LLMs for Accessibility

SVILUPPO DI APPLICAZIONI PER

DISPOSITIVI MOBILI - AA 2023/2024

Matteo Manzoni – 33598A

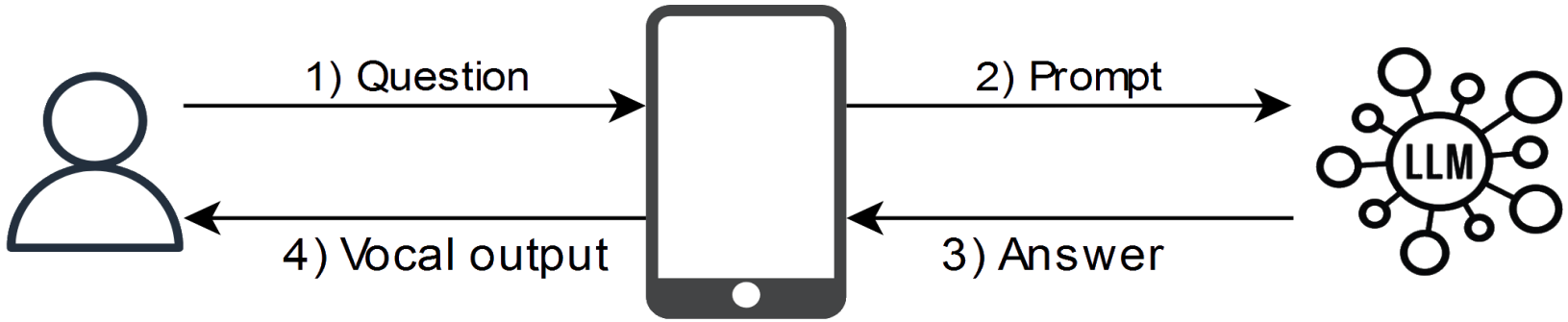
matteo.manzoni2@studenti.unimi.it

Project goal

Prototype an application to test if LLMs can be used as accessibility tools for visually impaired people in augmented reality mobile applications.

Users should be able to:

- Get a description of both the real and virtual environment
- Ask specific questions about the environment.



LLM

The app was tested with two LLMs: DescriPix and GPT-4o. Both models were able to describe the input image, but it was chosen GPT-4o because it is much more responsive to changes in the text prompt, allowing the user to ask specific questions about the environment.

Example:

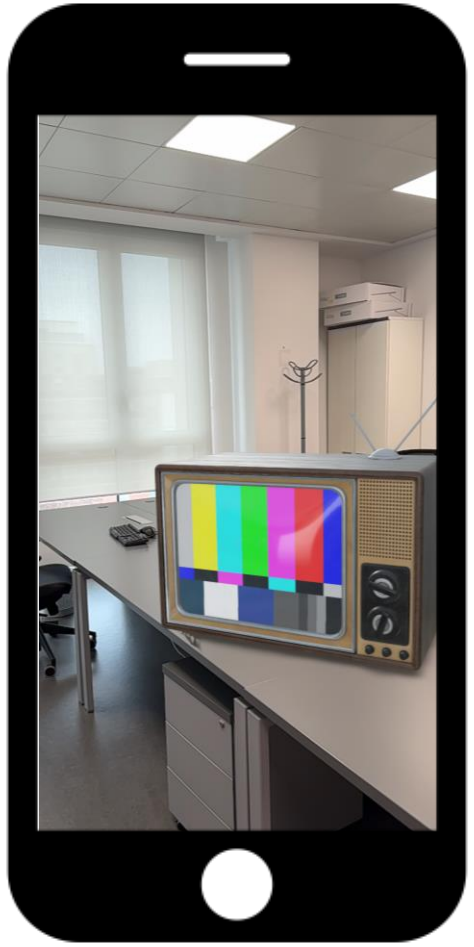
This is a picture of the environment around me taken from an AR application. The image contains real objects and virtual ones. I cannot see them; what's in the background? Give me a description in the form of an alt text for the image.

- DescriPix:

"The image contains real objects and virtual ones. The background includes a window and a trash can."

- GPT-4o:

"The background of the image shows an office environment. There are tall windows with white blinds partially drawn, letting in natural light. To the right, there is a white cabinet with some boxes on top. Next to the cabinet is a coat rack with no items hanging on it. The office has a clean and organized appearance."



Scene image acquisition

- Real world:

Each frame, ARKit provides the image acquired from the device camera, which is used to generate the AR frame.

- Entire scene:

The ARView class provides a method called “snapshot” to directly capture a screenshot of the entire scene.

- Virtual scene:

While it’s not possible to directly capture an image of the virtual elements of a scene without the real-world elements, ARKit allows you to hide the background. This way, you can capture an image of the entire scene with the background hidden, and then immediately restore the background.

Environment description

Real world

A modern office space with a minimalist design. There are two large gray desks arranged in an L shape, each with computers and office chairs. Two of the chairs are placed at the desks, while one is slightly pulled away. A small storage unit is located underneath one of the desks. A large window with sheer blinds is positioned on the left wall, providing natural light. On the right side, a white wardrobe stands against the wall...

Entire scene

Virtual elements

A 3D-rendered vintage television set with a wooden frame, a screen displaying colored vertical bars in a test pattern, and dials on the lower right side. An antenna extends from the top.

The environment is a modern office space. There are large windows with light-colored blinds, allowing natural light to filter in. A set of desks is arranged in the middle of the room, equipped with office chairs. On the desks, a keyboard and a mouse are visible. [...] In the foreground, a virtual antique television with a test pattern on the screen is superimposed on one of the desks, contrasting with the contemporary office setting...

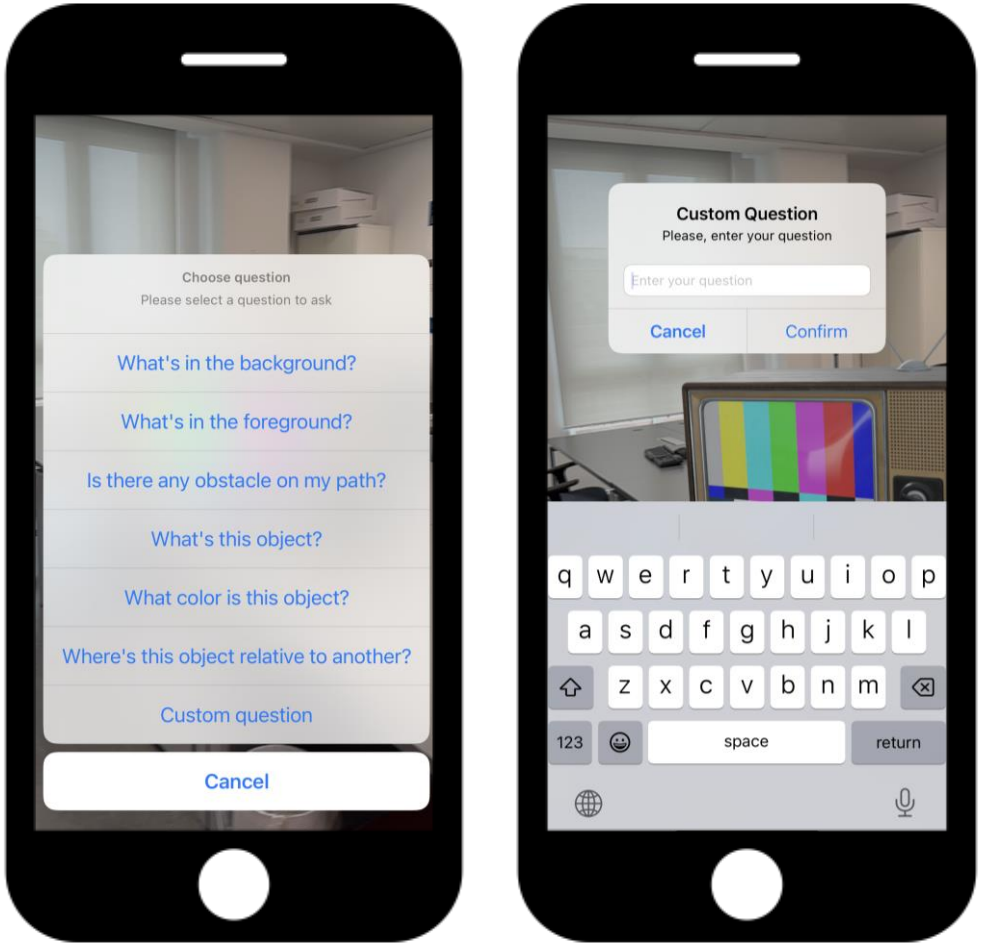
Specific questions

- Some questions are predefined and included in the application, but users can also type any question they wish to ask.
- Certain questions may require selecting one or more objects. Both real and virtual objects can be selected with a single tap.

Example:

What color is this object? (selecting the TV monitor)

Multicolored: yellow, cyan, green, magenta, red, blue, white, black, and shades of grey.



Future extensions

- Reducing response time
- Integrating a speech-to-text system for custom questions
- Testing the system's effectiveness with users
- Distributing it as an AR accessibility library

- Demo:

https://youtu.be/qEaWN3SsycM

- Source code:

https://github.com/Matteo-3033/Mobidev