

StreetReaderAI

Making Street View Accessible via
Context-Aware, Multimodal AI



**JON E.
FROEHLICH**
Google
Research



**ALEX
FIANNACA**
Google
DeepMind



**NIMER
JABER**
Google

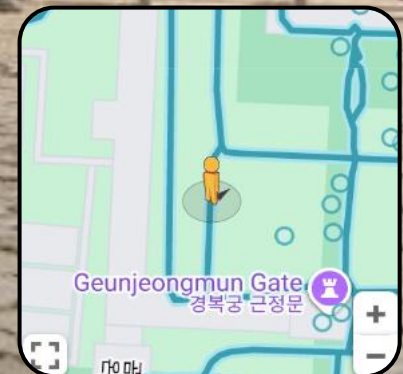


**VICTOR
TSARAN**
Google



**SHAUN
KANE**
Google
Research

You're standing in
Gyeongbokgung Palace,
facing northwest with a clear
view of a majestic two-tiered
palace building adorned with
intricate, colorful patterns...



Google Street View

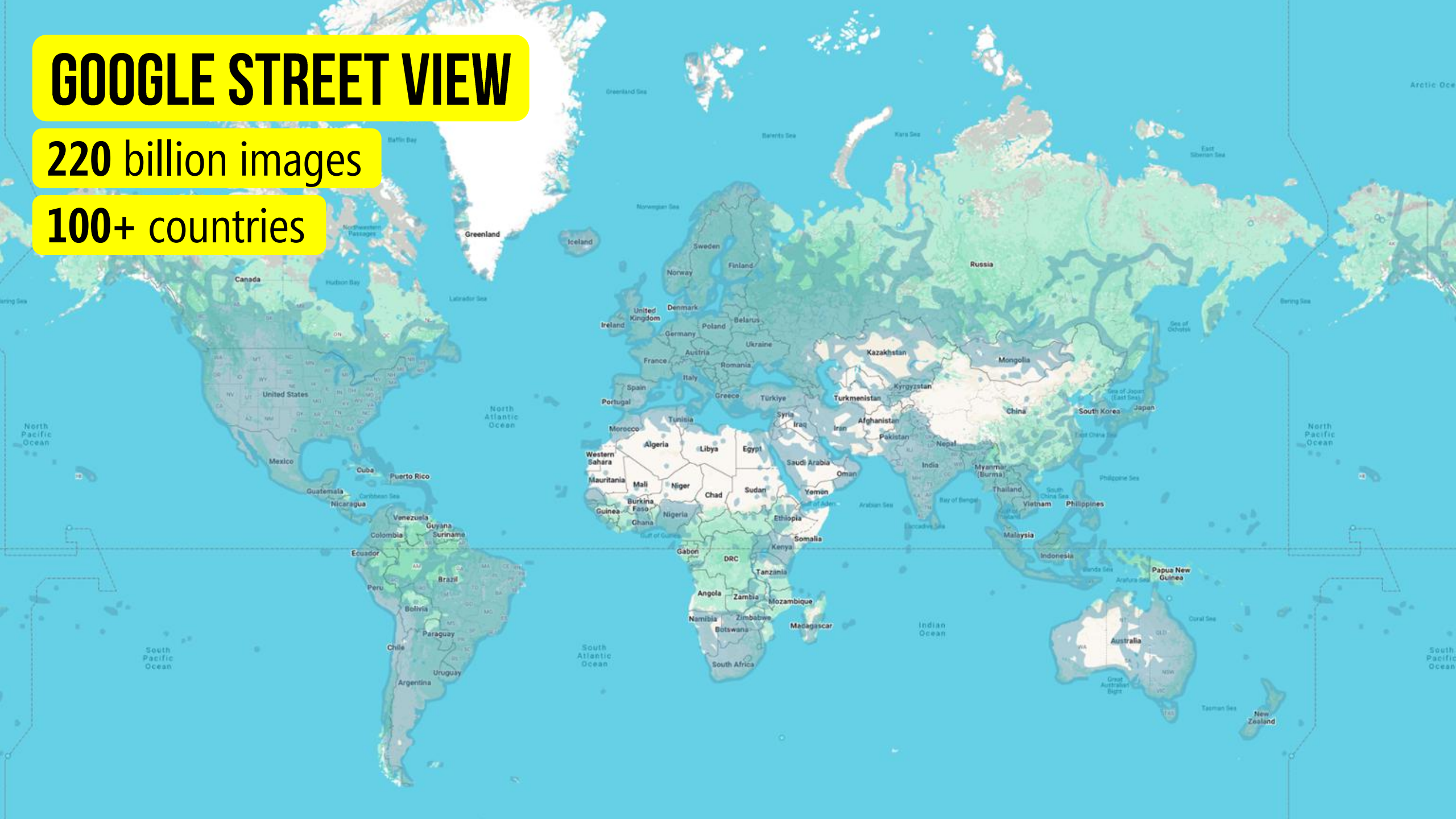
TRANSFORMED HOW PEOPLE **LEARN** ABOUT THE **WORLD**



GOOGLE STREET VIEW

220 billion images

100+ countries



Fundamentally **Inaccessible**



GOOGLE STREET VIEW

Current Screen Reader Experience

←

2411 7th Ave
New York

Google Street View

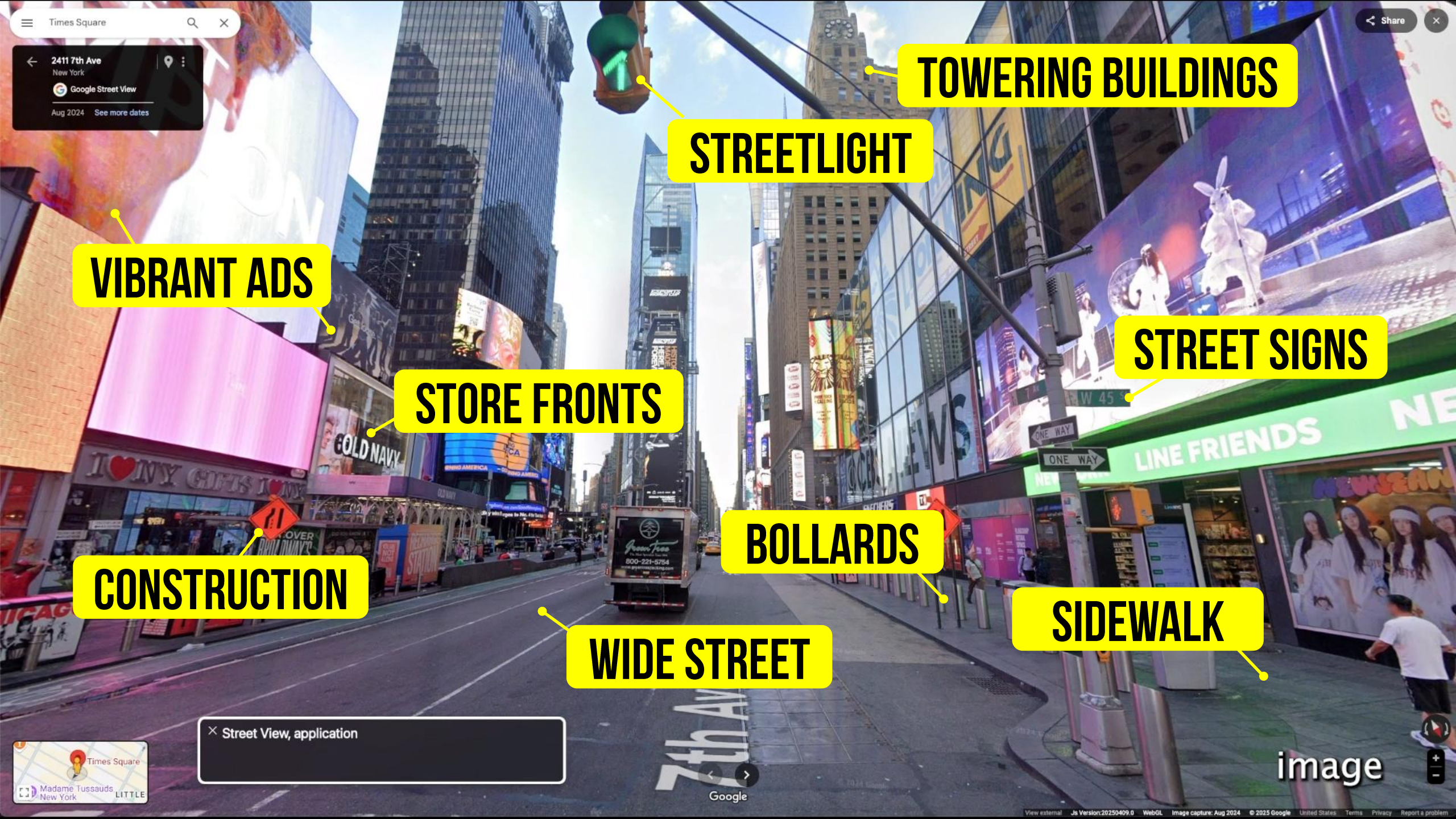
Aug 2024 See more dates

× Street View, application

Times Square

Madame Tussauds
New York

LITTLE



← 2411 7th Ave
New York
Google Street View
Aug 2024 See more dates

VIBRANT ADS

TOWERING BUILDINGS

STREETLIGHT

STORE FRONTS

STREET SIGNS

CONSTRUCTION

BOLLARDS

WIDE STREET

SIDEWALK

Times Square
Madame Tussauds
New York
LITTLE

× Street View, application

image

HOW CAN WE MAKE **STREETSCAPE** **TOOLS** ACCESSIBLE TO BLIND USERS?

× Street View, application

image

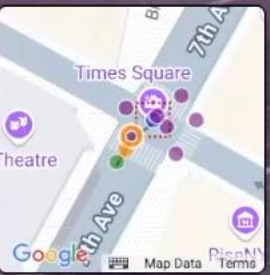


Let's try this
again but with
StreetReaderAI

Search for a location to go to...

Nearby Virtual Tours

Ask a question about this location...



AI Description
The image shows a street scene in Times Square, New York City, with a wide avenue stretching into the distance, lined with tall buildings covered in bright billboards and advertisements. In front of you is a crosswalk with white stripes, and a delivery truck is driving away from you in the center of the road. To the right is a sidewalk with metal bollards. At the corner to the right is a sign that reads "W 45 St, One Way".

StreetReaderAI is not just a context-aware description tool for streetscape images

StreetReaderAI is an accessible, immersive, and interactive street view map experience

Search for a location to go to...



Nearby



Virtual Tours

Generate Description

what's the nearest bus stop?

The nearest bus stop is just to your right, across the street. It's a bit further down the block, but it's the only one visible in your current view.

Listening...

End Conversation

EXAMPLE

Where's the bus stop?



Search for a location to go to...



Nearby



Virtual Tours



Generate Description

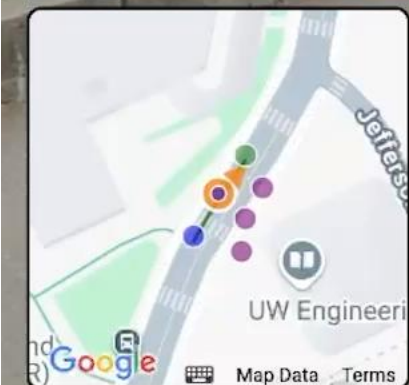
what's the nearest bus stop?

The nearest bus stop is just to your right, across the street. It's a bit further down the block, but it's the only one visible in your current view.

Listening...

End Conversation

Where's the bus stop?



Search for a location to go to...



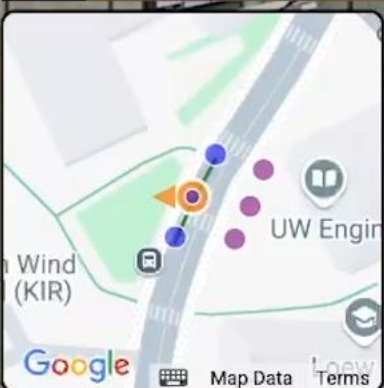
Nearby

Virtual Tours

Fluidly interacting in street view

Engaging in context-aware chat

Learning critical geo-info latent in images



Oh, covered. Beautiful. So no rain.



Generate Description

only one visible in your current view.

Where's the bus stop?

The bus stop is behind you. It's on the other side of the street.

Okay.

Great. Let me know if you have any other questions.

turn me to towards the bus station

Done

what's ahead of me.

The bus stop is directly ahead of you. There are benches and a covered area.

are there any shelters in this stop?

Yes, the bus stop has a covered shelter with benches.

Ask a question about this location...

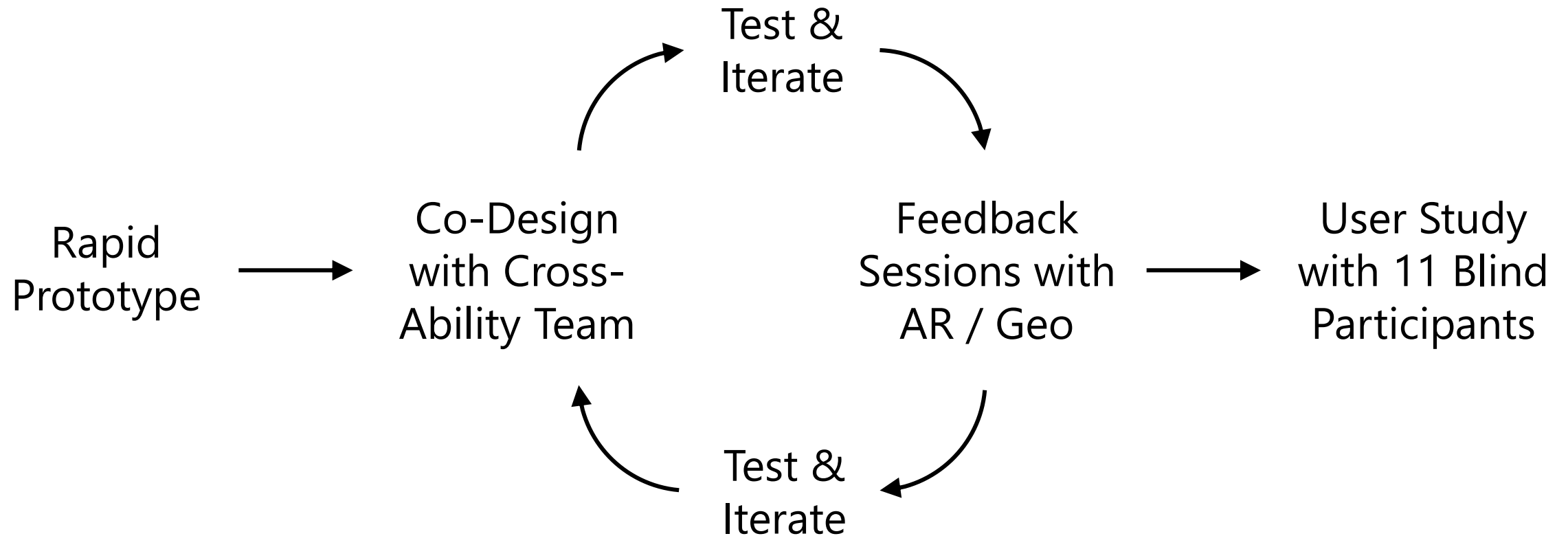


End Conversation

A background image of the Eiffel Tower in Paris, France, with a dark blue sky and green trees in the foreground. The tower is centered and its base is visible. The image is slightly dimmed to make the text stand out.

DESIGNING STREETREADERAI

ITERATIVE DESIGN PROCESS

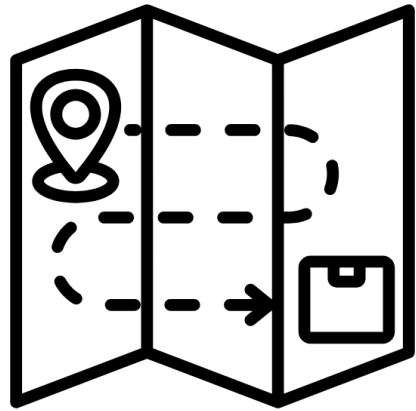


FOUR CORE USER TASKS



POI Investigations

What does a place look like?



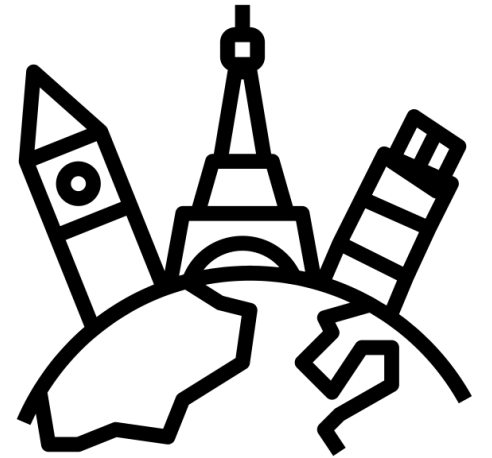
Route Planning

What does my route look like?



Open-Ended Exploring

I just want to explore...



Virtual Tourism

I've always wanted to visit...



Search for a location to go to...

Virtual Tours

Generate Description

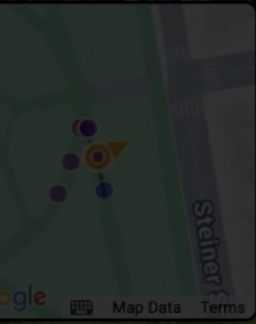
Ask a question about this location...

Are there curb cuts at the intersection in front of me?

How many lanes of traffic are there on Hayes Street?

Are there any tactile markings to indicate the crosswalk?

SYSTEM OVERVIEW



Now Facing
Northeast.

Search for a location to go to...



Nearby



Virtual Tours

LOCATION SEARCH

AI DESCRIBER

Generate Description

AI CHAT

Ask a question about this location...



Are there curb cuts at the intersection in front of me?

How many lanes of traffic are there on Hayes Street?

Are there any tactile markings to indicate the crosswalk?

MINI-MAP

STATUS MESSAGES

Now Facing
Northeast.

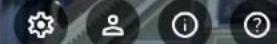


Search for a location to go to...



Nearby

Virtual Tours



AI DESCRIBER

Combines **GIS data** +
image analysis

Generate Description

Ask a question about this location...



Are there curb cuts at the intersection in front of me?

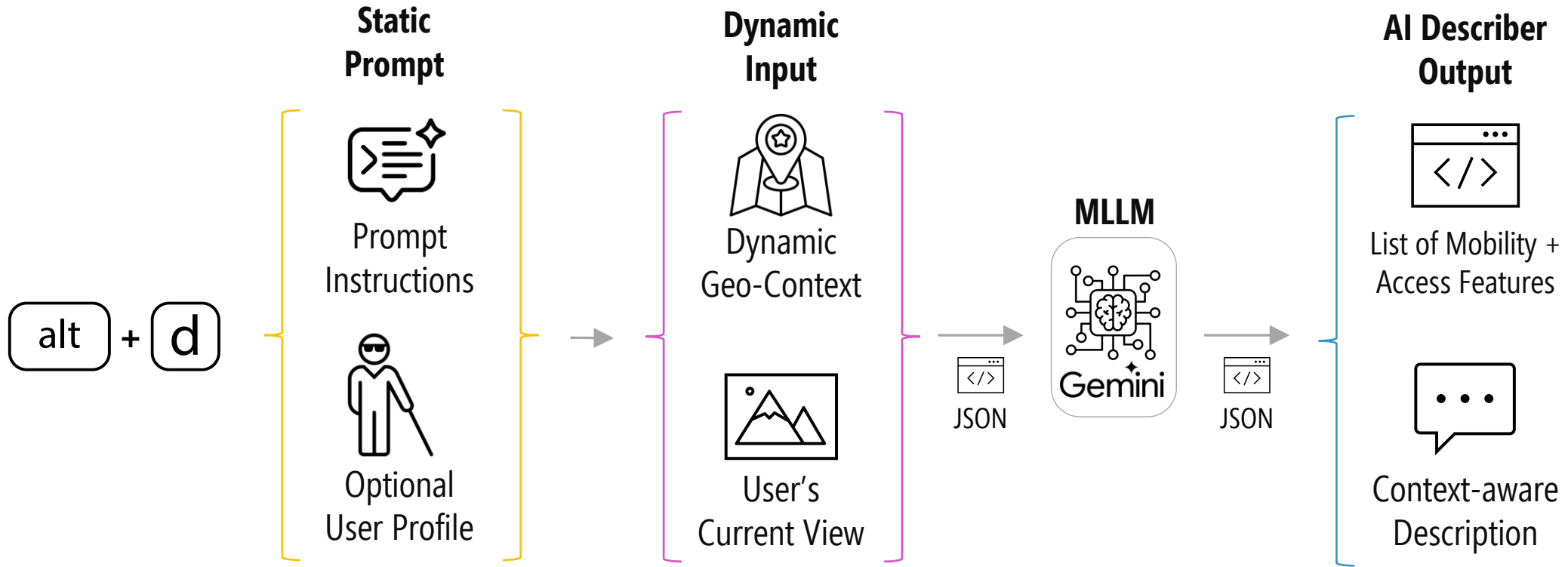
How many lanes of traffic are there on Hayes Street?

Are there any tactile markings to indicate the crosswalk?

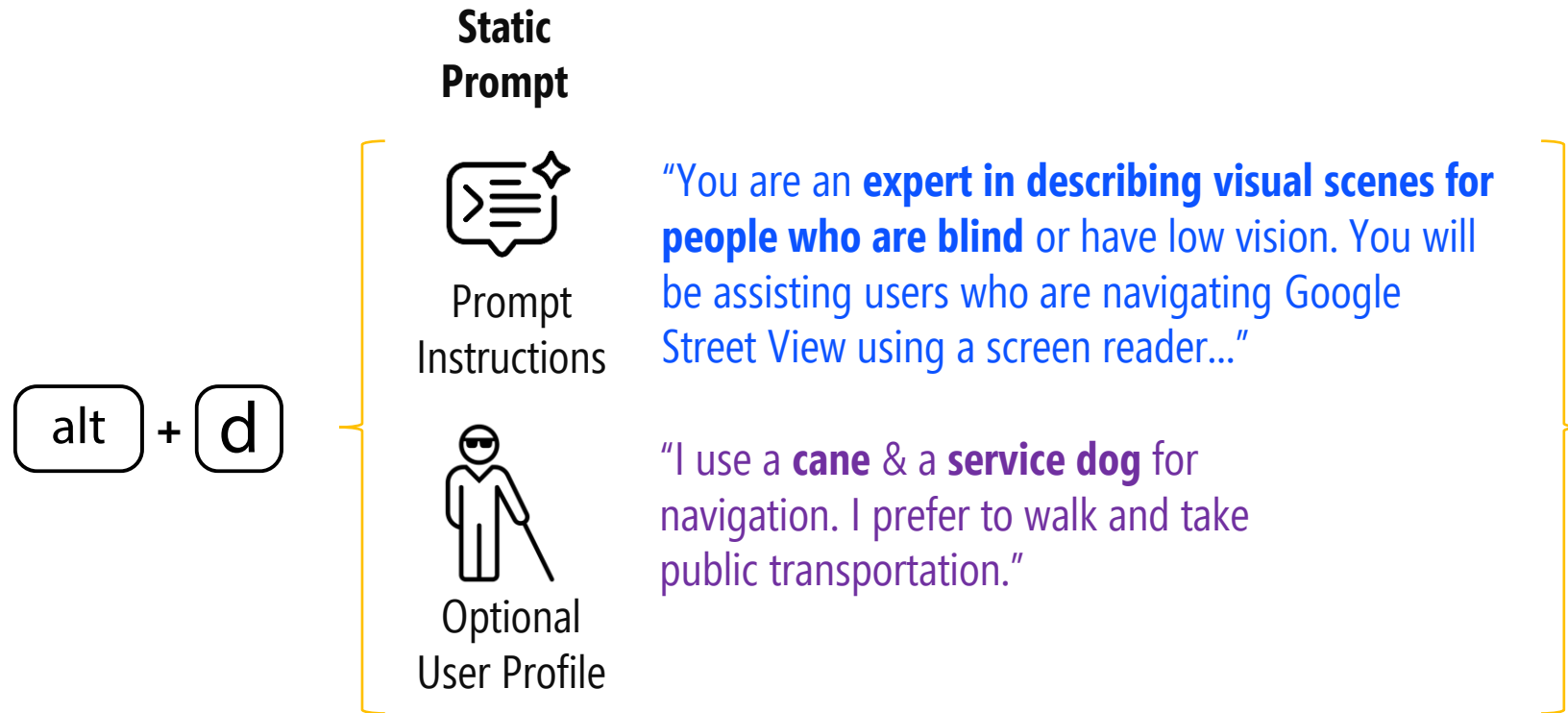


Google Map Data Terms

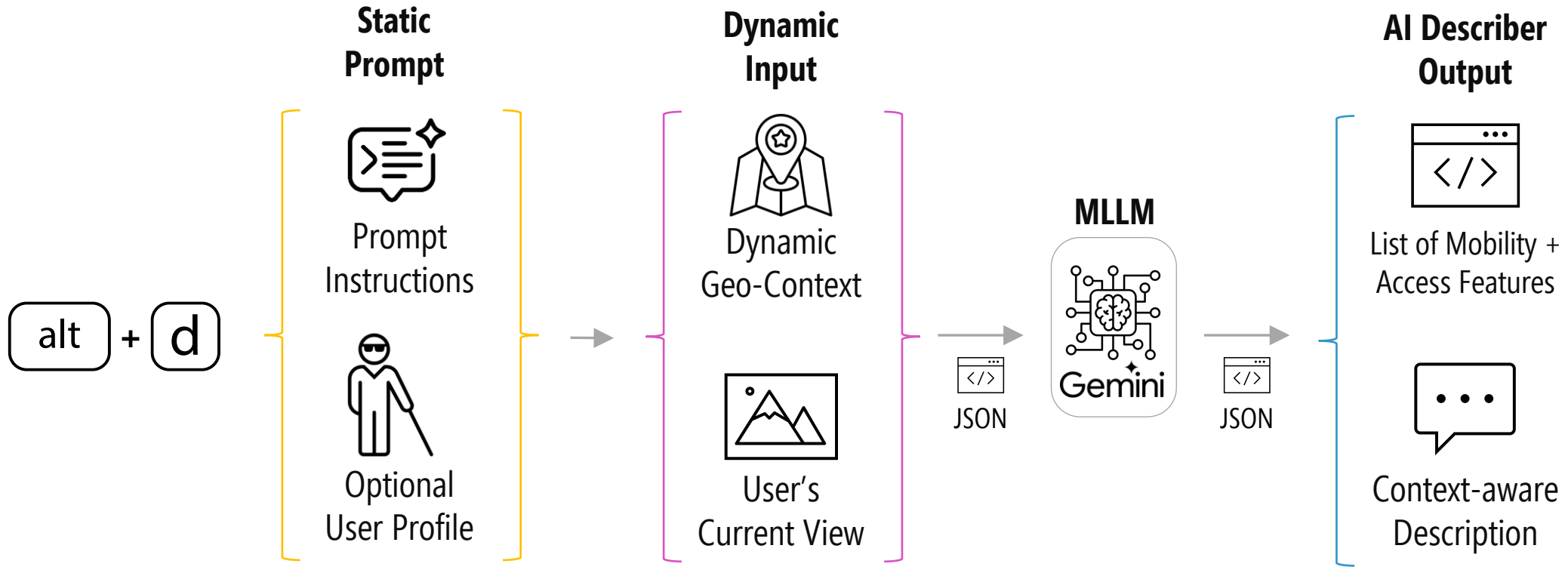
AI DESCRIPTER



AI DESCRIBER: STATIC PROMPT



AI DESCRIPTER



AI DESCRIBER: DYNAMIC INPUT

Dynamic
Prompt

Dynamic
Prompt
Instructions




Dynamic
User
Profile

Dynamic Input

USER LOCATION










-  <lat, lng>
-  Heading
-  Road Address
-  City, State, Country

VIEW METADATA

-  Date
-  Photographer
-  Pitch

NEARBY PLACES

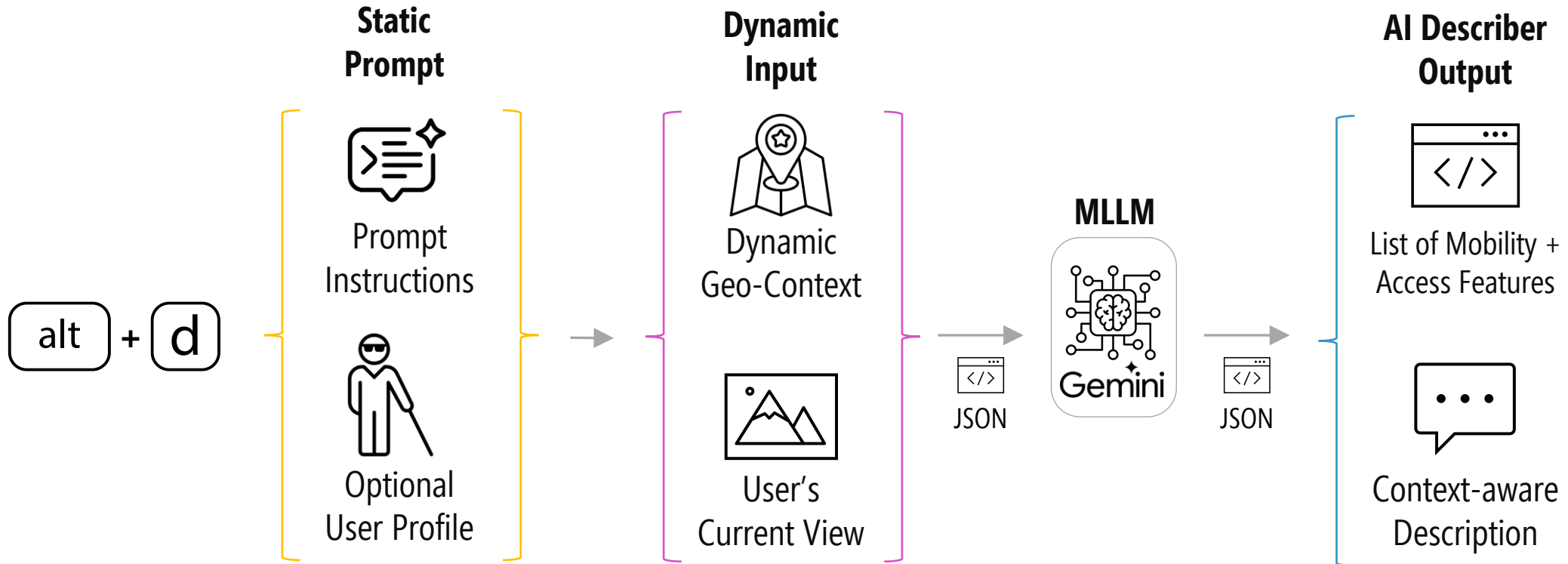


-  Place Name
-  Place Name
-  Place Name
-  Place Type
-  Editorial Summary
-  Place Address
-  <lat, lng>
-  Relative distance
-  Relative heading

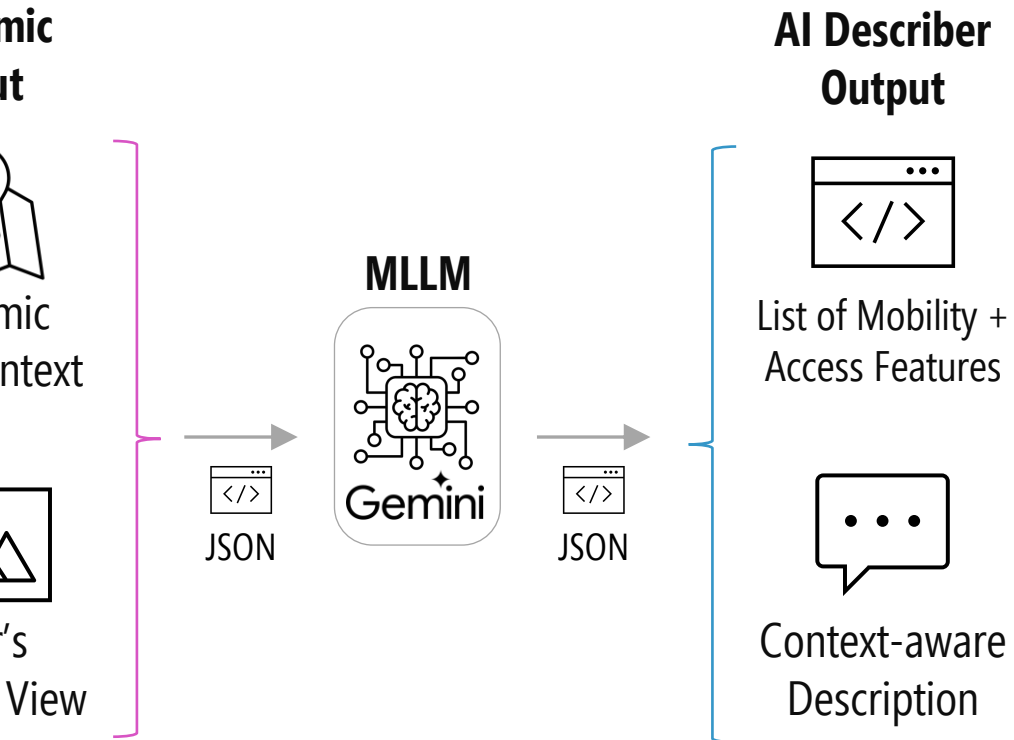
CURRENT USER VIEW



AI DESCRIPTER



AI DESCRIPTOR: OUTPUT



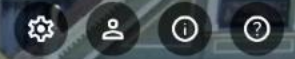
```
{
  mobilityFeatures: [There is a crosswalk in front of you,
    down the hill", "The crosswalk has painted white
    stripes", "There is a stop sign", ...],
  obstacles: ["There are cars parked on the street", "You
    need to walk down a hill"],
  safetySummary: "Use caution when crossing...",
}
```

"The image shows a **view from Alamo Square in San Francisco** with the user facing **northeast**. In front of the user, **across a street with crosswalks**, is a row of Victorian houses, known as the **Painted Ladies**. To the right, **about 35 meters away**, is the Painted Ladies Observatory."



Search for a location to go to...

Nearby Virtual Tours



Hayes St &, Steiner St, San Francisco, CA 94117, USA

Summary Mobility Obstacles Safety

The image shows a view from Alamo Square in San Francisco, with the user facing northeast. In front of the user, across a street with crosswalks, is a row of Victorian houses, known as the Painted Ladies. To the right, about 35 meters away, is The Painted Ladies Observatory.

Loaded from cache Regenerate

AI CHAT

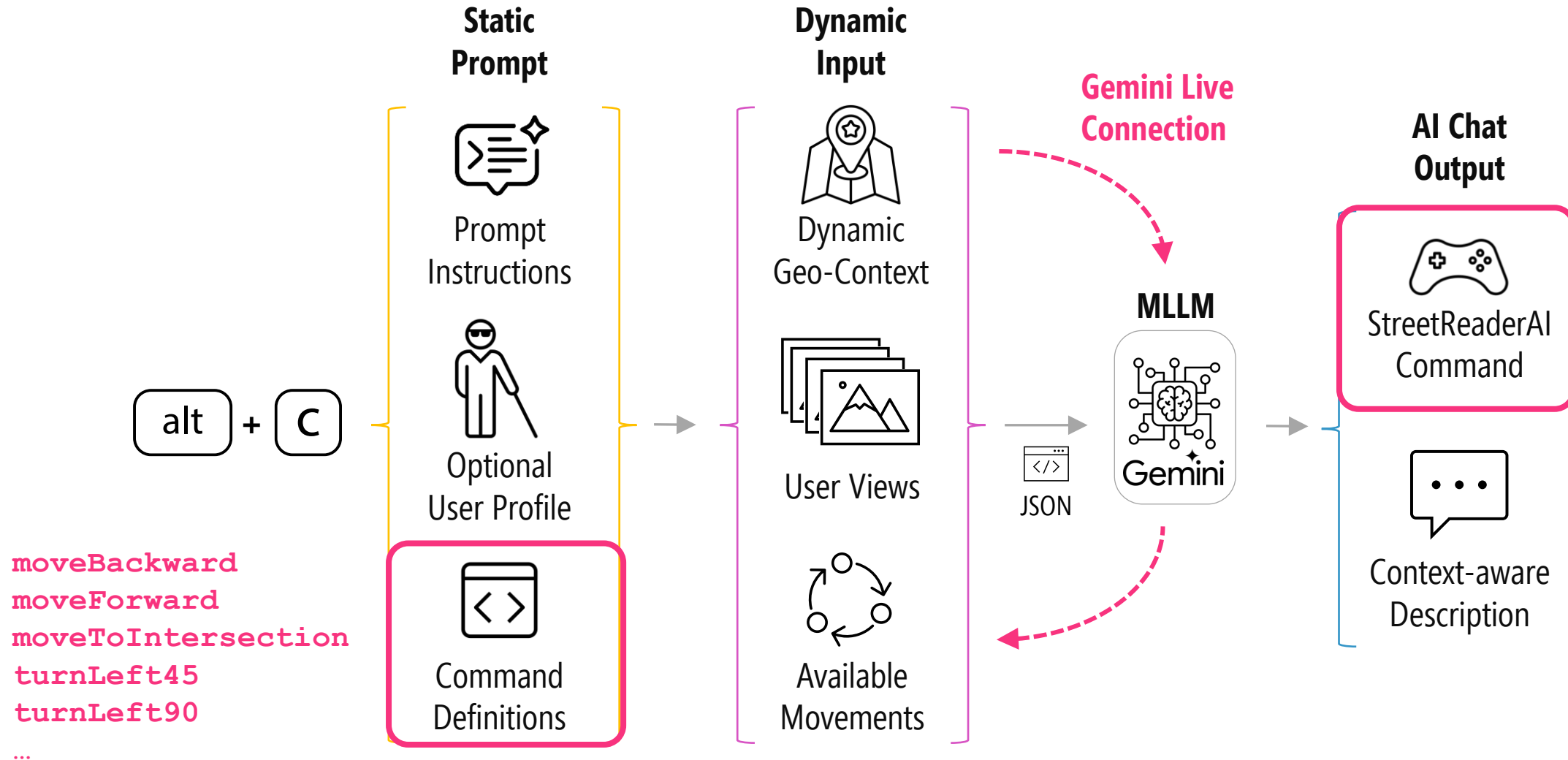
Ask a question about this location...

End Conversation



Google Map Data Terms

AI CHAT: LIVE BIDIRECTIONAL UPDATES



USER STUDY

LAB-BASED USER STUDY



11 Blind Participants

All cane users
Mixed familiarity with AI



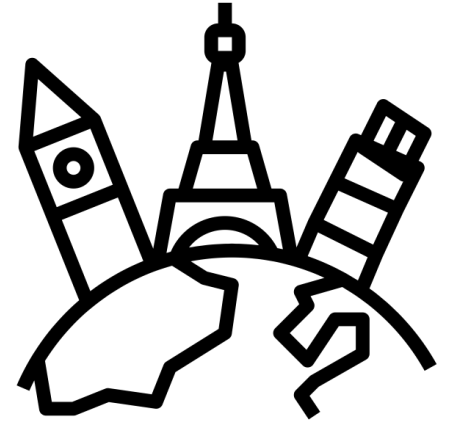
POI Investigations

What does a place look like?



Open-Ended Exploring

I just want to explore...



Virtual Tourism

I've always wanted to visit...

Search for a location to go to...



Nearby



Virtual Tours



Generate Description

Yes, there are a few obstacles including some tables and chairs on the sidewalk as well as a sandwich board sign on the right. There is also a vehicle parked on the street. Be careful when walking in this area.

Are there any curb ramps in the image?

Yes, there is a curb ramp visible in the image in front of you. It appears to be on the other side of the street.

Are there any crosswalks in this area?

Yes, there is a crosswalk in front of you. It appears to be a standard painted crosswalk.

Are there any accessible pedestrian signals around this area?

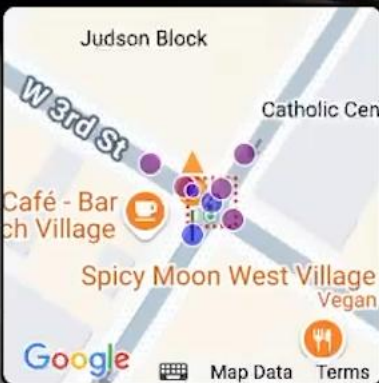
Based on the image, I do not see any accessible pedestrian signals. There are standard traffic signals at the intersection but I cannot confirm if they have an audio component.

Ask a question about this location...



End Conversation

So [it's like] if you just physically moved your body...



LAB-BASED USER STUDY



11 Blind Participants

All cane users
Mixed familiarity with AI



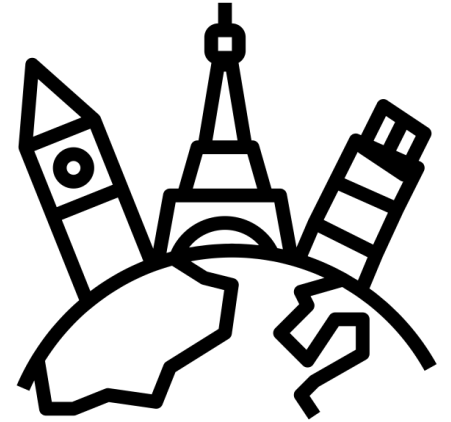
POI Investigations

What does a place look like?



Open-Ended Exploring

I just want to explore...



Virtual Tourism

I've always wanted to visit...

TASK 1: POI INVESTIGATIONS



11 Blind Participants

All cane users
Mixed familiarity with AI



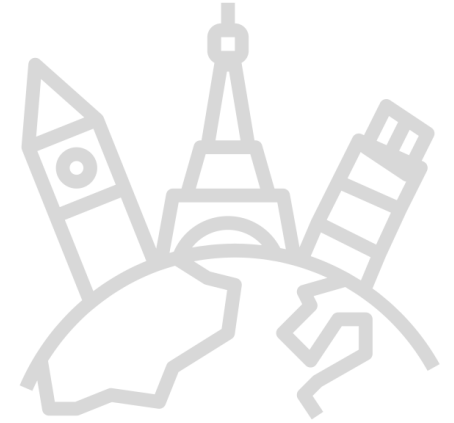
POI Investigations

What does a place look like?



Open-Ended Exploring

I just want to explore...



Virtual Tourism

I've always wanted to visit...



Search for a location to go to...



Nearby

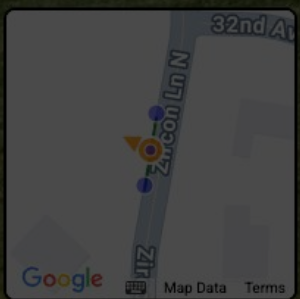
Virtual Tours

Ask a question about this location...



"You plan to take your **two young nieces** to an **unfamiliar playground**. Your goal is to **learn about** the playground **before going**"

TASK 1: PLAYGROUND SCENARIO



Google

Search for a location to go to...



Nearby



Virtual Tours

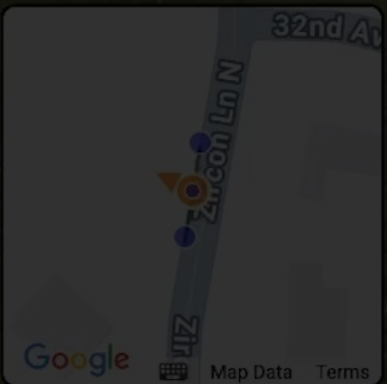
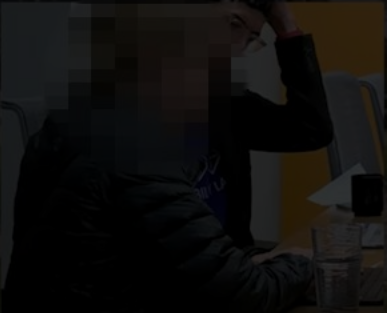
Generate Description

Ask about this location...



Observe the **diversity** and
type of **questions** asked, and
the system's response

PLAYGROUND EXAMPLES



2.0x

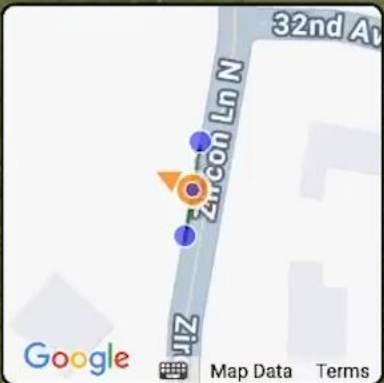
Search for a location to go to...



Nearby Virtual Tours

Generate Description

Ask a question about this location...



2.0x

TASK 1: POI INVESTIGATIONS



11 Blind Participants

All cane users
Mixed familiarity with AI



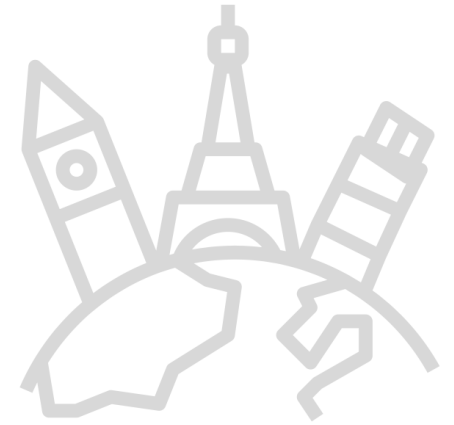
POI Investigations

What does a place look like?



Open-Ended Exploring

I just want to explore...



Virtual Tourism

I've always wanted to visit...

TASK 2: OPEN-ENDED EXPLORING



11 Blind Participants

All cane users
Mixed familiarity with AI



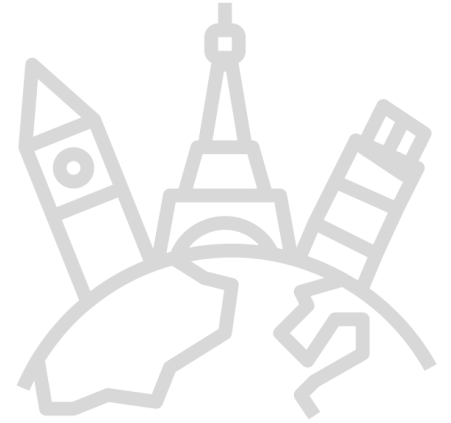
POI Investigations

What does a place look like?



Open-Ended Exploring

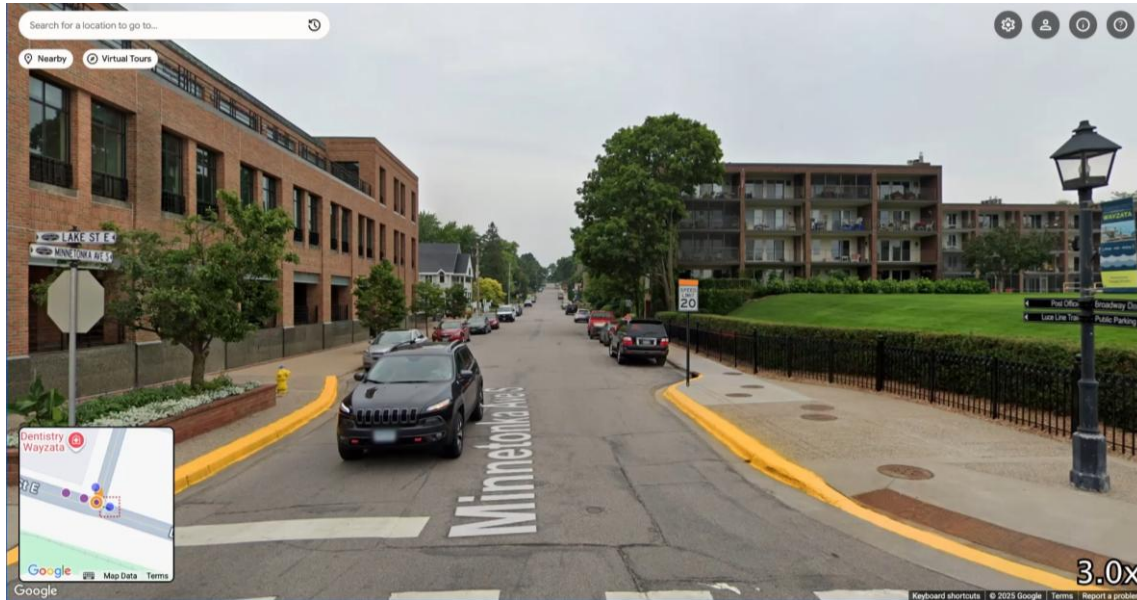
I just want to explore...



Virtual Tourism

I've always wanted to visit...

TWO EXPLORATION SCENARIOS

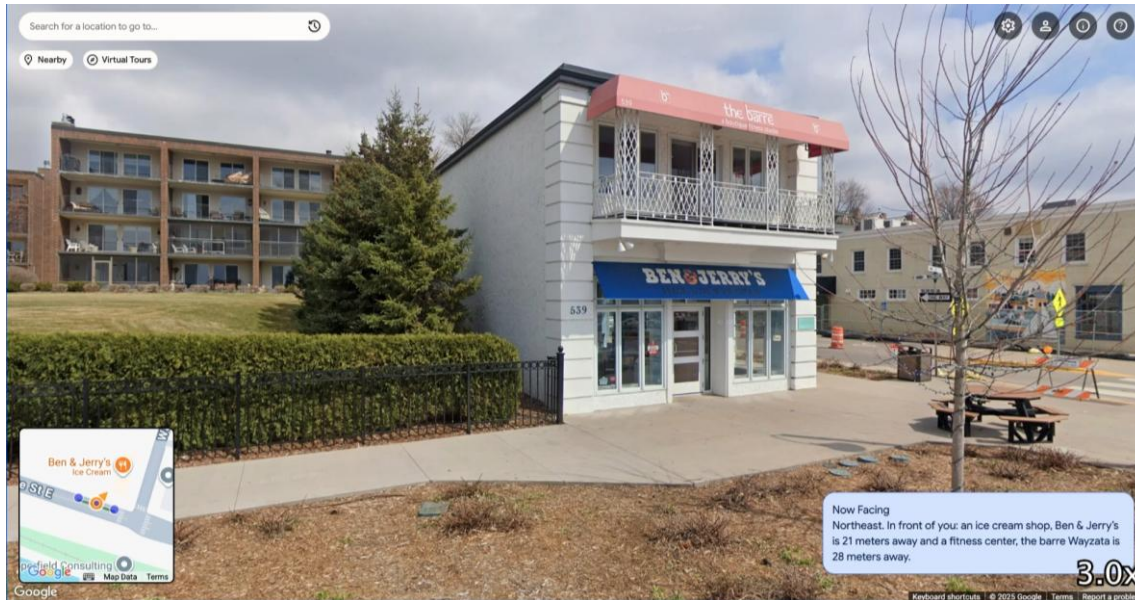


Explore Route to Ice Cream Shop

Walk east to explore the route

110 meters: ~4 heading changes, 12 steps

TWO EXPLORATION SCENARIOS



Explore Route to Ice Cream Shop

Walk east to explore the route

110 meters: ~4 heading changes, 12 steps



Explore Route to Japanese Restaurant

Walk north then west to explore the route

75 meters: ~7 heading changes, 5 steps

USER STUDY RESULTS

OVERALL REACTIONS

Usefulness

6.4/7

Median=7/7 (SD=0.9)

Use in Future?

6.6/7

Median=7/7 (SD=0.8)

*"Most navigation systems can get you to the last 5-10 feet but this **helps get you to the door** and even **describes it**"*

-P10

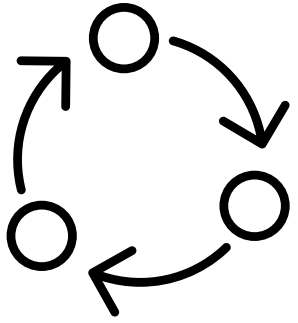
*"This is just **so exciting**—having a conversation about the street [with AI]... this is really amazing"*

-Screen Reader User

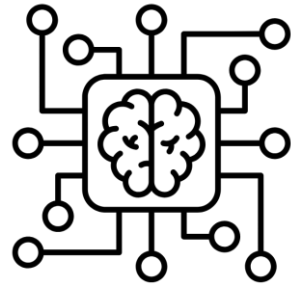
StreetReaderAI

Collaborator

KEY PROBLEMS



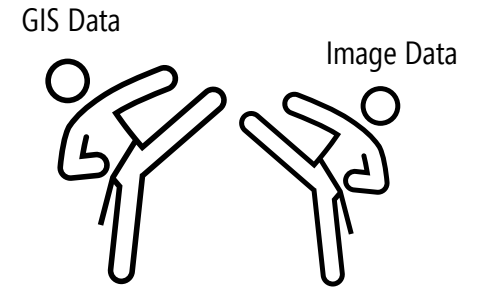
Orientation



Truth?



System Limits



Conflicting Info

USER STUDY RESULTS

AI SYSTEM USES (N=1,053)

AI DESCRIBER

12.9%

136 Instances

vs.

AI CHAT

87.1%

917 Instances

Strong tendency
towards AI Chat



AI SYSTEM USES (N=1,053)

AI DESCRIBER

12.9%

136 Instances

vs.

AI CHAT

87.1%

917 Instances

Strong tendency
towards AI Chat

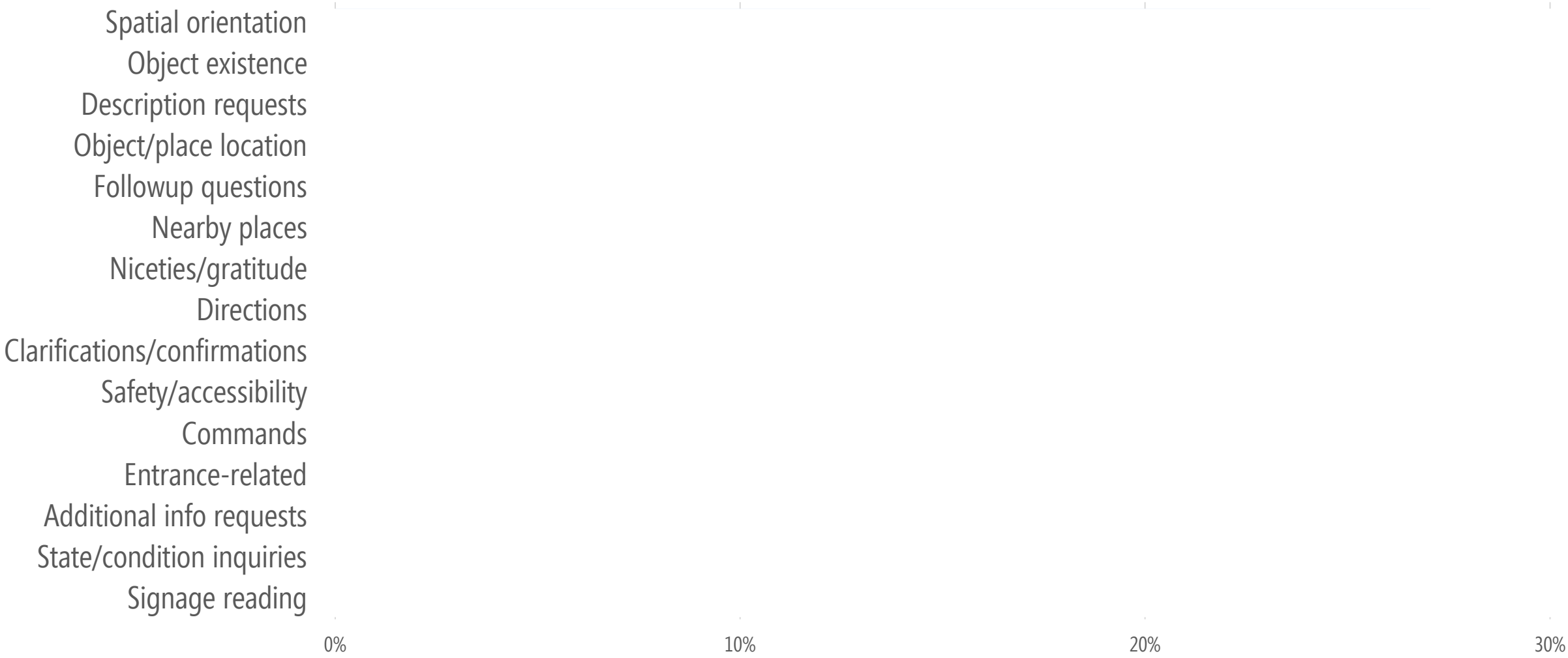


*"The more you **ask**,
the more you **learn**"*

-P10

**What questions do blind people
ask a street view system?**

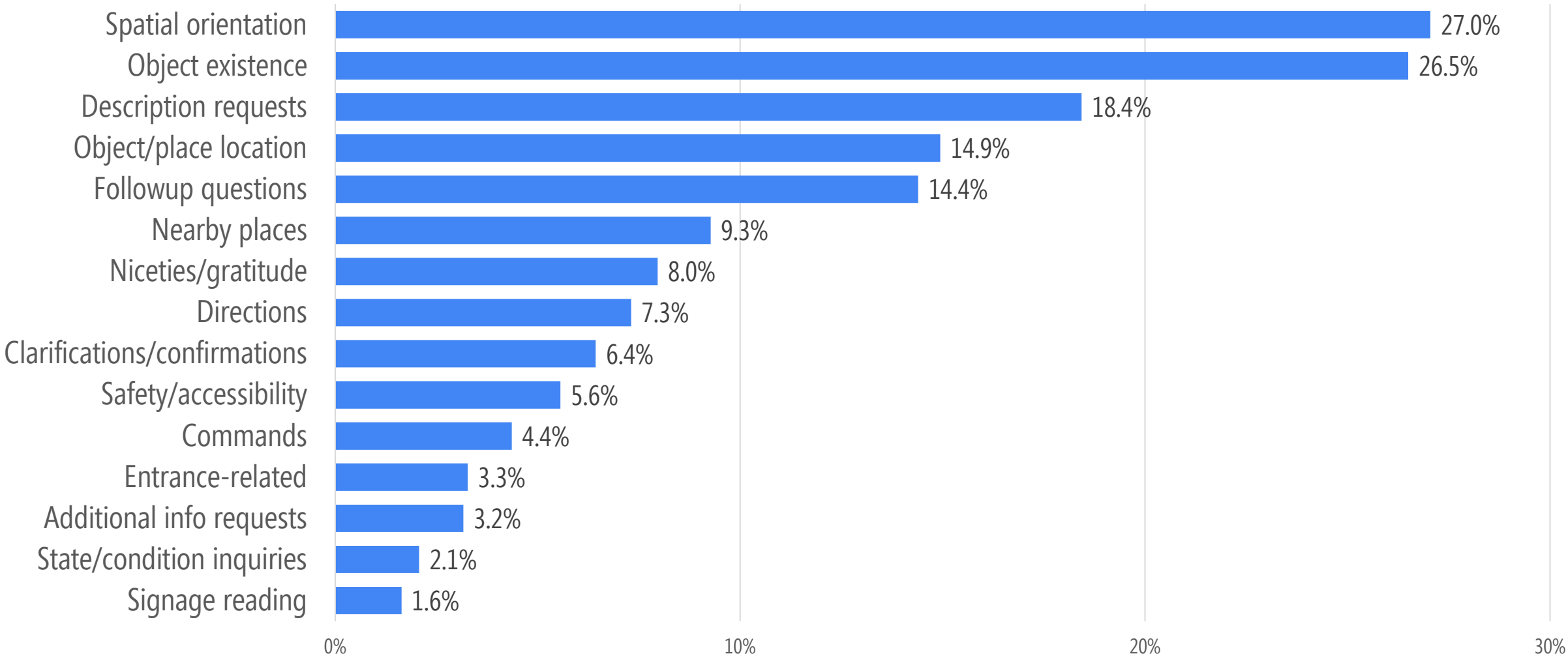
QUESTION TYPES (N=917)



*Each question could be tagged with up to 3 categories

Percentage of AI Chat Interactions

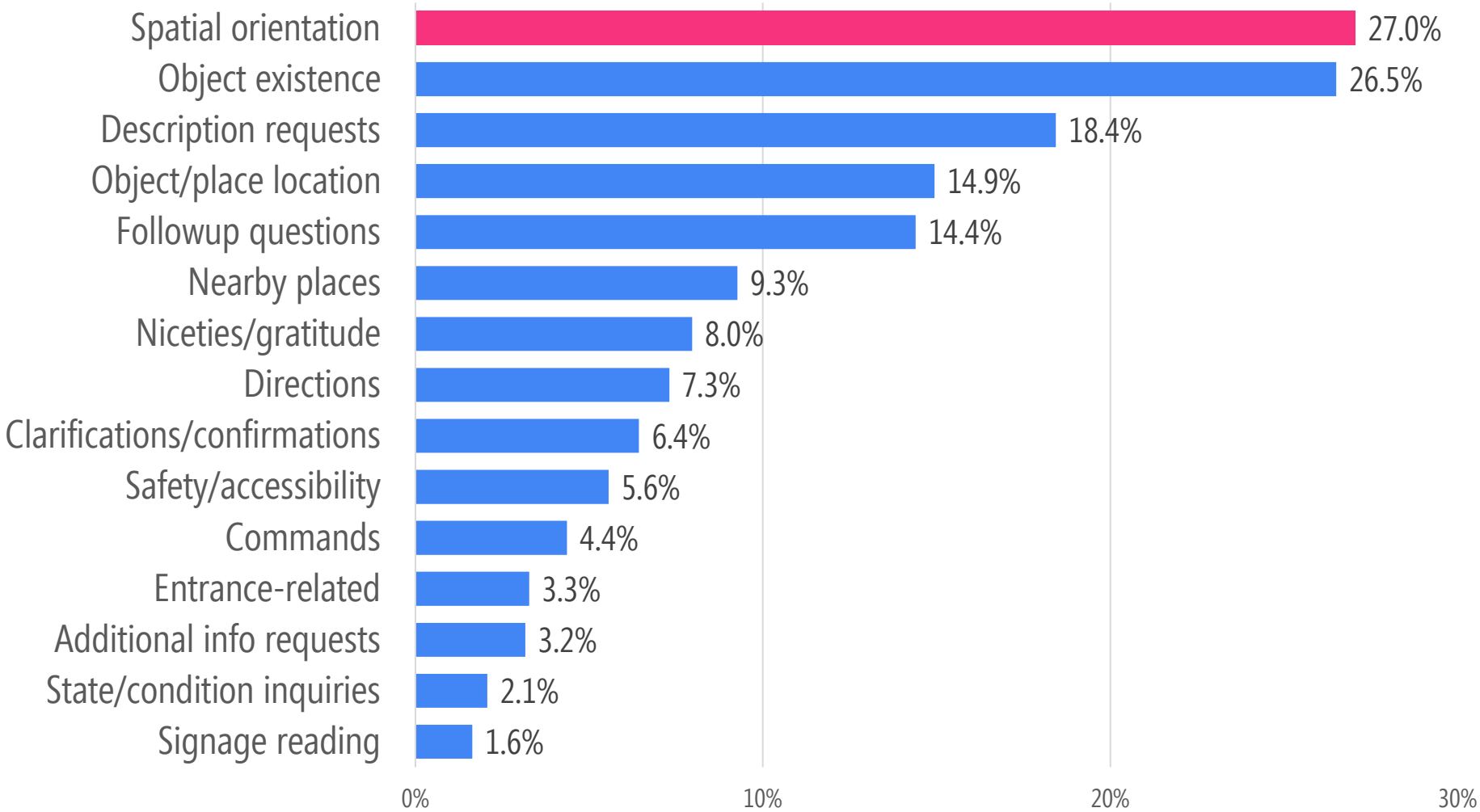
QUESTION TYPES (N=917)



*Each question could be tagged with up to 3 categories

Percentage of AI Chat Interactions

QUESTION TYPES (N=917)



“How far is the bus stop to where I’m standing?”

“Which side is the bus stop sign on?”

“What’s to the left of this building”

...

*Each question could be tagged with up to 3 categories

Percentage of AI Chat Interactions

Search for a location to go to...



Nearby

Virtual Tours



Generate Description

Is there an intersection at the park?

No, there is no intersection visible in the image of the park.

Thank you.

You're welcome! Let me know if you have any other questions.

Is there any benches in the park?

I don't see any benches in the image.

Is there any slides in the park?

Yes, there is a slide visible on the play structure in the park.

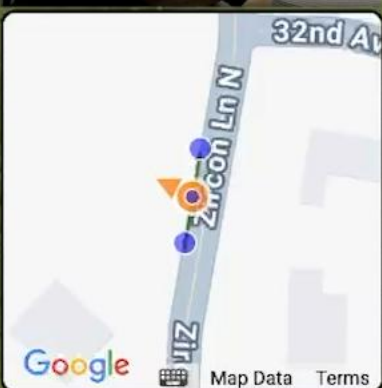
See I get a point for that one.

Indeed you do! You are very observant. Is there anything else I can help you with today?

Listening...

End Conversation

What direction am I facing?



Combine Street View interaction
with AI Chat to **spatially orient**

Search for a location to go to...



Nearby

Virtual Tours



Generate Description

questions.

Is there any benches in the park?

I don't see any benches in the image.

Is there any slides in the park?

Yes, there is a slide visible on the play structure in the park.

See I get a point for that one.

Indeed you do! You are very observant. Is there anything else I can help you with today?

What direction am I facing?

You are currently facing Northwest.

Is the sidewalk going Northwest?

Yes, the sidewalk in front of you appears to be going in a Northwest direction.

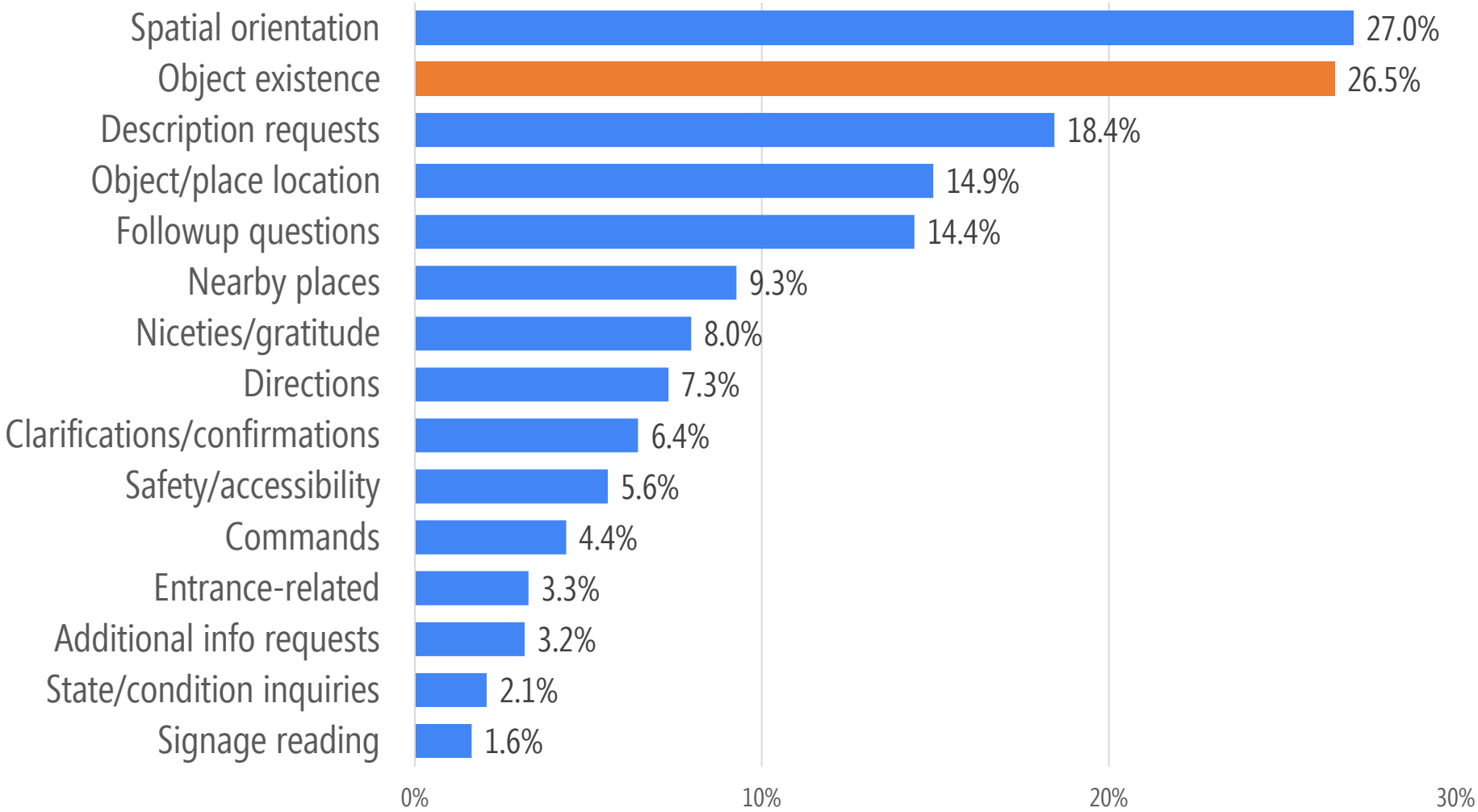
Listening...

End Conversation

Yes, the sidewalk in front of you appears to be going in a northwest direction.

2.0x

QUESTION TYPES (N=917)



"Is there a bench?"

"Are there soccer fields?"

"Is there a crosswalk?"

...

*Each question could be tagged with up to 3 categories

Percentage of AI Chat Interactions

Search for a location to go to...



Nearby

Virtual Tours



Generate Description

Could you tell me what's around me please?

Certainly! You are at a park with a playground. There is a sidewalk in front of you, and the play structure is a bit further ahead in the distance.

Could you describe the play structure for me?

The play structure appears to have a slide, a climbing area, and some swings. It looks like it's made of wood and plastic.

Could you tell me what other structures are in the park?

Based on the image, there are some trees and a house in the background along with the play structure.

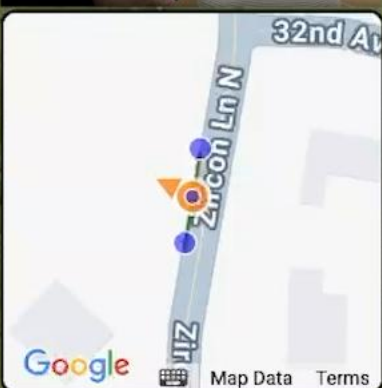
Is that the only structure to play on in the park?

Yes, it appears that the play structure is the only structure to play on in the park.

Listening...

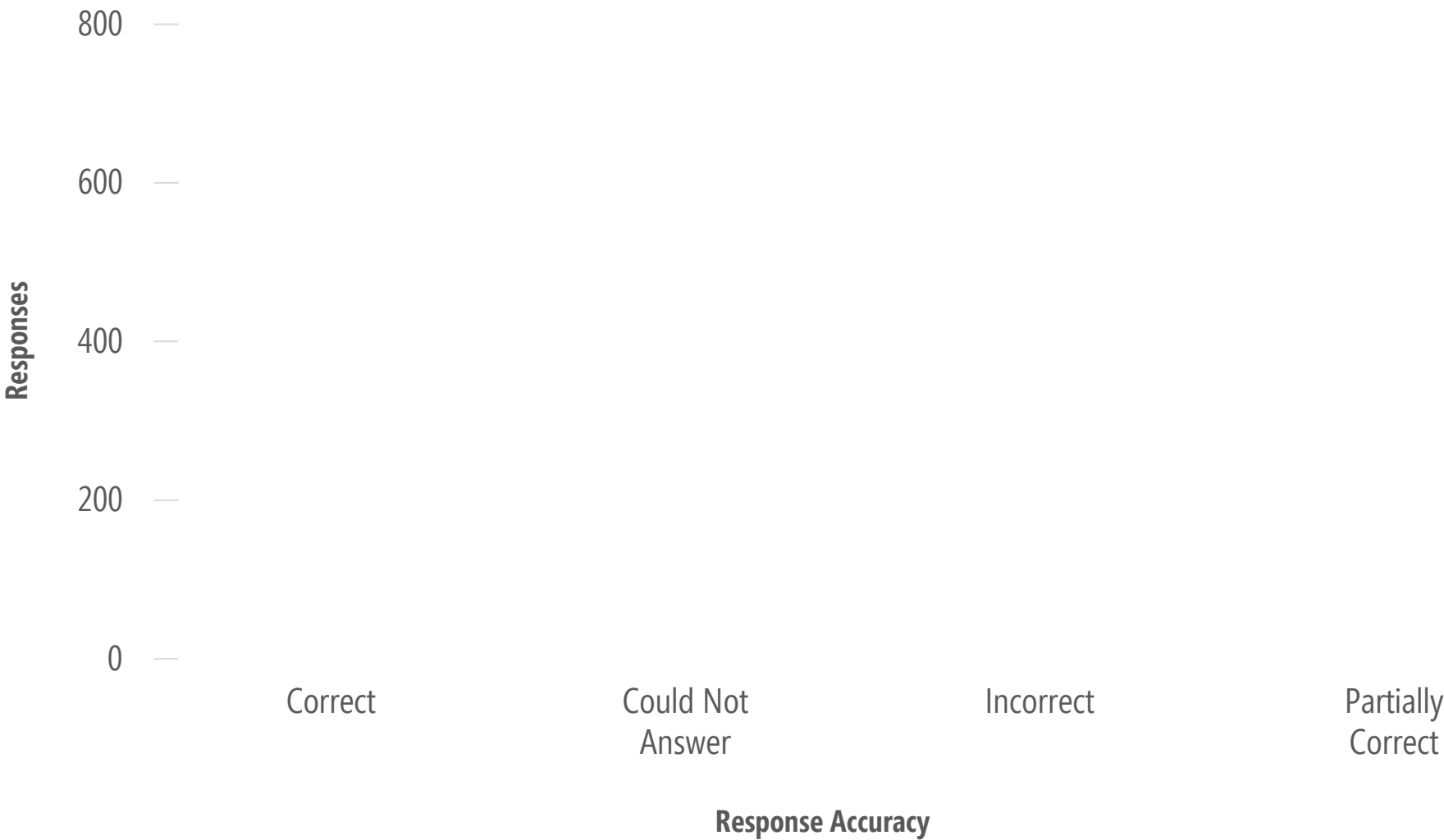
End Conversation

Is there a baseball field
in the park?

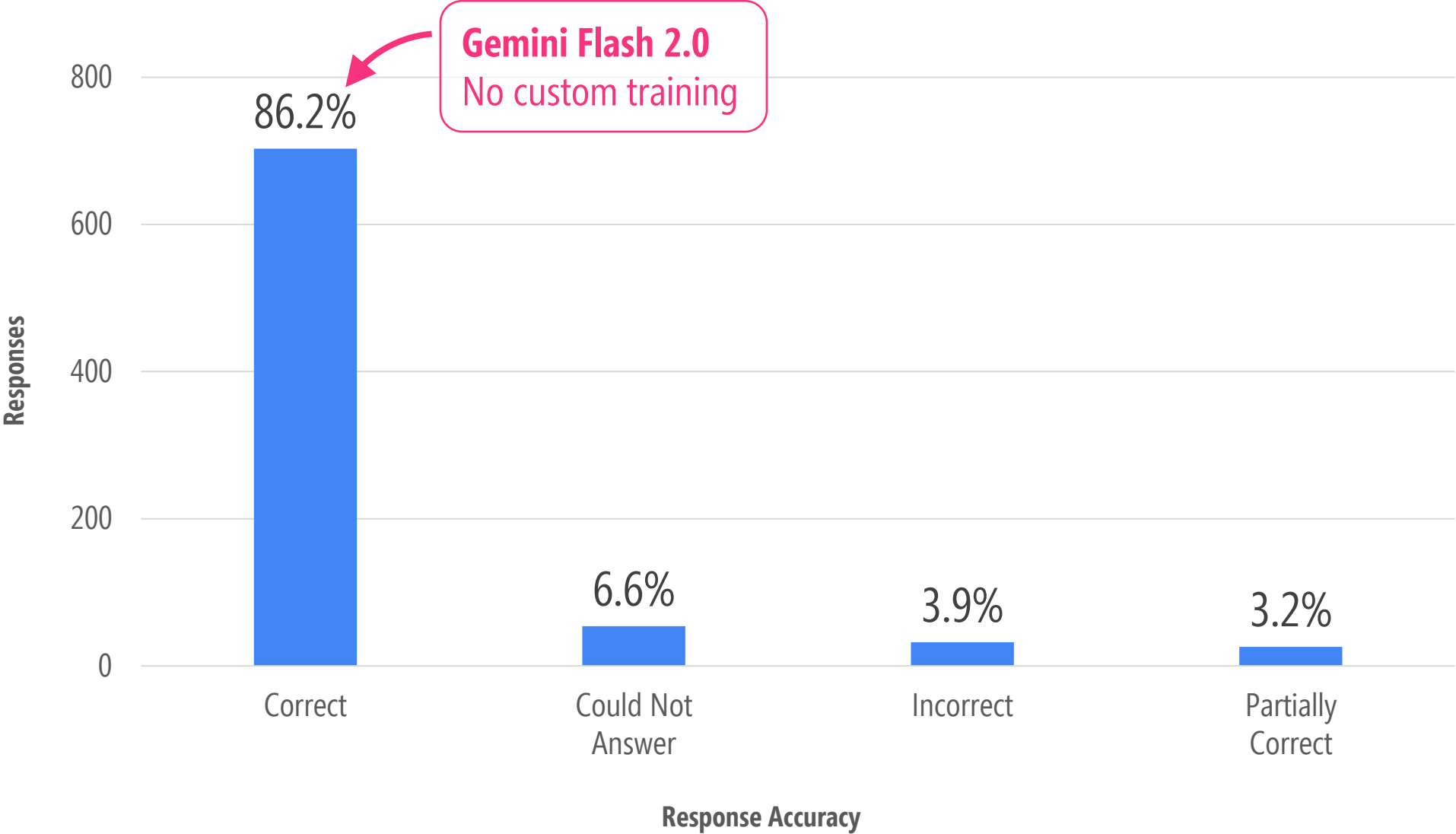


AI model accuracy **is critical**

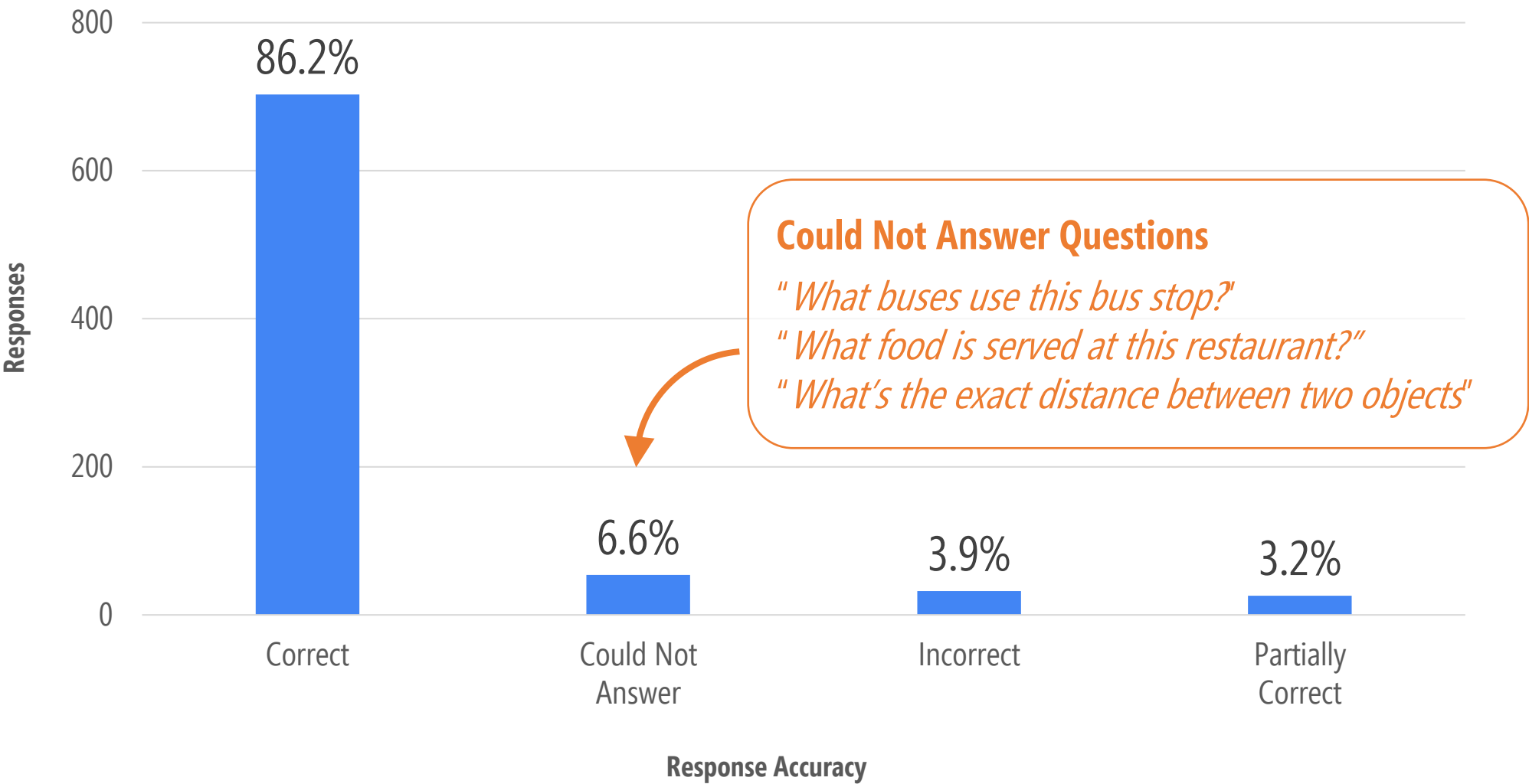
QUESTION RESPONSE ACCURACY (N=816)



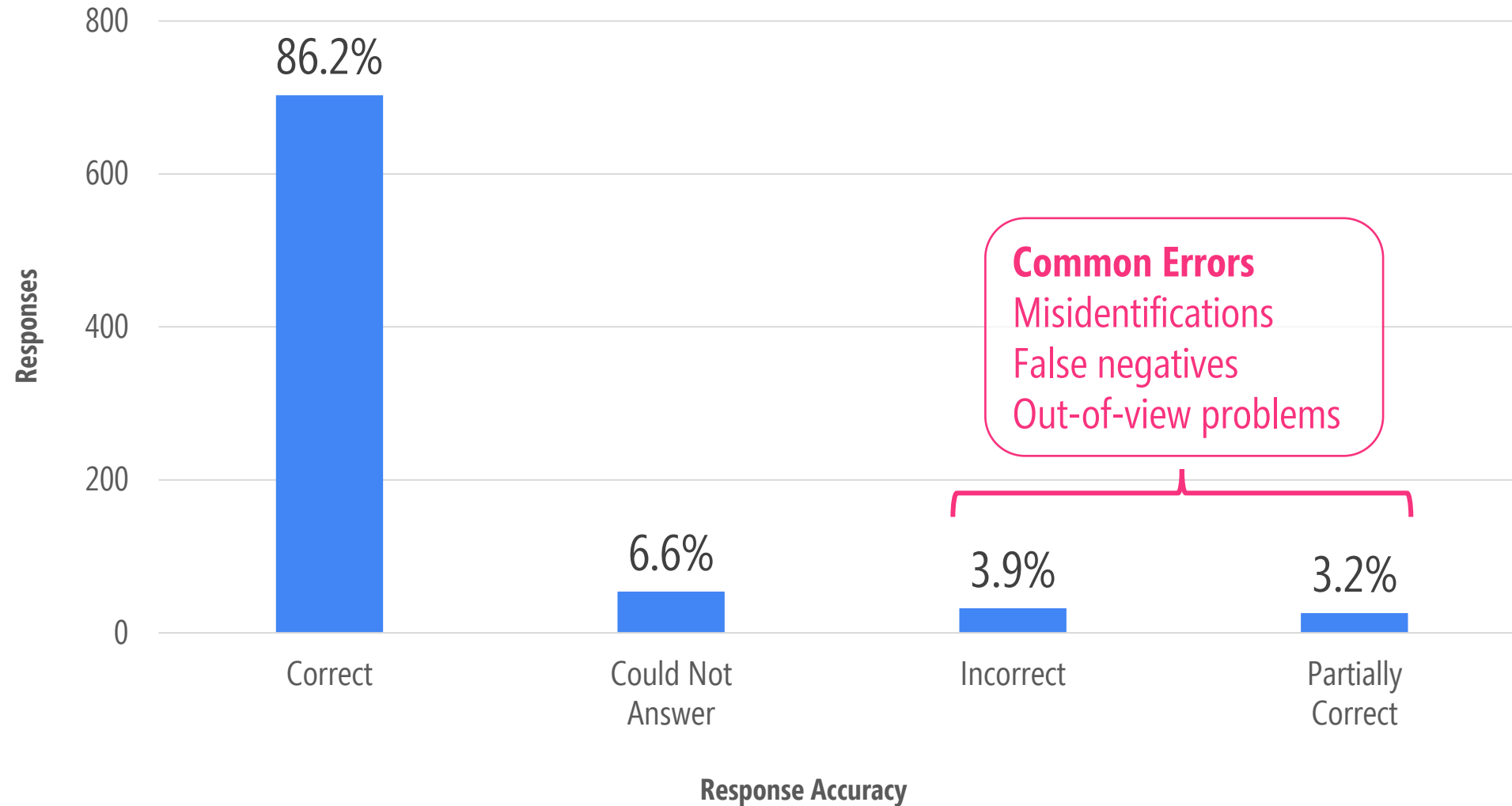
QUESTION RESPONSE ACCURACY (N=816)



QUESTION RESPONSE ACCURACY (N=816)



QUESTION RESPONSE ACCURACY (N=816)



Search for a location to go to...



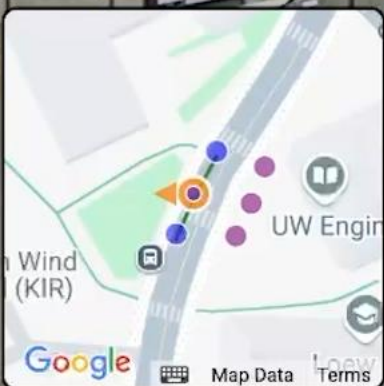
Nearby



Virtual Tours

✗ INCORRECT

Correct answer: no pedestrian signals



Generate Description

the street.

Okay.

Great. Let me know if you have any other questions.

turn me to towards the bus station

Done

what's ahead of me.

The bus stop is directly ahead of you. There are benches and a covered area.

are there any shelters in this stop?

Yes, the bus stop has a covered shelter with benches.

is there an APS system nearby?

I'm sorry, but I don't have the ability to detect APS systems.

Listening...

End Conversation

In some cases, system behaved
inconsistently and **unpredictably**

Search for a location to go to...

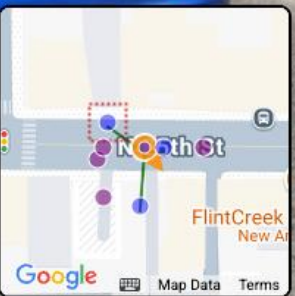
Nearby Virtual Tours



Ask a question about this location...



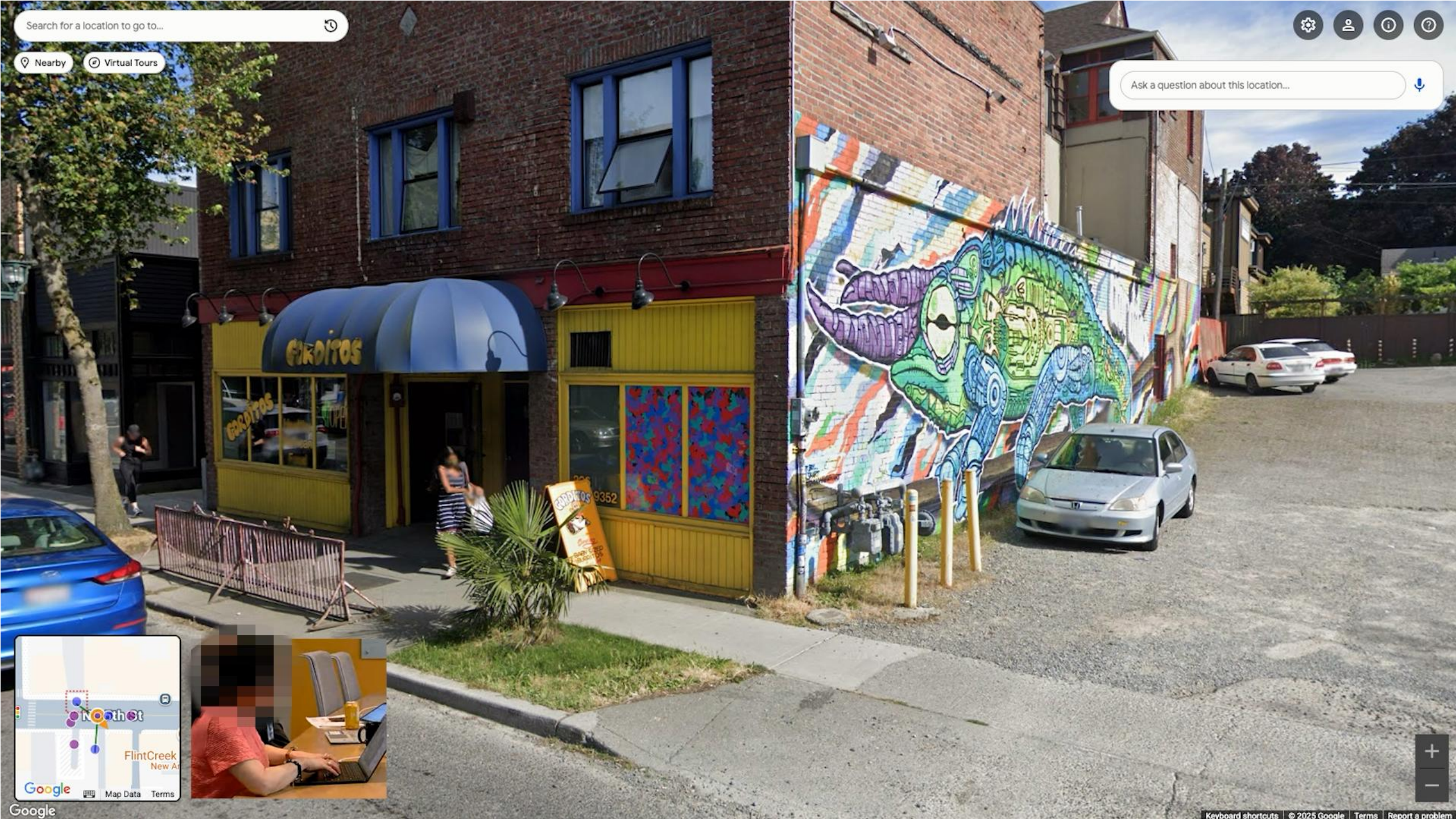
OLD BIKE RACK



Search for a location to go to...

Nearby Virtual Tours

Ask a question about this location...

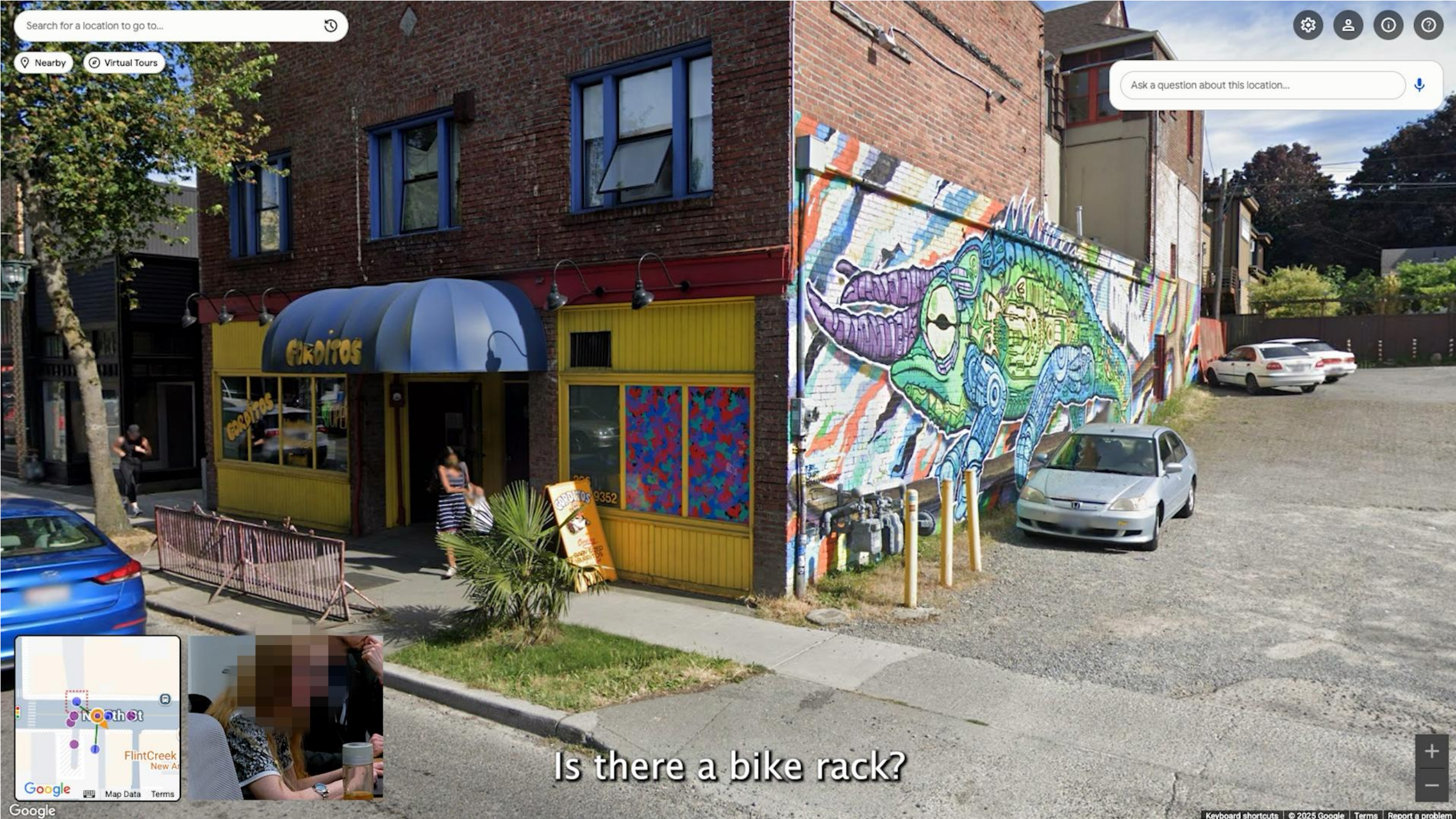


But StreetReaderAI also gets it **right**

Search for a location to go to...

Nearby Virtual Tours

Ask a question about this location...



Is there a bike rack?

User asks a question **before**
the target is visible in image

Search for a location to go to...

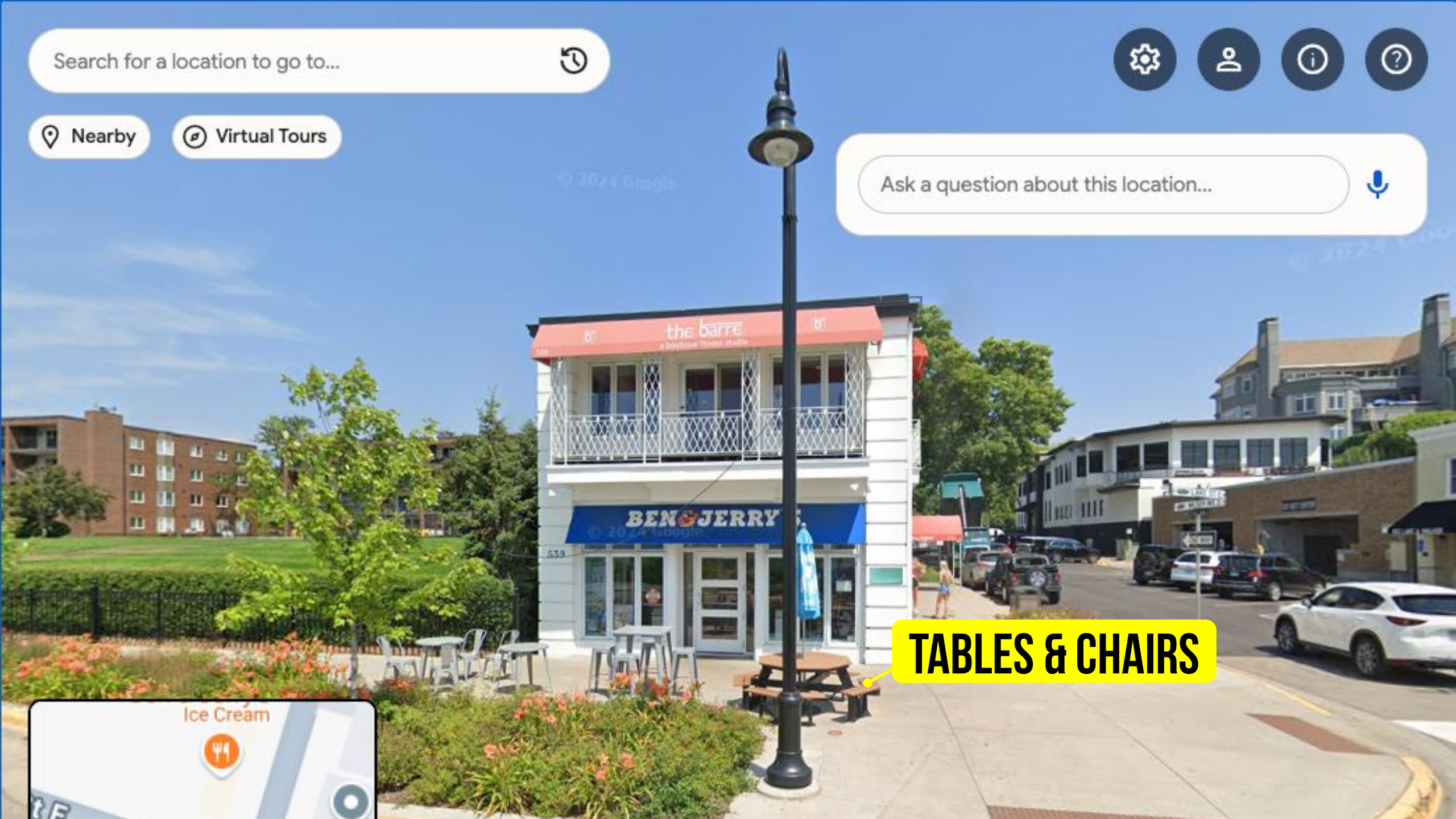


Nearby



Virtual Tours

Ask a question about this location...



TABLES & CHAIRS

Ice Cream



Search for a location to go to...



Nearby

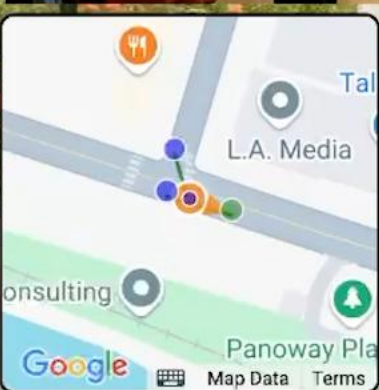
Virtual Tours



✗ INCORRECT

Correct answer: there are tables & chairs

Solution: AI needs better viewpoint



Generate Description

Turn East.

Okay

move forward 100 m

Okay, I have moved you forward.

Verizon move forward

Okay, I have moved you forward.

forward to Ben & Jerry's

Okay, I have moved you forward. Ben & Jerry's is to your left.

Listening...

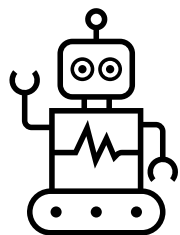
End Conversation

Okay. I have moved you forward.

2.0x

CV4A11y

ICCV'25 Workshop



Agentic AI

“Does the cafe entrance look accessible? Where is the door?” Towards Geospatial AI Agents for Visual Inquiries

Jon E. Froehlich^{1,2} Jared Hwang¹ Zeyu Wang¹ John S. O’Meara¹ Xia Su¹ William Huang³
Yang Zhang³ Alex Fiannaca⁴ Philip Nelson² Shaun Kane²

¹University of Washington ²Google Research ³UCLA ⁴Google DeepMind

jonf@cs.uw.edu

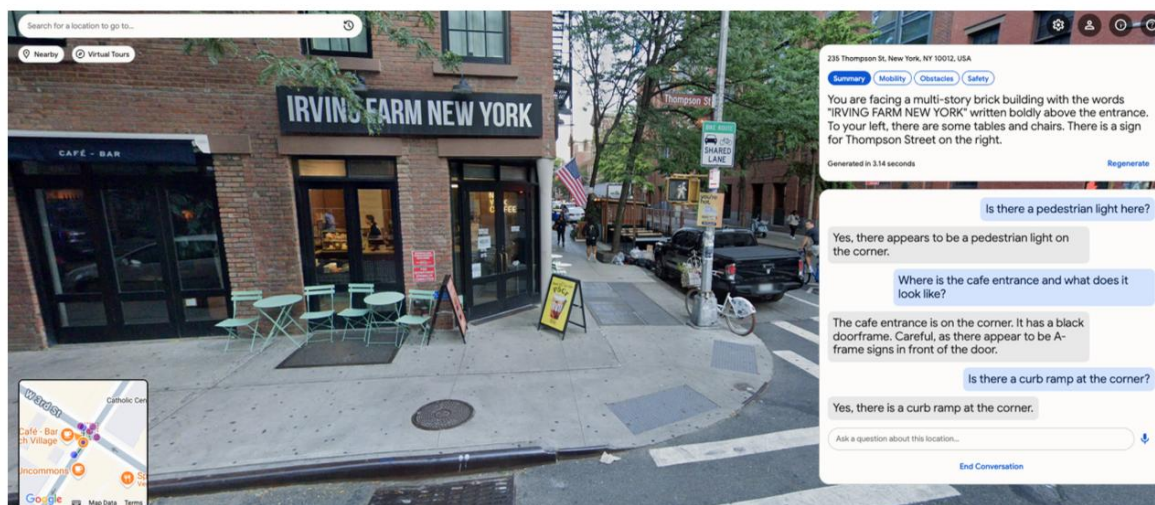


Figure 1. We introduce our vision for *Geo-Visual Agents*—multimodal AI agents capable of understanding and responding to nuanced visual-spatial inquiries about the world by analyzing large-scale repositories of geospatial images combined with traditional GIS data sources. For example, *StreetViewAI* [14] (above) makes street view accessible to blind users by combining geographic context, user information, and dynamic street view images into an MLLM, accessed via an AI chat interface and accessible screen reader controls.

Abstract

Interactive digital maps have revolutionized how people travel and learn about the world; however, they rely on pre-existing structured data in GIS databases (e.g., road networks, POI indices), limiting their ability to address geovisual questions related to what the world looks like. We introduce our vision for Geo-Visual Agents—multimodal AI agents capable of understanding and responding to nuanced

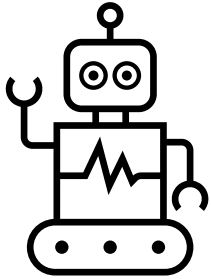
1. Introduction

Over the last two decades, precise location sensing, pervasive internet connectivity, and interactive digital maps have transformed human mobility from travel planning to *in situ* navigation.. Despite these advances, current mapping systems are confined to pre-existing structured geospatial data, leaving a vast repository of visual information—latent within street-level, aerial, and user-contributed imagery—

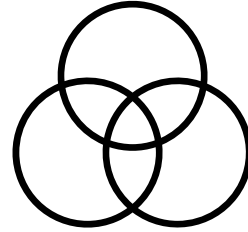
Proposes a
more **agentic**
approach

FUTURE

FUTURE WORK



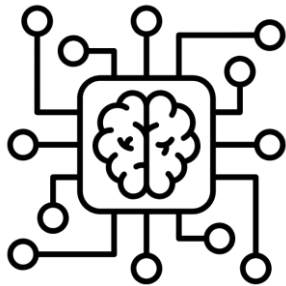
Agentic AI



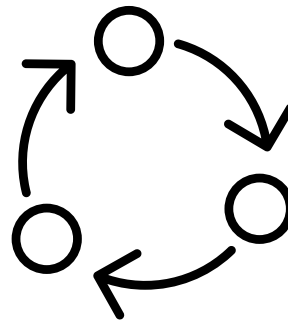
**Multimodal
Data Sources**



Data Provenance



**Model
Transparency**



**Improved SV
Immersion**



**Handoff to In-Situ
Experience**

A man with short brown hair, wearing a dark jacket over a grey t-shirt, is sitting at a desk and typing on a laptop. In the foreground, the back of a person wearing a red jacket is visible, their face blurred. The background shows an office environment with a whiteboard and a window. The text "CLOSING QUOTE" is overlaid in the center.

CLOSING QUOTE

Having the freedom to get lost.



Having the freedom to get lost.

StreetReaderAI

Making Street View Accessible via
Context-Aware, Multimodal AI

Come find me for a
demo at UIST!



**JON E.
FROEHLICH**
Google
Research



**ALEX
FIANNACA**
Google
DeepMind



**NIMER
JABER**
Google



**VICTOR
TSARAN**
Google



**SHAUN
KANE**
Google
Research



Technology
AI Society
and Culture



AI for
Social
Good



People + AI
Research

