



Artwalk: An AI-Enhanced Urban Art Discovery App

Katharina Klostermeier¹, *Conceptualization*

<https://orcid.org/0009-0000-4757-4253>

Silvia Trovato¹, *Conceptualization*; Mert Kayaburun¹, *Conceptualization*; Hue Nhi Phan¹, *Conceptualization*;

¹ Informatics and Design Bachelors Degree, Hochschule München University of Applied Sciences, 80335 Munich, Germany; mert.kayaburun@hm.edu, katharina.klostermeier@hm.edu, hue-nhi.phan@hm.edu, silvia.trovato@hm.edu

Abstract

Artwalk is a mobile application that enables users to discover and understand urban street art through an AI-powered multi-agent system. This project introduces four collaborative agents— Camera Crew, Info Crew, Audio Crew, and Route Planner—which interact to identify artwork, provide contextual data, generate spoken descriptions, and guide users through curated city routes. Built with React Native for the frontend and FastAPI for the backend, Artwalk integrates local language models (LLaMA3, LLaVA) via Ollama to analyse and interpret user-submitted images. The system sources data from streetartcities.com and supports audio and visual accessibility. This article presents a technical overview of the system architecture and agent functionality, highlighting the app's potential for educational, cultural, and inclusive urban exploration.

Keywords

urban art; AI agents; computer vision; mobile app; route planning; accessibility.

1. Introduction

Street art is an ephemeral yet culturally rich form of expression that often remains unnoticed in everyday urban life. The Artwalk app seeks to bridge this gap by transforming cities into open-air galleries, inviting users to explore public artworks through an AI-assisted mobile experience. By using a decentralised architecture composed of specialised AI agents, Artwalk offers identification, interpretation, and navigation features in real time. This paper outlines the conceptual and technical framework behind the app's development.

2. Referencing Style

Artwalk's user experience is driven by a distributed AI architecture where each "crew" represents an autonomous functional unit. These agents collaborate in sequence to process input from users and generate contextual, multimodal output in return.

2.1. Camera Crew (`classify_image`)

The Camera Crew serves as the initial interface between the user and the Artwalk app. When a user submits a photo of a street artwork, the agent processes it using the visual-language model LLaVA. In the first run, the model attempts to extract key information such as the artist's name, the title of the artwork, and a brief description.

If the initial result contains only a generic description without identifying the artist or artwork, a second LLaVA run is triggered. This follow-up uses the previously generated description as contextual input to improve the accuracy of identification or fill in missing metadata.

In cases where the first run does return names but they are uncertain or incomplete, the second run also serves as a validation step, confirming or refining the initial

output. Once both runs are complete, the aggregated and verified information is passed on to the Info Crew (`detail_agent`) for further enrichment and structuring.

2.2. Info Crew (`detail_agent`)

Building on the Camera Crew's output, the Info Crew agent augments the data with descriptive metadata. This includes interpretations, biographical notes, stylistic analysis, and references to similar artworks. The enriched response is generated through local LLaMA3 models and formatted as a structured JSON object for downstream use.

2.3. Audio Crew (`storyaudio`)

To enhance accessibility and user experience, the Audio Crew transforms the structured output of the Info Crew (`detail_agent`) into an engaging, podcast-style audio narration. This process begins with the JSON file generated by the Detail Agent, which contains metadata such as the artist's name, artwork description, historical context, and stylistic information.

This JSON data is first processed by a dedicated Audio Writer Agent, which composes a coherent and natural-sounding narrative text—similar in tone and flow to a short podcast episode. This narrative is then passed to a text-to-speech (TTS) model, which synthesises the spoken version of the text and outputs it as an MP3 file.

The final audio is served via a REST API and made accessible within the app's interface, allowing users to listen to the artwork's story while navigating or exploring their surroundings.

2.4. Route Planner (`route_planner`)

The Route Planner constructs personalised walking routes using geolocation, user preferences, and artwork metadata. If an artwork lacks stylistic information, a classification agent infers it from the description. The system dynamically generates routes based on proximity, time constraints, and user-selected criteria, transforming exploration into a curated cultural experience

2.5. Map and Save function

Beyond guided routes, the app also features an open map view that displays all known artworks within the database. For simplicity and performance reasons, the current map implementation is limited to artworks located within Munich, based on data from streetartcities.com. Each mapped artwork can be selected for more information or added to a personal collection (Figure 1).

Users can save artworks in two ways:

2.5.1. Scanned photos, which are processed through the AI pipeline and include personalised detail pages.

2.5.2. Map-selected artworks, which come with geolocation and metadata.

All saved items—whether scanned or map-selected—are accessible through a dedicated 'Saved' view within the app. This page allows users to revisit their discoveries, replay audio guides, or plan future tours based on their curated selections.

3. Technical Overview

3.1. System Architecture and Technologies

The frontend was developed using React Native and Expo, ensuring cross-platform compatibility. The backend relies on FastAPI and coordinates all AI agents through Ollama, which manages local instances of LLaVA (image analysis) and LLaMA3 (text generation).

The AI crews are containerised and orchestrated using Docker Compose, while environment configurations are managed via `.env` files across the frontend and backend. Art data is sourced from a JSON File provided by streetartcities.com, a public open-data repository.

Audio outputs are generated using a standalone Text-to-Speech (TTS) service. The entire system operates through modular REST endpoints, allowing scalable interaction among agents.

4. Figures



Figure 1. Description Page of Map Artworks. Source Artwalk App.

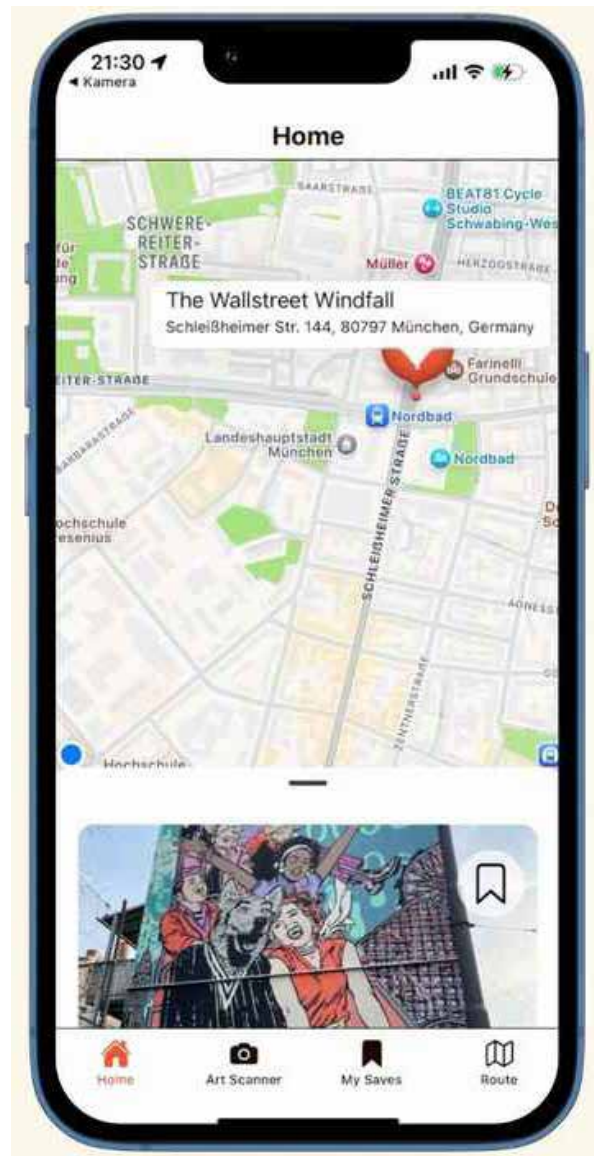


Figure 2. Map Artworks: Artwalk App

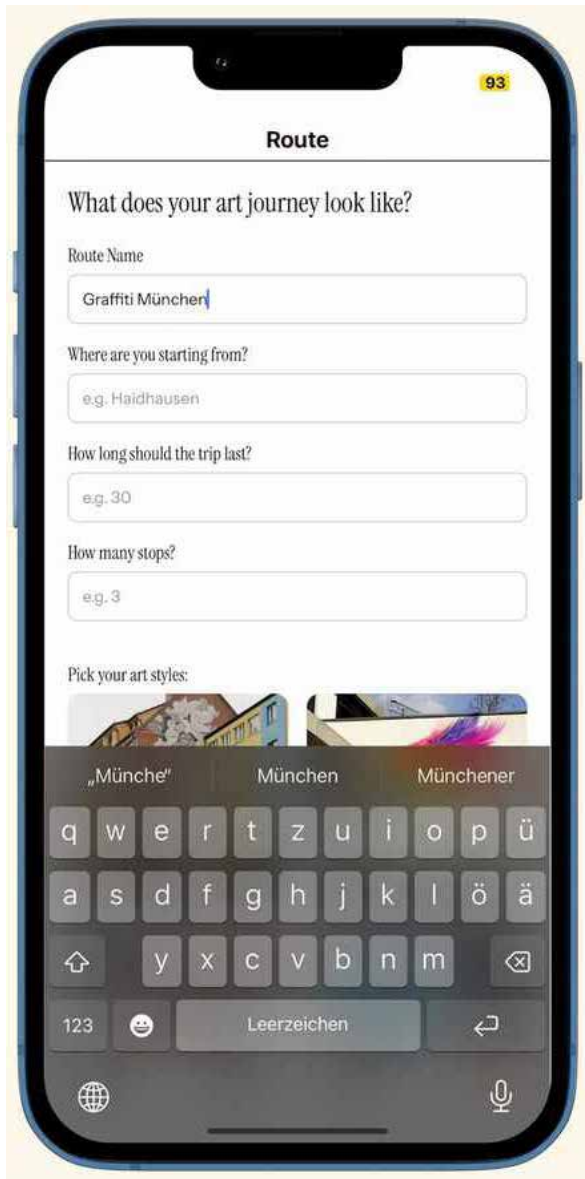


Figure 3: Route Planner input fields

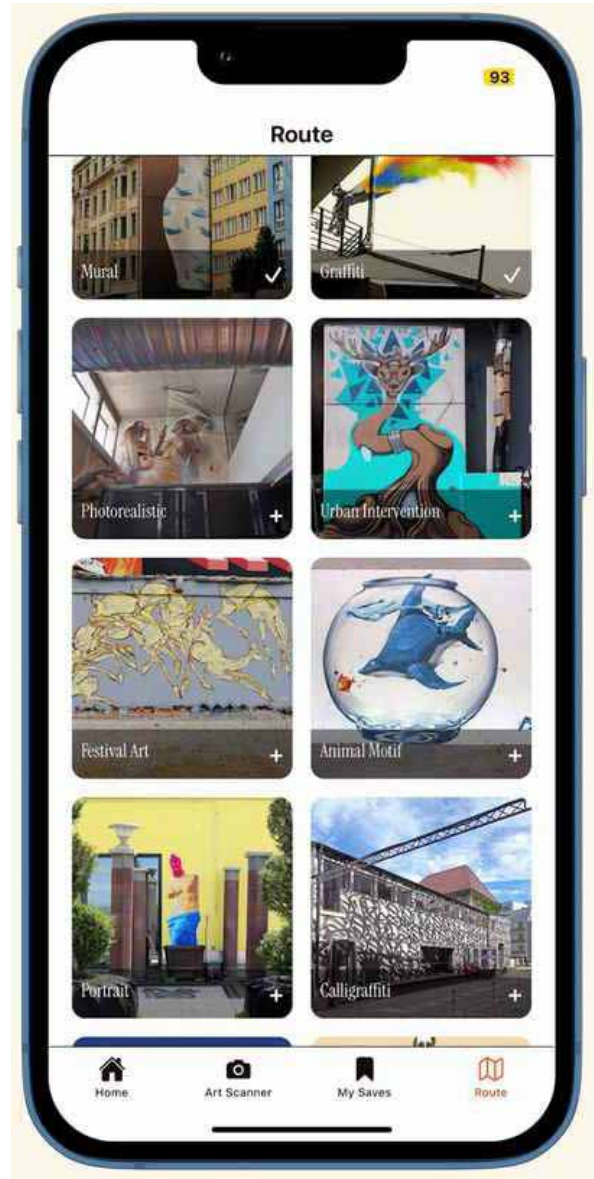


Figure 4. Select the Art Genre for Route Planner. Source: Artwalk App



Figure 5: Overview over generated Route



Figure 6: Camera View of the Image classifier Agent. Source: Artwalk App



Figure 7. Detail Page of the Scanned Artworks. Source: Artwalk

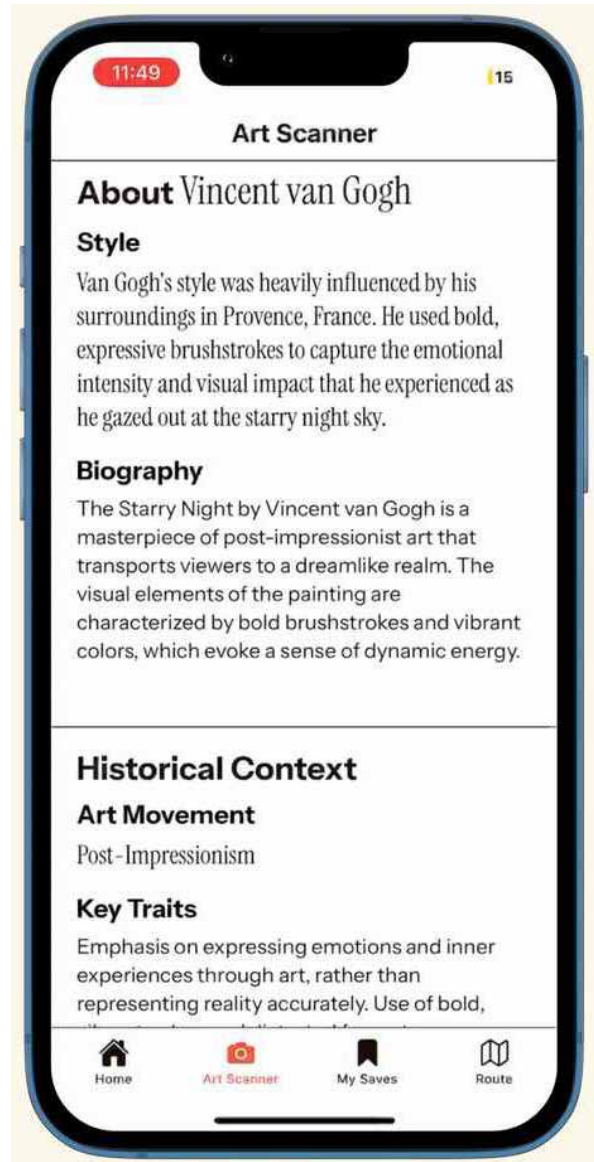


Figure 8. Detail Page of the Scanned Artworks Part 2. Source: Artwalk

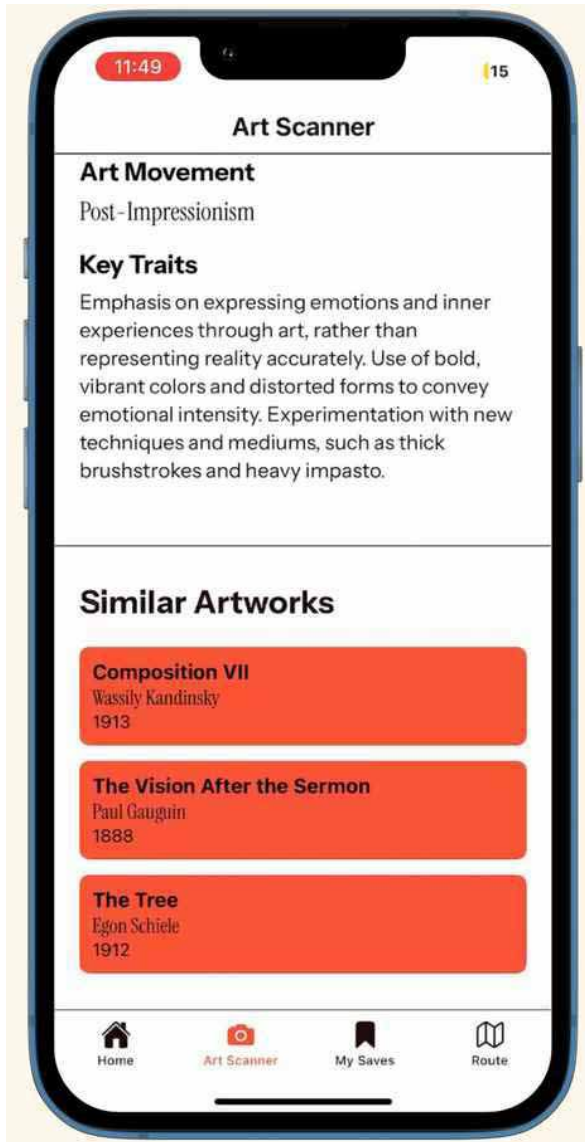


Figure 9. Detail Page of the Scanned Artworks Part 3. Source: Artwalk



Figure 10: Audio Guide Player Page. Source: Artwalk App

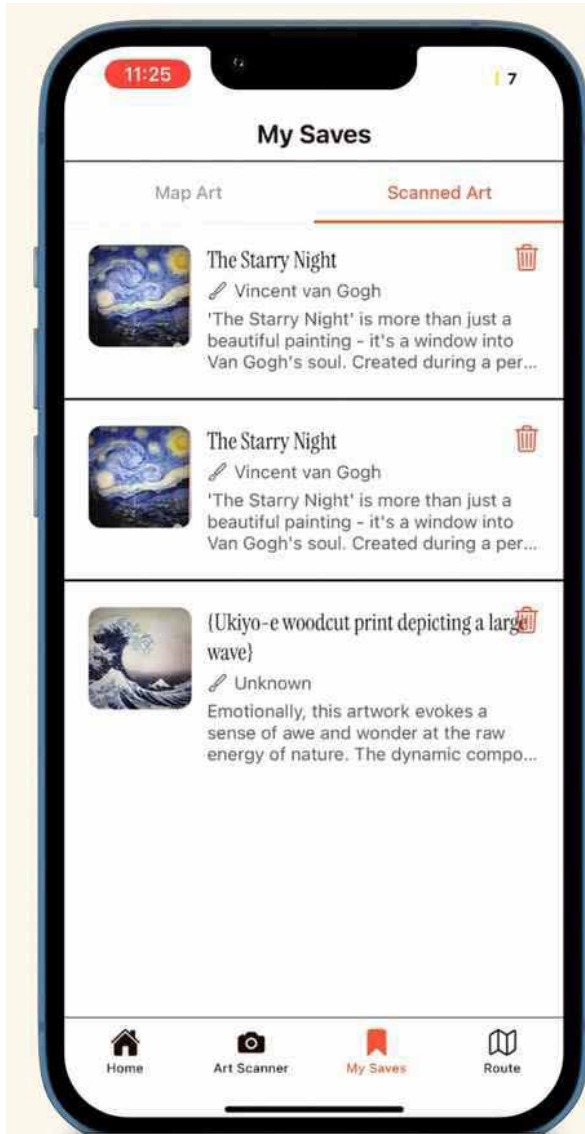


Figure 11. Saved Scanned Artworks. Source: Artwalk App



Figure 12. Saved Map Artworks. Source: Artwalk App

5. Conclusions

Artwalk demonstrates how distributed AI systems can enhance urban exploration by making hidden or overlooked street art accessible and meaningful. Through a collaborative agent architecture, the app identifies artworks, provides interpretive content, and creates audio narratives and curated routes. This approach not only augments the aesthetic experience but also promotes accessibility and inclusivity—particularly for users who benefit from audio guidance.

Technically, the integration of containerised services, local language models, and open data sources has proven effective for building a modular and extensible system. However, the limitations of current AI models, particularly in recognising lesser-known works, indicate room for future improvement. Continued training with local datasets and expanding the app's regional scope are potential directions for future development. Ultimately, Artwalk shows how technology can elevate the everyday cityscape into a dynamic, user-driven cultural experience.

Conflict of Interests and ethics

The author(s) declare no conflict of interests. The author(s) also declare full adherence to all journal research ethics policies, namely involving the participation of human subjects, anonymity and/or consent to publish. No personal data was collected during the development or testing of the Artwalk application.

Acknowledgements

We would like to thank our instructors and peers from the Department of Informatics and Design at Hochschule München for their continuous support and critical feedback throughout the development of this project. Special thanks go to the creators and maintainers of streetartcities.com for providing the open data foundation on which our application builds. We also acknowledge the collaborative input from the testing groups Cricis Companion and MaSaPaLou, whose suggestions significantly improved our app's usability and accessibility.

References

- Ollama. (2025). Ollama: Local AI runtime for models. Retrieved from <https://ollama.com>
- Street Art Cities. (2025). Open data platform. Retrieved from <https://streetartcities.com>

