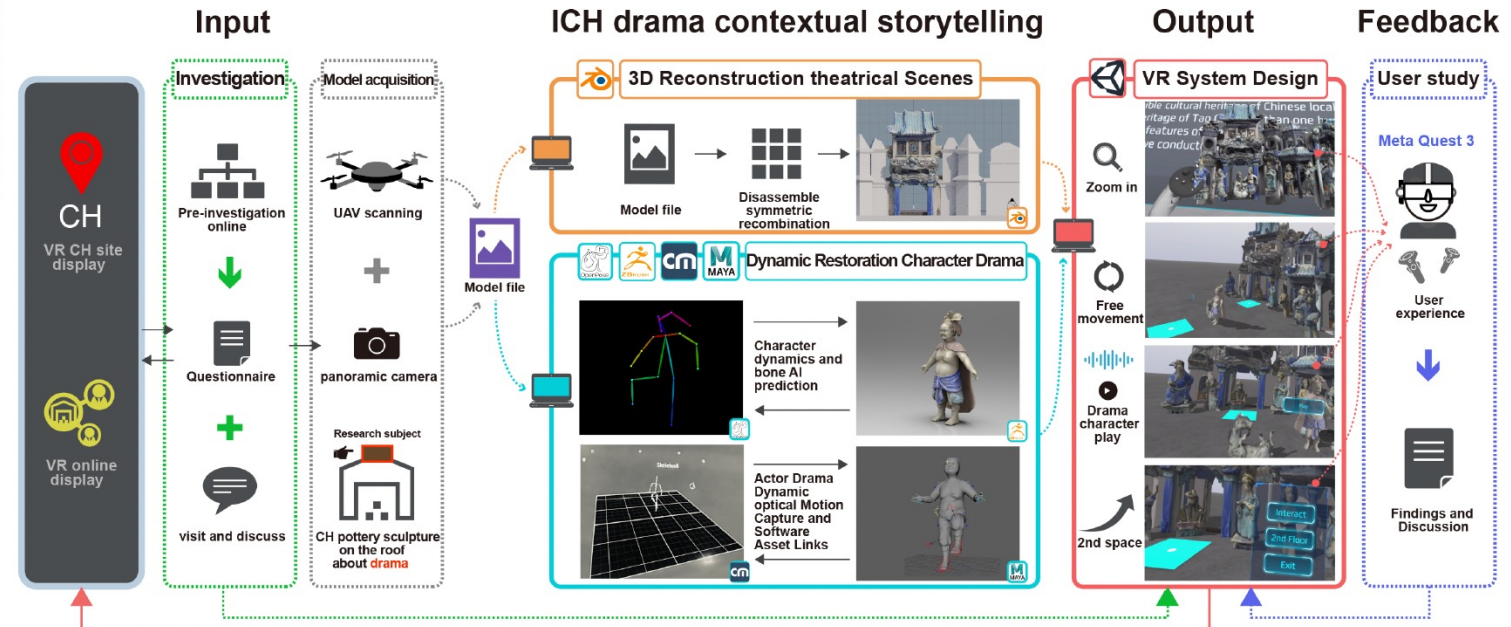
*The Contemporary Art of Image Search:
Iterative User Intent Expansion via Vision-Language Model*

CSCW 2024



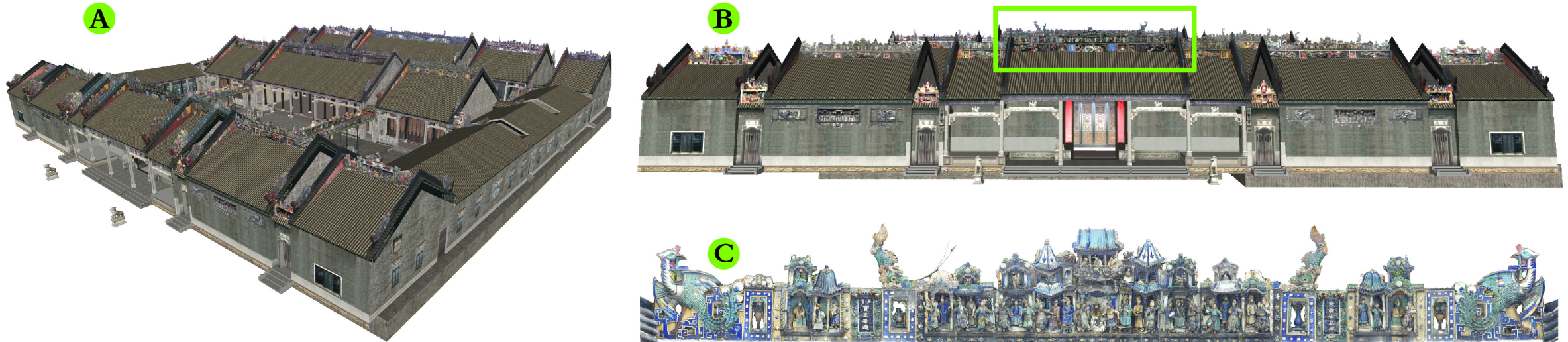
Centennial Drama Reimagined

An Immersive Experience of Intangible Cultural Heritage through Contextual Storytelling in Virtual Reality



Background

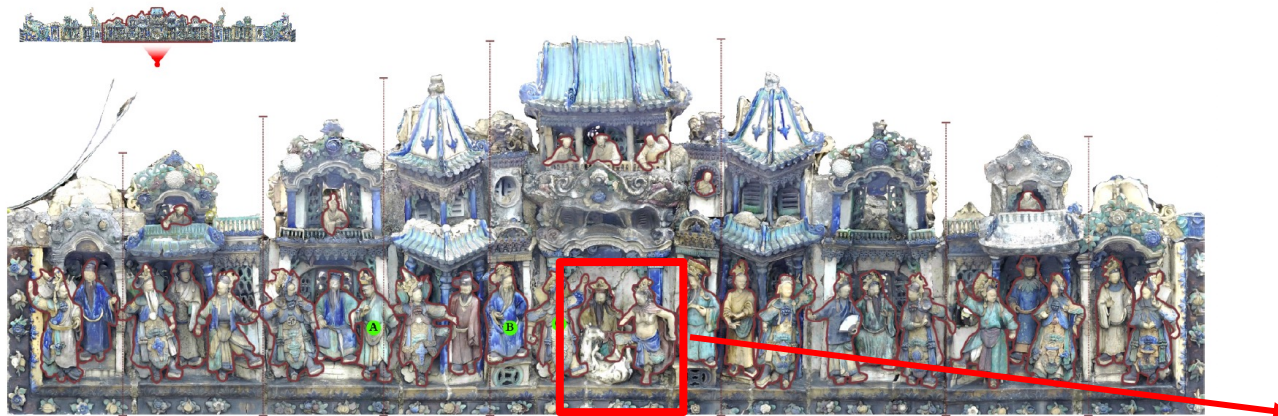
- Shiwan pottery sculptures (石湾瓦脊) are predominantly located on the roofs of Lingnan buildings (岭南建筑), several meters high, and are used for decoration and blessings.
- Chen Clan Ancestral Hall (陈家祠) has a history of over a hundred years.



3D prototype view of Chen Clan Ancestral Hall

Background

- Cantonese opera (粵劇) is a distinct genre of drama known for its diverse theatrical performances and storytelling arts.
- "Li Yuanba Subduing the Dragon Colt" (李元霸降龙伏驥).



The relationships of gazes among pottery sculpture characters, and the grouping diagram of character plots



Dramatic characters and spatial storytelling relationship



"Small Sheng" 小生



"Old Sheng" 老生



"Wu Sheng" 武生

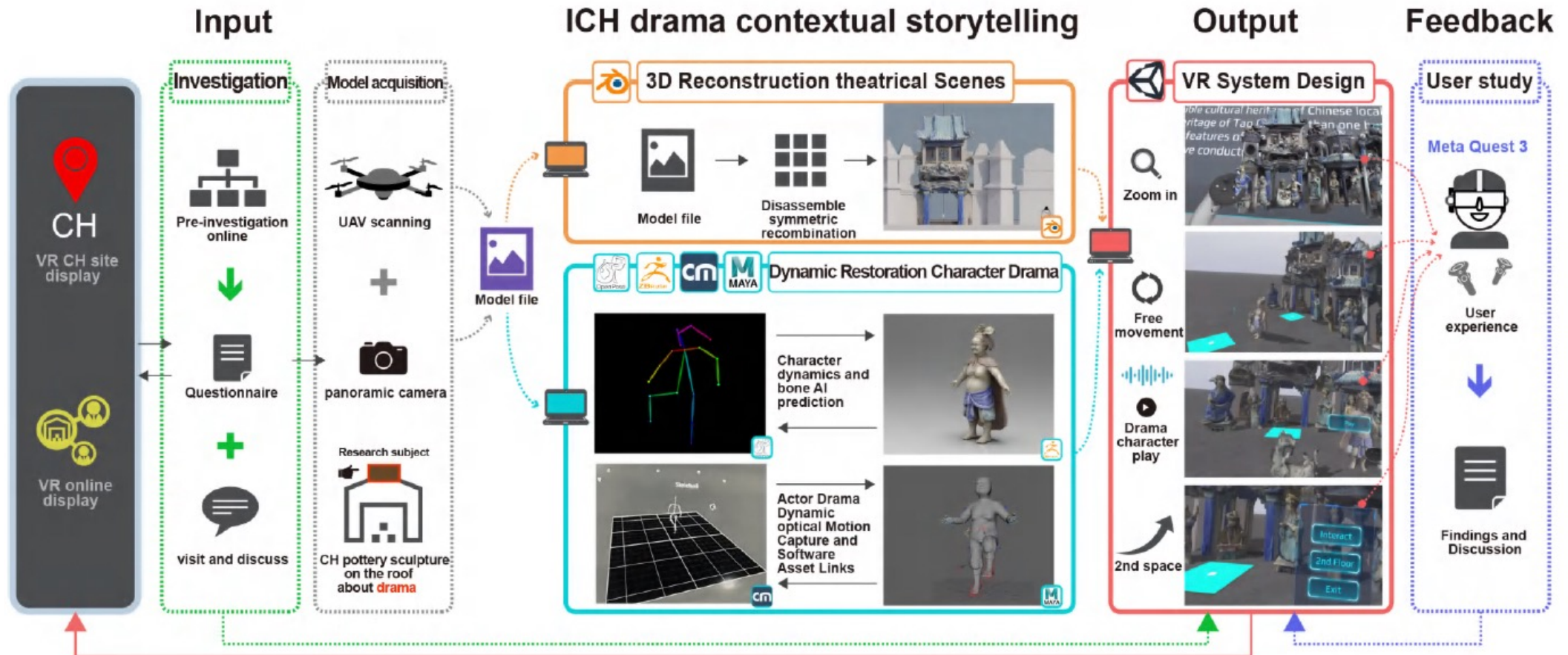


Motivation

- **Goal:** to utilize VR contextual storytelling techniques to address these challenges and vividly recreate the vibrant scenarios characteristic of ICH.
 - RQ1: What kind of knowledge and experience are audiences seeking at cultural heritage sites to learn about ICH?
 - RQ2: How should contextual storytelling techniques be applied to reconstruct the dynamic scenarios and narratives within ICH virtually?
 - RQ3: What essential design factors should be considered to enhance audience interaction and universally increase interest in ICH?

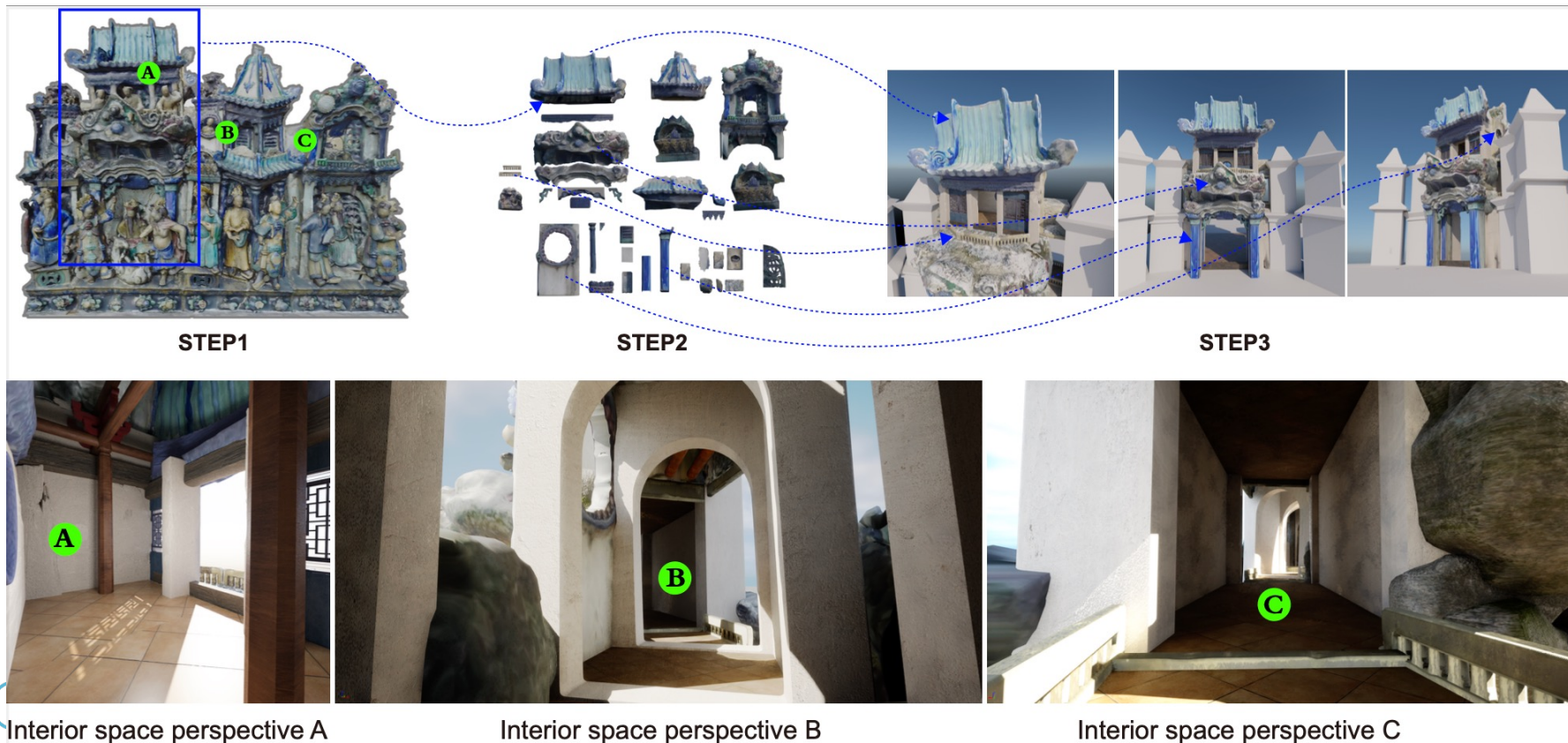
Framework

- Overall framework for a VR contextual storytelling system



Framework

- 3D Reconstruction of Theatrical Scenes
 - Step 1: 3D Model Laser Information Collection.
 - Step 2: Model Segmentation.
 - Step 3: Storytelling Scene Assembly and Reshaping.

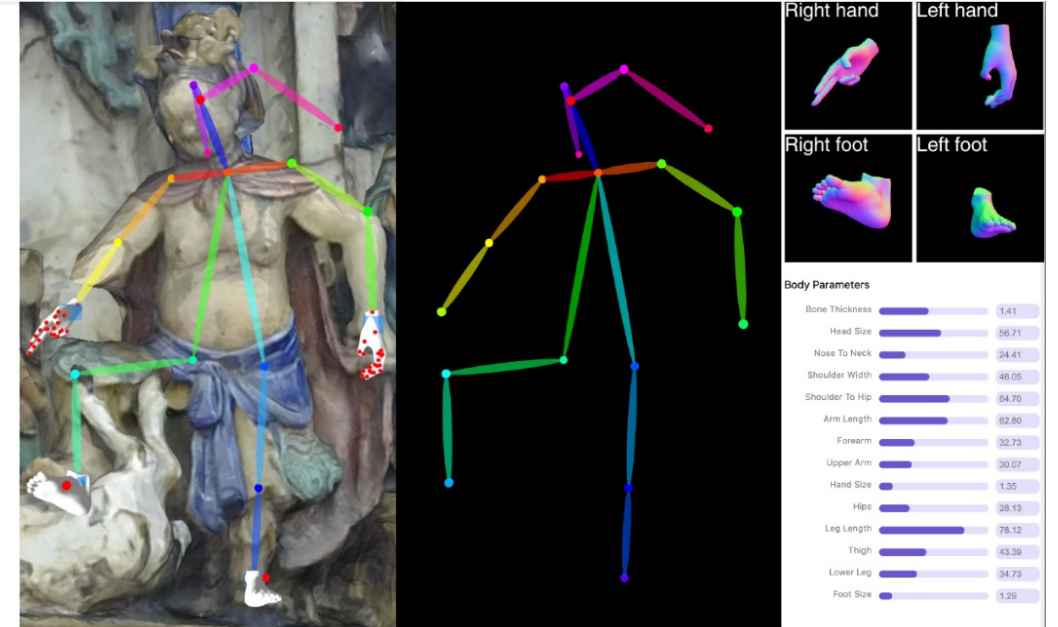


Framework

- Dynamic Restoration of Character Drama
 - OpenPose to analyze and predict the skeleton framework of the “Li” character model



Multi-angle view of the "Li" drone scanning model.

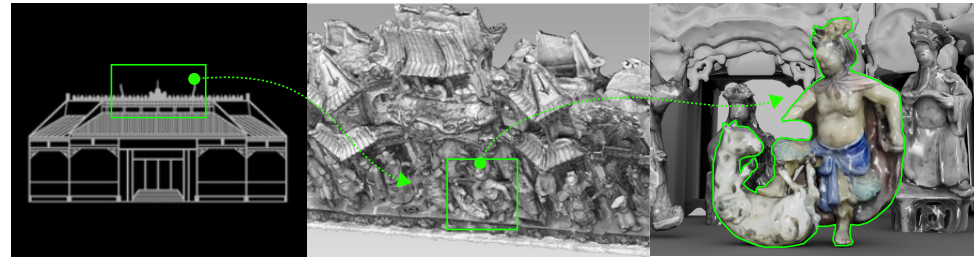


Use AI to predict the skeleton framework of "Li" character model.

Framework

- Dynamic Restoration of Character Drama

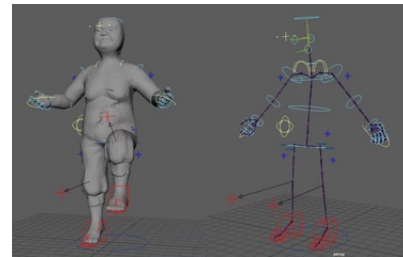
- From static to dynamic:
combination with
CH character
Cantonese opera
motion capture.



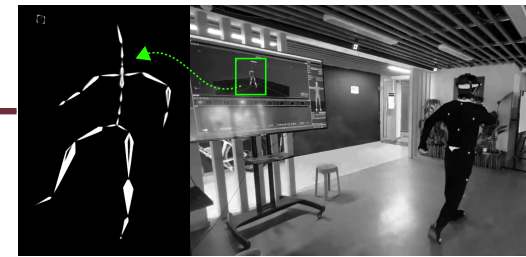
STEP1: Cultural heritage drone scanning model character features capture



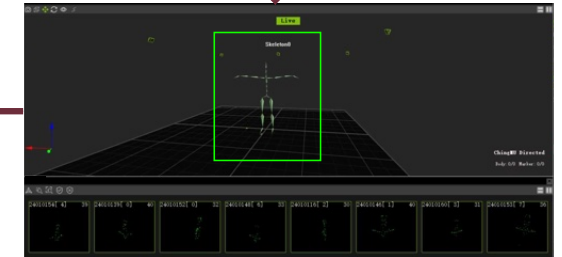
STEP2: Character T post model creation



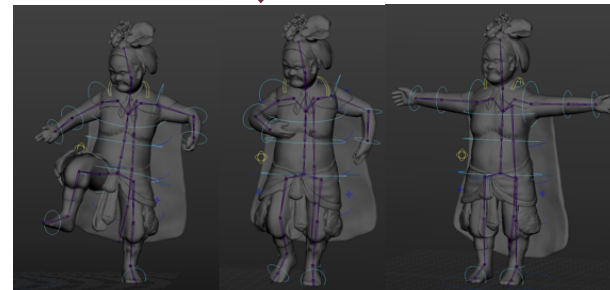
STEP5: Adjust skeleton and controller



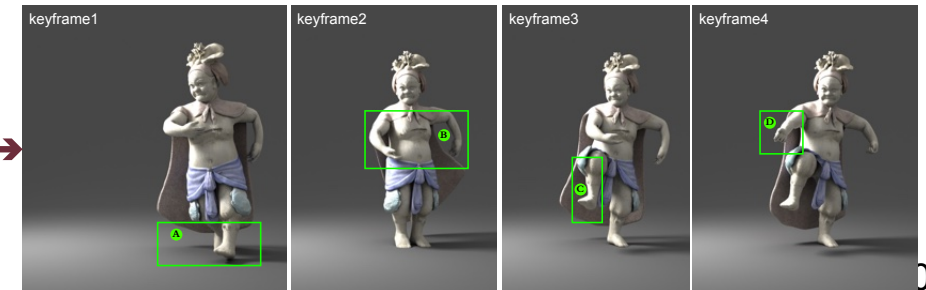
STEP4: Actors use optical motion capture to achieve bone dynamics



STEP3: Create Bone Stick templates, Skin Templates



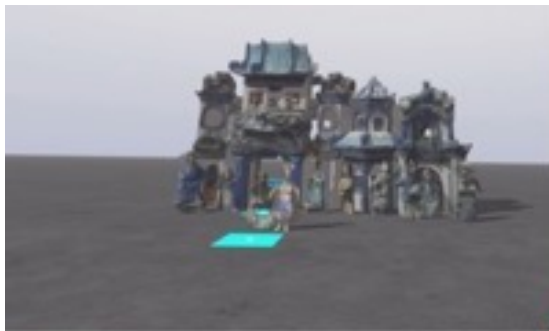
STEP6: Realize the dramatic dynamics of characters in cultural heritage



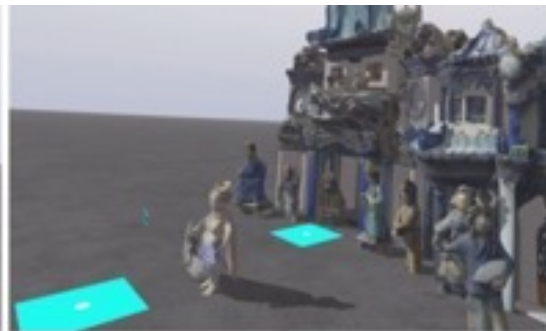
STEP7: Cantonese drama movement: A—Paddock; B—Appearance; C—Stampede; D—Angry finger

Framework

- An interactive VR system
 - integrates motion-captured performances with 3D digitized heritage artifacts



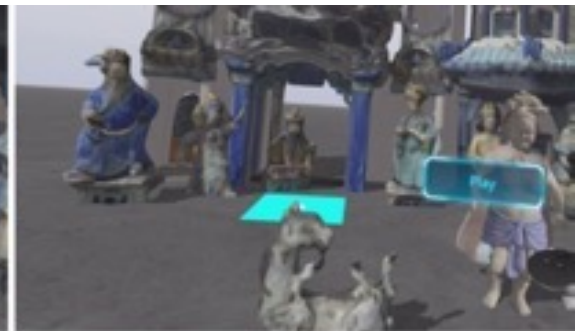
START



STEP1: Walk into the scene



STEP2: Selective interaction



STEP3: Character drama paly



STEP4: 2nd floor view



STEP5: Enter the 2nd floor space scene



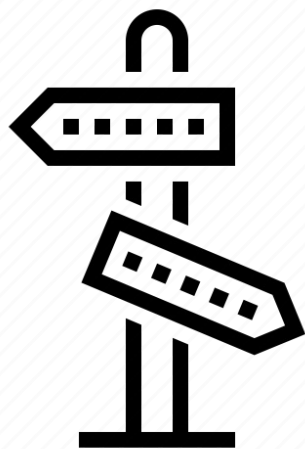
STEP6: 2nd floor interior



STEP7: 2nd floor overlook

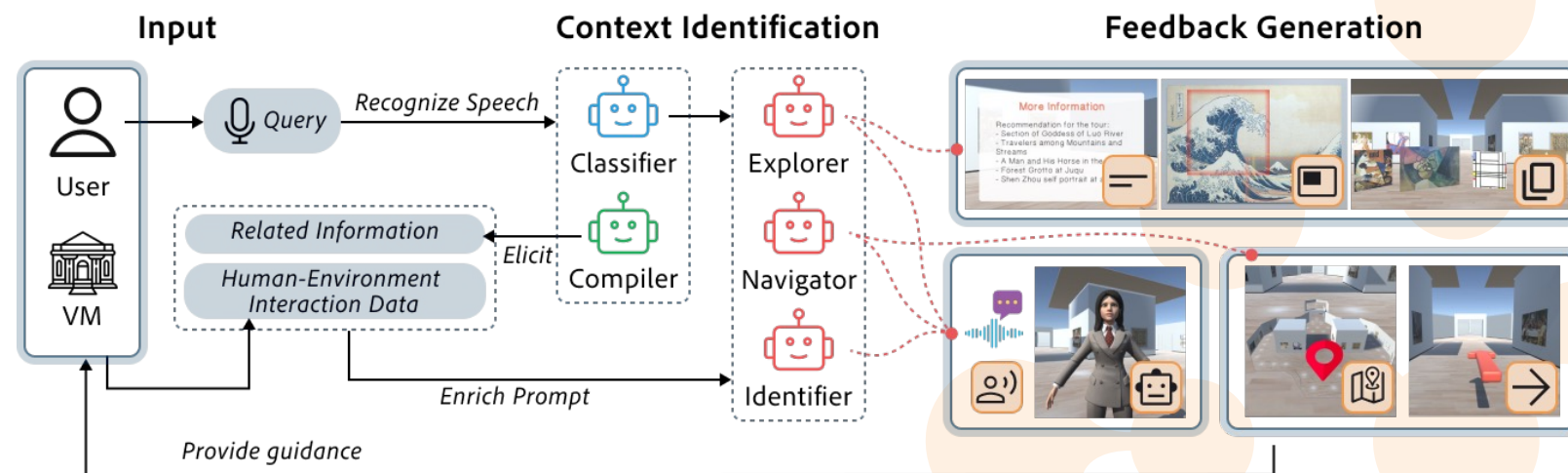
Demo





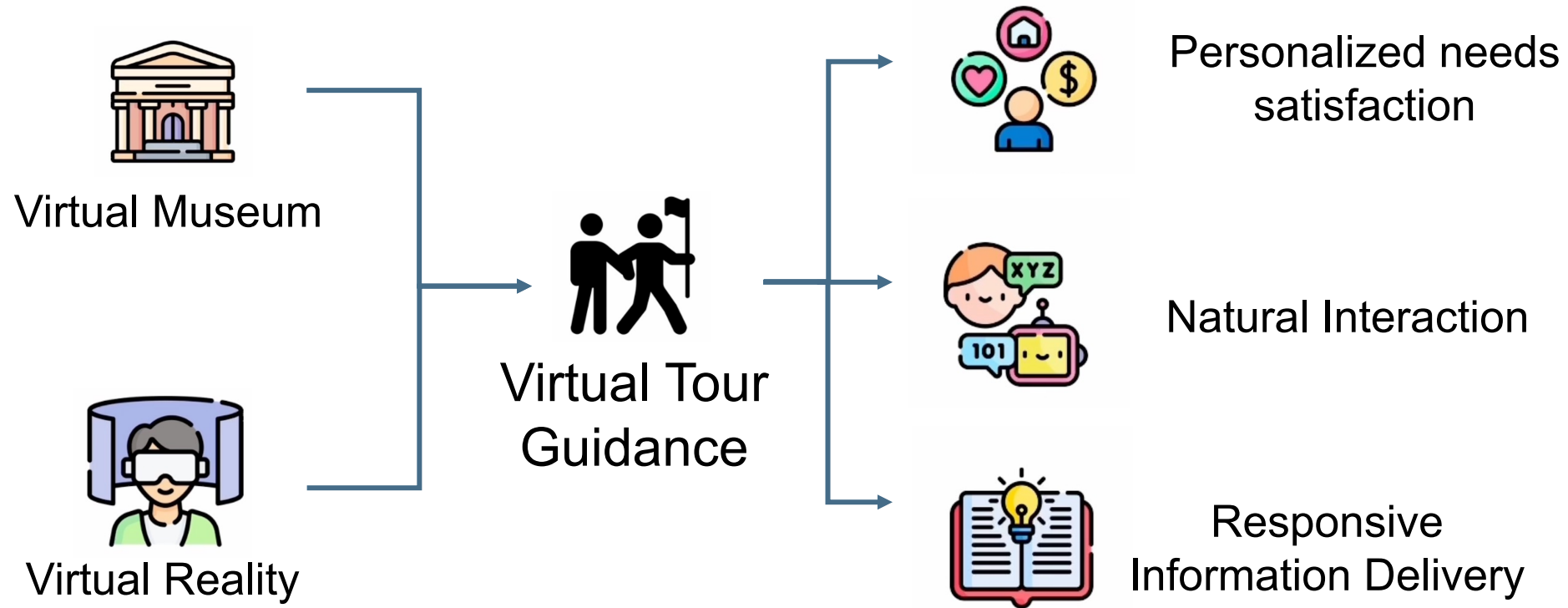
VirtuWander

Enhancing Multi-modal Interaction for
Virtual Tour Guidance through Large Language Models



Background

- Virtual tour guidance can boost the audience visiting experience by offering diverse, flexible, and virtual assistance.



Challenges

Understand user intent

Engage in multi-turn dialogues

Support various downstream tasks



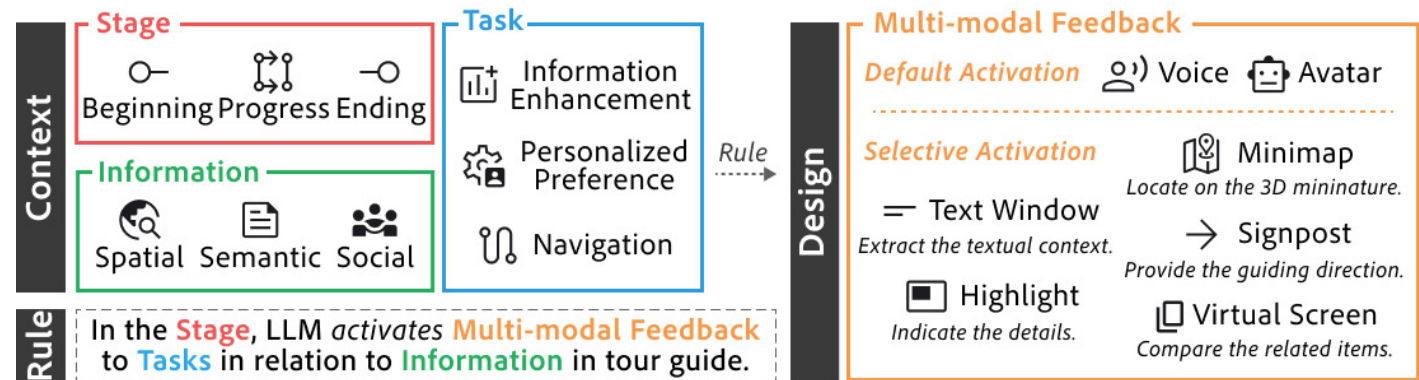
1. Accommodate various guidance-seeking scenarios
2. Translate LLM's natural language outputs into multi-modal feedback

Design Space

Preliminary Study

- 12 participants
- 3 virtual museums
- guidance needs + imagined interactions

Design Framework



Stage: different timings at which visitors seek guidance of their visit.

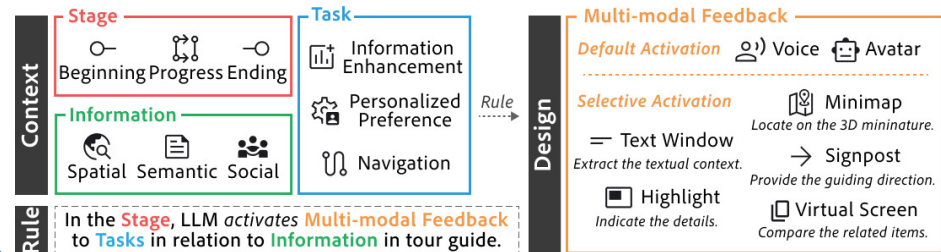
Information: data or insights visitors seek.

Task: what guidance should facilitate visitors to achieve with the above information.

Feedback: LLM's responses to visitors' various guidance-seeking contexts.

Design Space

Design Framework



Five Common Feedback Combinations summarized from our design framework

Input Examples		→ Stage + Information + Task → Feedback			
C1	Please show me Chinese paintings first. I want to see other different paintings in other places.				Voice + Avatar
C2	How many paintings in this museum? Introduce this painting to me. Is there popular paintings I haven't visited?				Voice + Avatar + Text Window
C3	What are the most interesting details in this painting?				Voice + Avatar + Text Window + Highlight
C4	Is there any abstract painting in this museum? Is there any other painting of the similar style? Summarize this tour.				Voice + Avatar + Text Window + Virtual Screen
C5	Guide me to the most popular paintings.				Voice + Avatar + Minimap + Signpost

Multi-modal Feedback Design

- VirtuWander** is an interactive voice-controlled system that enhances LLMs with domain-specific knowledge to improve virtual tour guidance.

Design Mechanism

Input Examples →		Stage + Information + Task →			Feedback
C1	Please show me Chinese paintings first.	○	📄	⚙️	👤🗣️
	I want to see other different paintings in other places.	🔗	🔍📄	⚙️	👤🗣️
C2	How many paintings in this museum?	○	📄	📊	👤🗣️ =
	Introduce this painting to me.	🔗	🔍📄👤	📊	👤🗣️ =
C3	Is there popular paintings I haven't visited?	○	🔍👤	📊	👤🗣️ =
	What are the most interesting details in this painting?	🔗	🔍📄👤	📊	👤🗣️ = 📄
C4	Is there any abstract painting in this museum?	○	📄	📊	👤🗣️ = 📄
	Is there any other painting of the similar style?	🔗	🔍📄👤	📊	👤🗣️ = 📄
C5	Summarize this tour.	○	🔍📄👤	📊	👤🗣️ = 📄
	Guide me to the most popular paintings.	🔗	🔍📄👤	📊	👤🗣️ = 📄 →

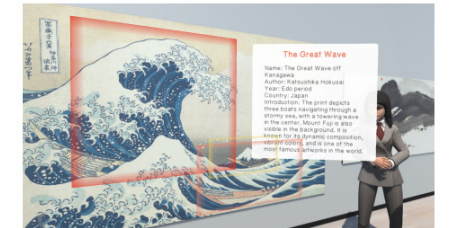
Multi-modal Feedback Design



Voice + Avatar



Voice + Avatar + Text Window



Voice + Avatar + Text Window + Highlight



Voice + Avatar + Text Window + Virtual Screen



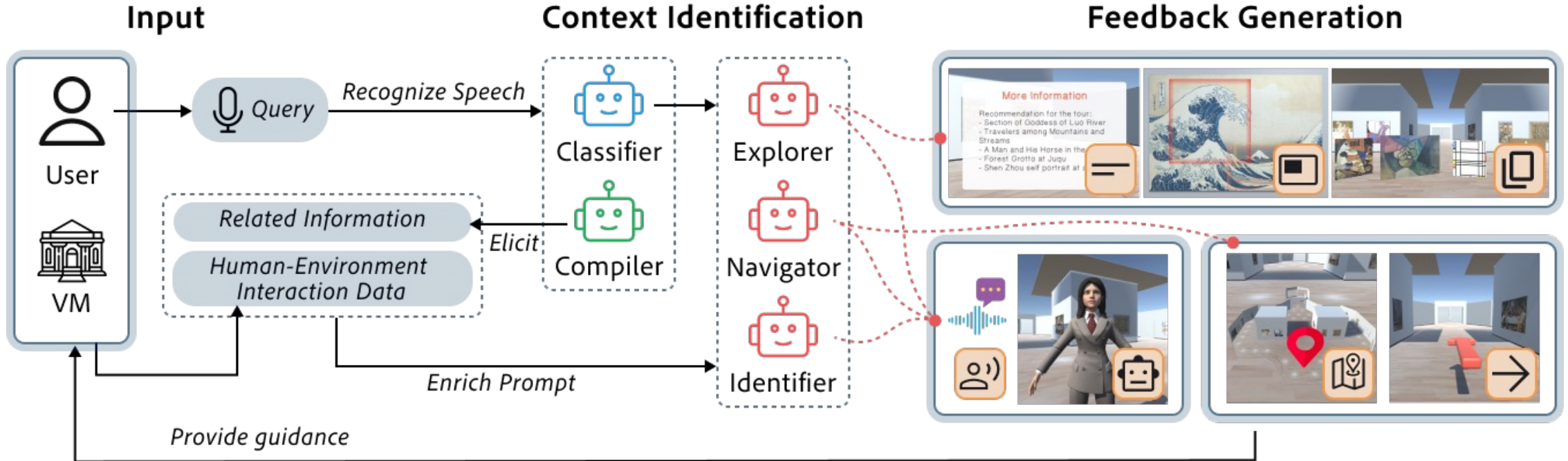
Voice + Avatar + Minimap + Signpost



(b) Virtual Museum

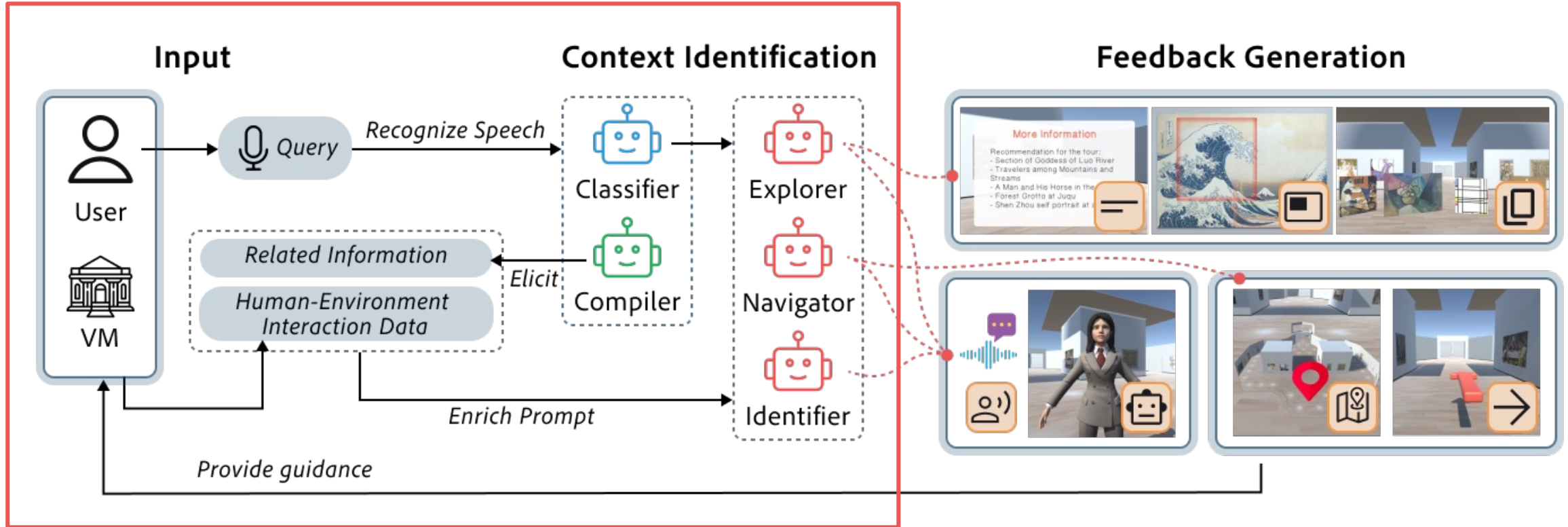
(a) Multi-modal Feedback Combinations

LLM-based Feedback Generation



A Two-stage Framework

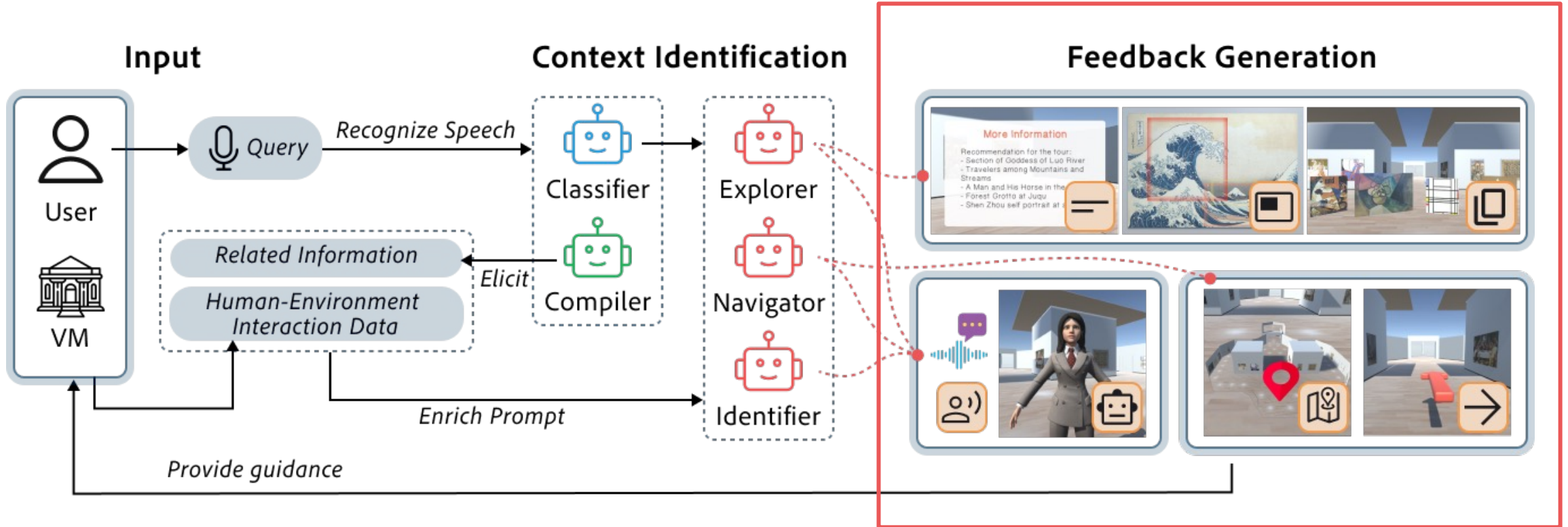
LLM-based Feedback Generation



Stage 1 - Context Identification

Transform user natural language input into guidance-seeking contexts with prompted LLMs.

LLM-based Feedback Generation



Stage 2 – Feedback Generation

Generate multi-modal feedback from LLM formatted responses.

Example Cases

Experience a Thematic Tour

"Please plan a tour in 30 minutes." → "Take me to visit them." → "Summarize the tour and suggest the next painting."

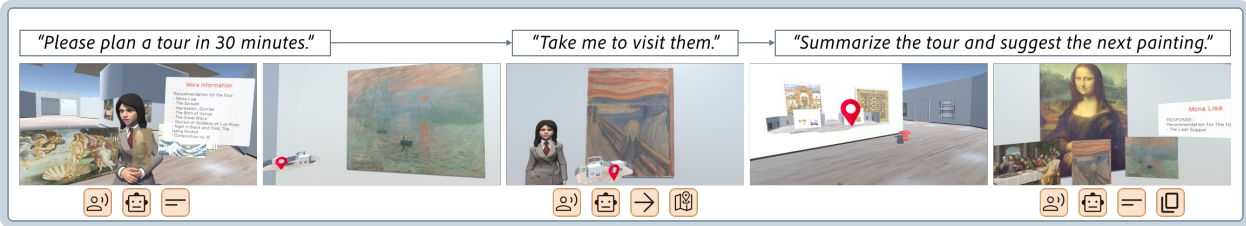
Explore a Single Artwork

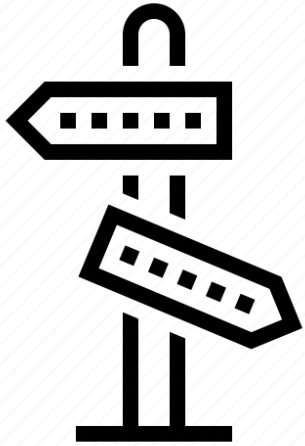
"I want to see The Birth of Venus." → "Introduce this painting." → "What are the interesting details?" → "Who is the people in the middle?" → "Any other paintings of the similar style?"

Customize a Personal Tour

"I really like Chinese paintings." → "Give me some recommendations." → "Introduce the second one." → "Show me this painting." → "Take me to the closet chinese painting."

Example Cases





Conclusion

Richer and more dynamic CH experiences

- VR/AR/XR as new visual mediums
- GenAI for new interaction methods

VR/AR/XR: New Visual Mediums

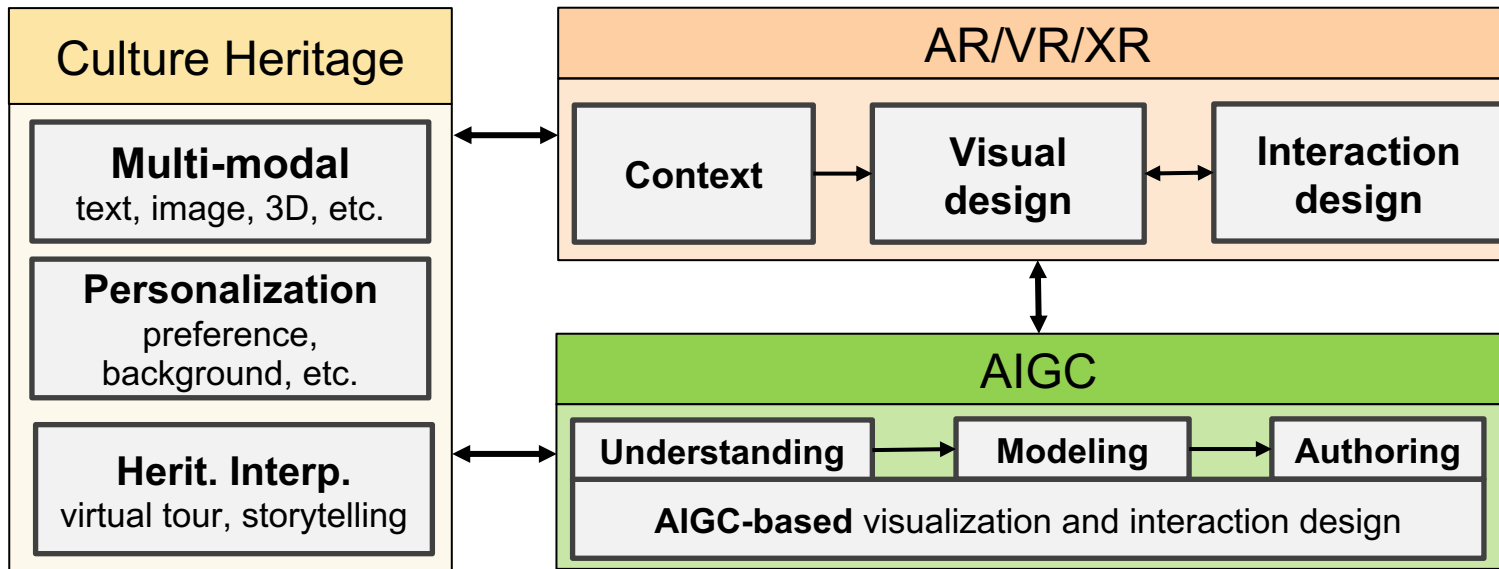
- Transforming Visual Engagement
 - VR: Immersive experiences that transport users to different times and places.
 - AR: Enhancing real-world views with digital overlays.
 - XR: Combining VR and AR for comprehensive experiences.
- **Impact**
 - Greater accessibility to cultural sites
 - Engaging storytelling and visual narratives
 - Enhanced educational tools

Generative AI: New Interaction Methods

- Redefining User Interaction
 - Context-Aware Interactions: Tailoring content based on user behavior and preferences.
 - Natural Language Processing: Facilitating intuitive user queries and responses.
 - Adaptive Learning: Evolving content delivery to match user engagement.
- **Impact**
 - Personalized cultural journeys
 - Interactive and dynamic content
 - Enhanced user engagement and retention

Integrating Technologies for Cultural Heritage

- Combination of VR/AR/XR and Generative AI:
 - Creating immersive and interactive cultural narratives
 - Enabling real-time feedback and content adaptation
 - Fostering deeper connections with cultural heritage





Dr. ZENG Wei

Assistant Professor

The Hong Kong University of Science and
Technology (Guangzhou)

E-mail: weizeng@hkust-gz.edu.cn

Personal Web: zeng-wei.com

Team Web: www.hkust-cival.com