

Apple is working on an AI agent that can describe Street View scenes to blind people

Apple engineers have detailed an AI agent that can accurately describe Street View scenes. If the research is successful, it could become a tool to help visually impaired people explore a virtual location in advance.

Apple engineers have detailed an AI agent that can accurately describe Street View scenes. If successful, it could become a tool for blind people to explore a virtual location in advance. Blind and visually impaired people already have tools to navigate their devices and local environments.

But Apple believes that it could be beneficial for those people to know about the physical features of a place before they visit. A paper published through Apple Machine Learning Research on Monday talks about SceneScout, an AI agent powered by a large multimodal language model .

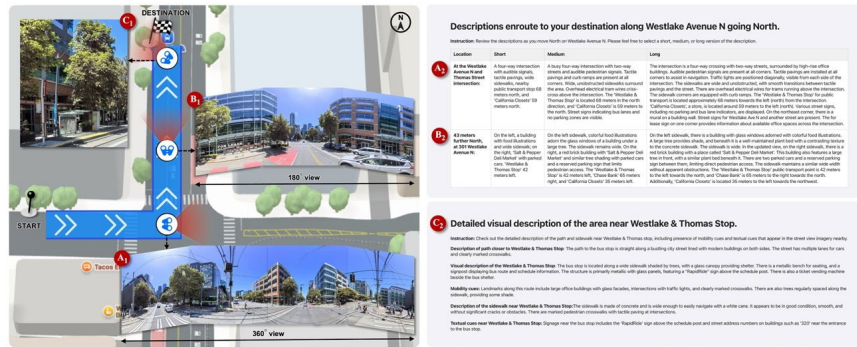
The key point of the agent is that it can be used to look at Street View imagery, analyze what it sees, and describe it to a human viewer. The paper was written by Apple's Leah Findlater and Cole Gleason, as well as Columbia University's Gaurav Jain. It explains that people with low vision may be reluctant to travel alone in unfamiliar environments, as they don't know in advance what physical landscape they'll encounter.

There are tools available to describe the local environment, such as Microsoft's Soundscape app from 2018. However, they are all designed to work on-site, not in advance. Currently, pre-travel advice provides details like landmarks and turn-by-turn navigation, which doesn't provide much landscape context for visually impaired users.

However, Street View-style imagery, such as Apple Maps Look Around, often provides sighted users with more contextual clues that are often missed by those who can't see.

SceneScout

This is where SceneScout comes in, as an AI agent to provide accessible interactions using Street View imagery. Scene Scout has two modes, with **Route Preview** providing detailed information about the elements it can observe on a route. For example, it can inform the user about trees at turns and other tactile elements.



The second mode, **Virtual Exploration**, is described as allowing free movement within Street View imagery, describing elements to the user as they move virtually. In user research, the team determined that SceneScout is useful for people with visual impairments, in terms of discovering information they would not be able to access using existing methods.

When it comes to descriptions, it is considered mostly accurate, at 72% of the time, and can describe consistent image elements 95% of the time. However, there are sometimes 'subtle and logical errors' that make descriptions difficult to verify without using sight.

When it comes to how to improve the system, testers suggested that SceneScout could provide personalized descriptions that could adapt across sessions. For example, the system could pick up on the types of information users like to hear. Changing the perspective for descriptions from a camera on a car roof to where pedestrians typically are could also help improve the information.

Another way to improve the system could also be done on the spot. Participants said they would like to see Street View descriptions provided in real time, to match where they were walking. Participants said this could be an app that provides visual information through bone conduction headsets or Transparency mode as they move.

Furthermore, users may want to use a combination of a gyroscope and compass in one device to point in a general direction for environmental details, rather than hoping they align the camera properly for computer vision.

Future applications

Like a patent filing, a paper detailing new uses for AI doesn't guarantee it will appear in a future product or service. But it does provide a glimpse into the applications Apple has considered for the technology.

While it doesn't use Street View imagery, a similar approach could be used in a number of rumored Apple products. Apple is said to be working on AirPods with built-in cameras, as well as Apple Glass smart glasses with their own cameras.

In both cases, the camera could provide Apple Intelligence with a view of the world, which would then be used to help answer user queries. It's not hard to imagine a similar system being used to describe the local environment to users, all using live data rather than potentially outdated Street View imagery.

You finished reading the article "**Apple is working on an AI agent that can describe Street View scenes to blind people**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us

regularly.
