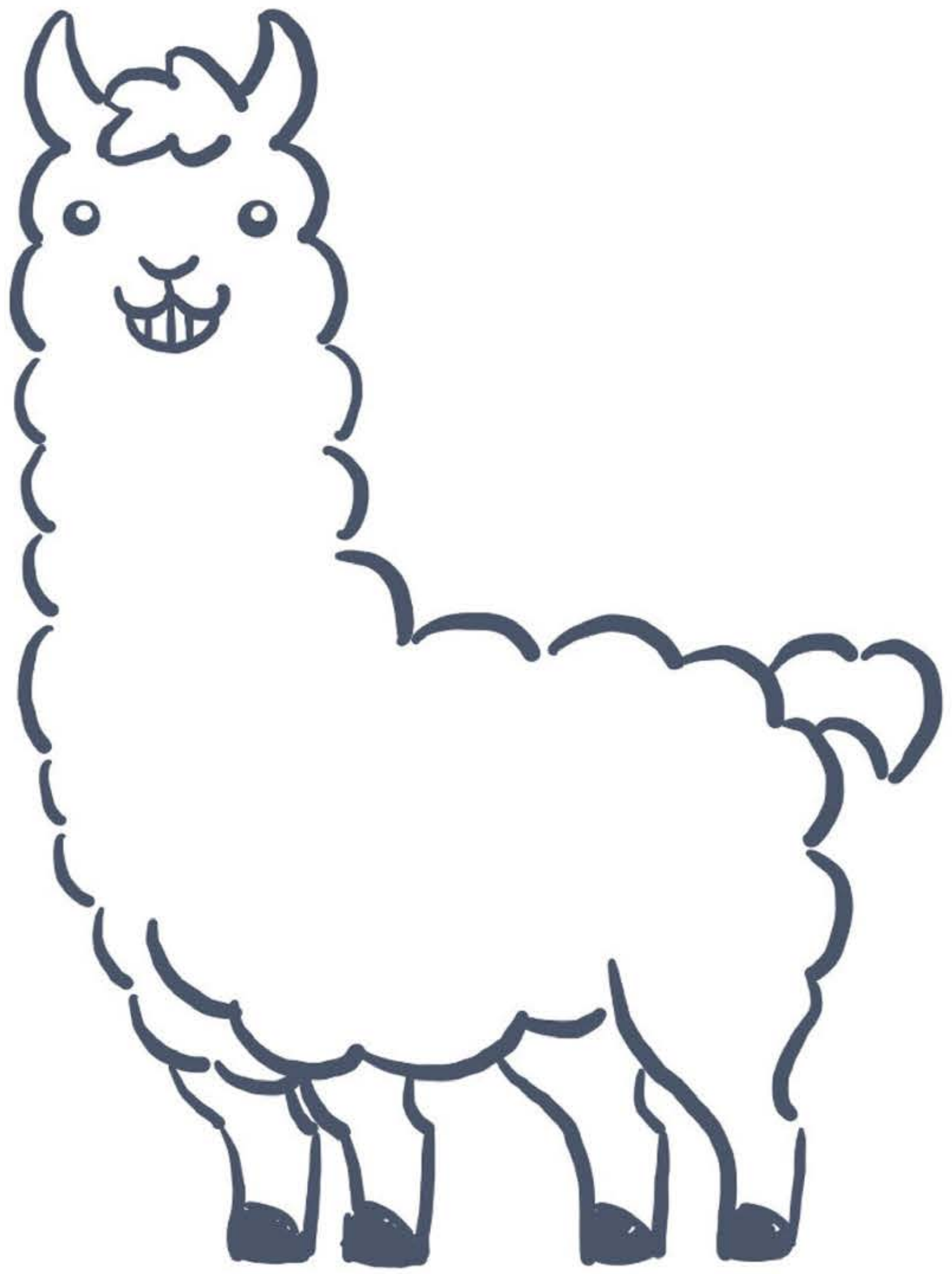
LabelAId Just-in-time AI Interventions for Improving Human Labeling Quality and Domain Knowledge

Chu Li*, Zhihan Zhang*, Michael Saugstad, Esteban Safranchik, Minchu Kulkarni, Xiaoyu Huang, Shwetak Patel, Vikram Iyer, Tim Althoff, Jon E. Froehlich

MAKEABILITY LAB
DUB DESIGN USE BUILD
CREATE Center for Research and Education on Accessible Technology and Experiences
W PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING
Berkeley UNIVERSITY OF CALIFORNIA

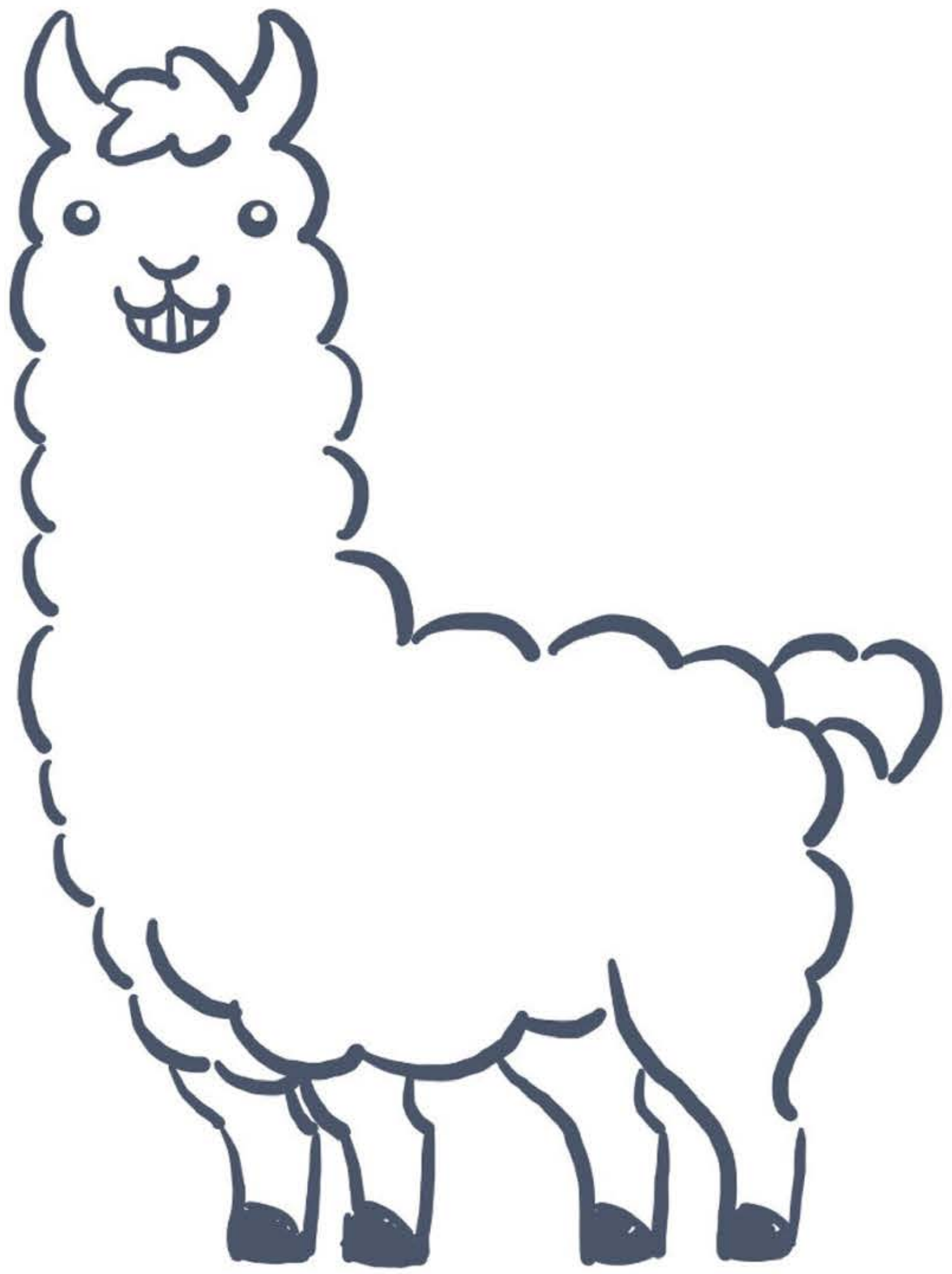


Llama

Alpaca

Are you sure this is a llama?

System feedback can influence labeling confidence, quality & knowledge

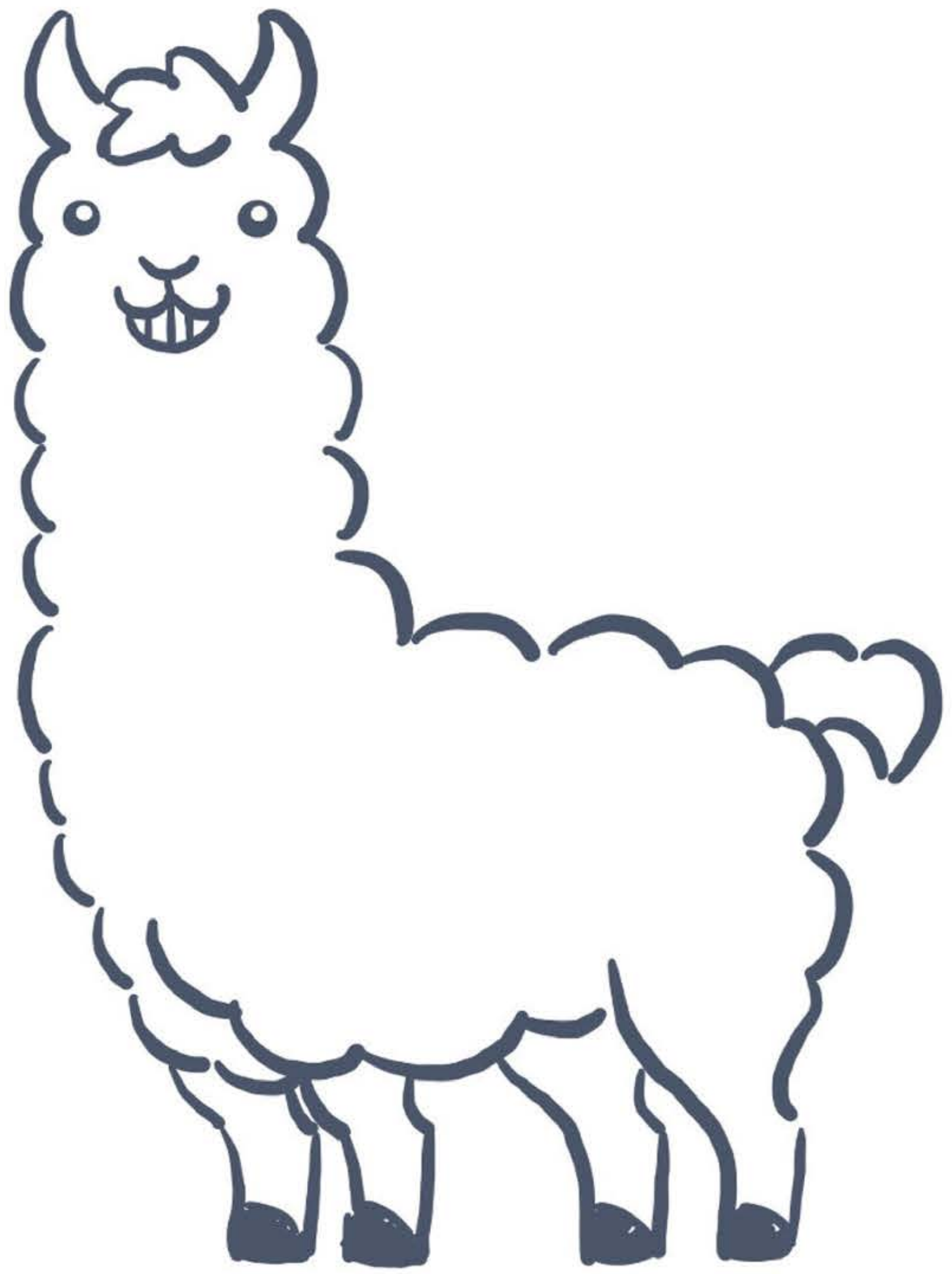


Llama

Alpaca

Are you sure this is an alpaca?

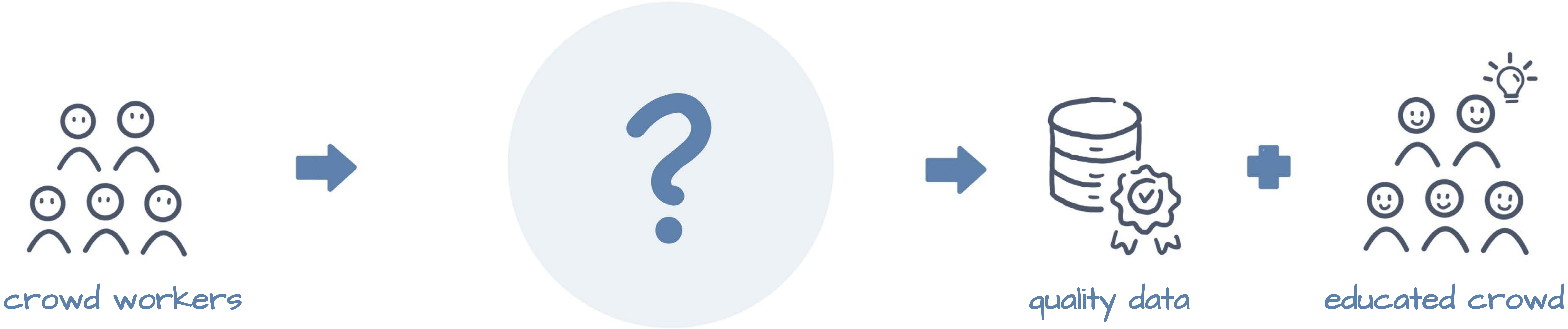
System feedback can influence labeling confidence, quality & knowledge

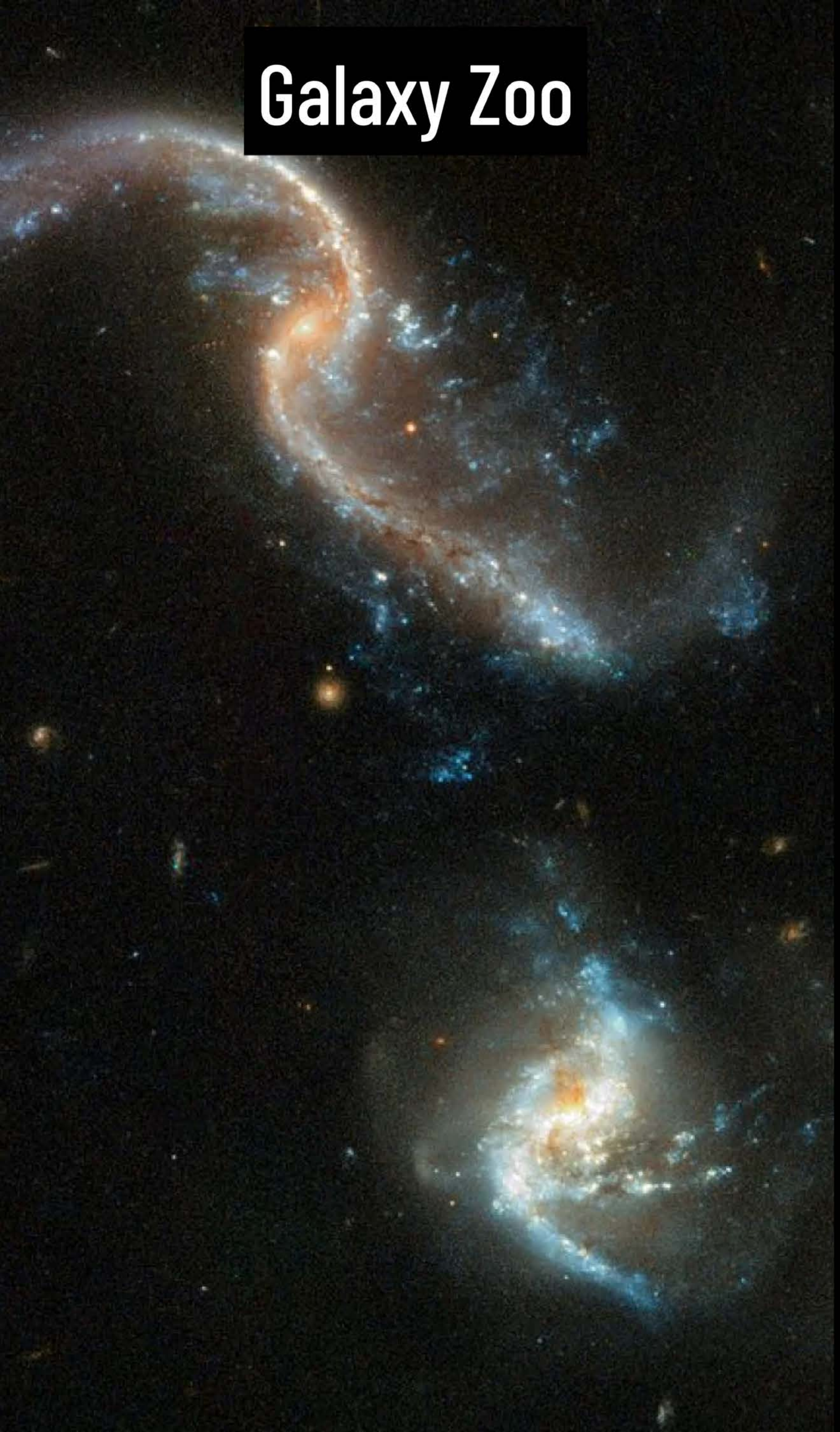


Llama

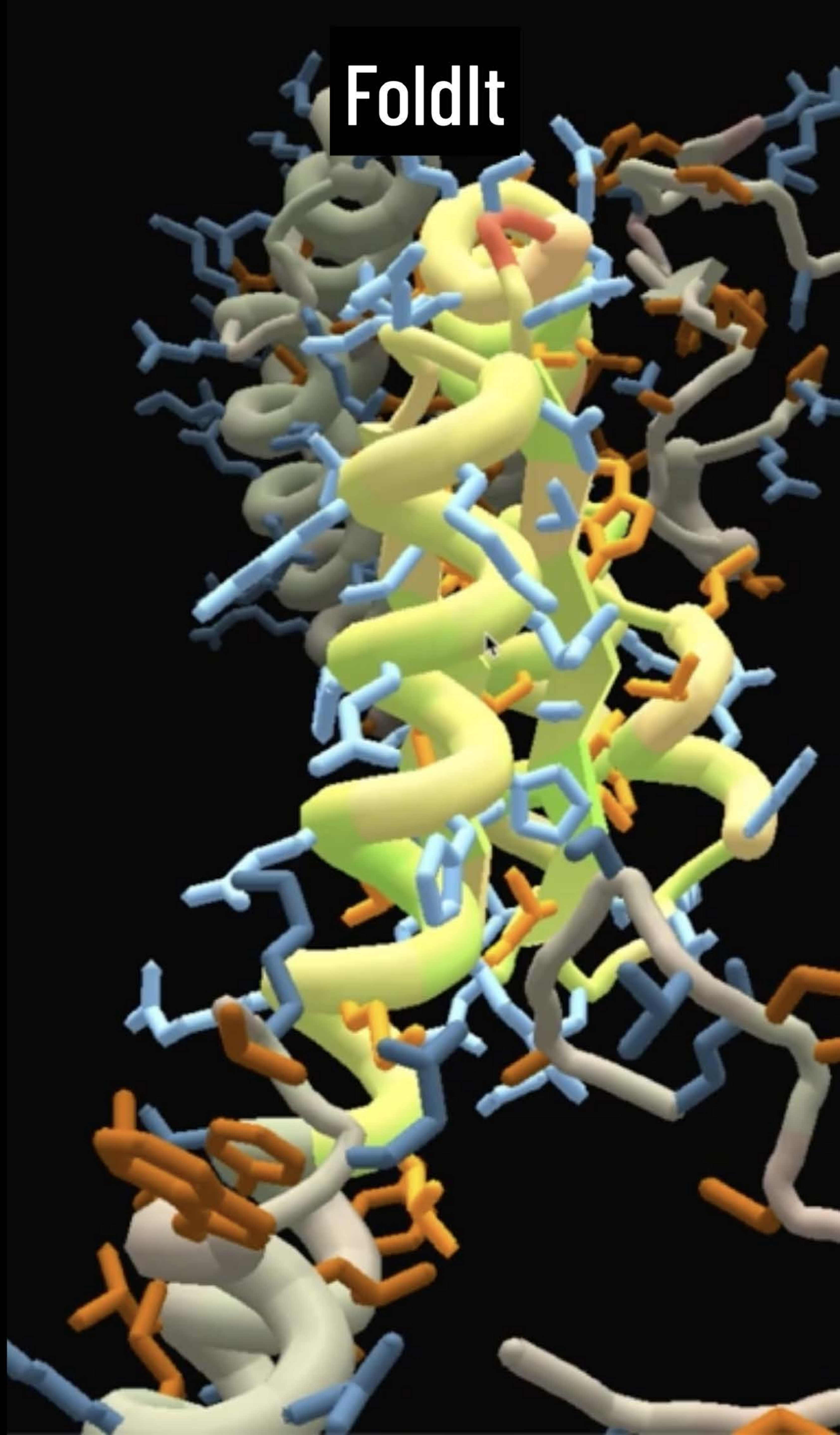
Alpaca

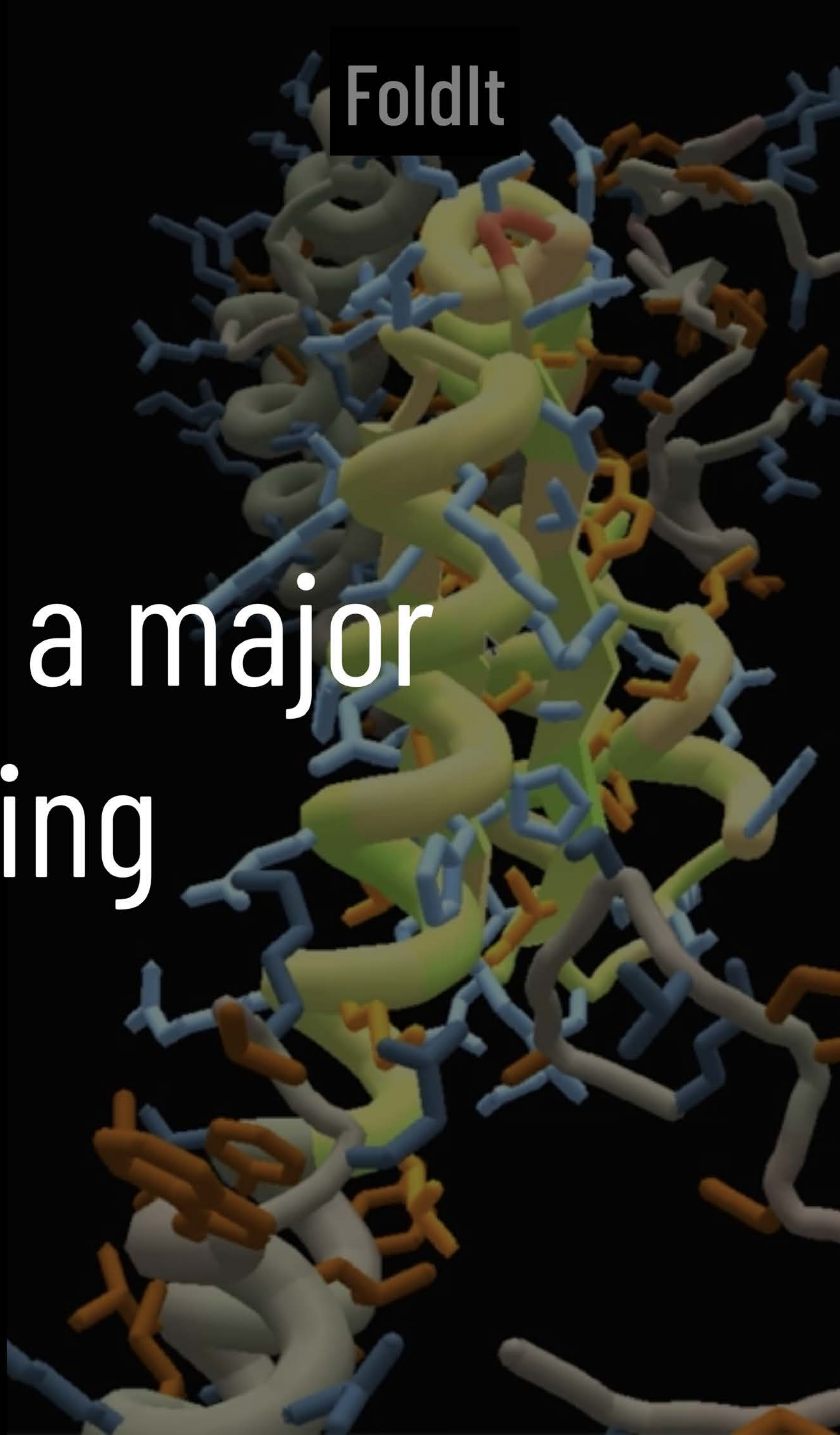
How can we leverage system feedback?




Galaxy Zoo


Open Street Map


FoldIt



Quality control remains a major challenge in crowdsourcing

Prior Work

Quality Control + Learning Experience

Example Projects

- Shepherd
  Dow et al., 2012
- Review vs. Doing
  Zhu et al., 2014
- Learning from the Crowd
  Mamykina et al., 2016

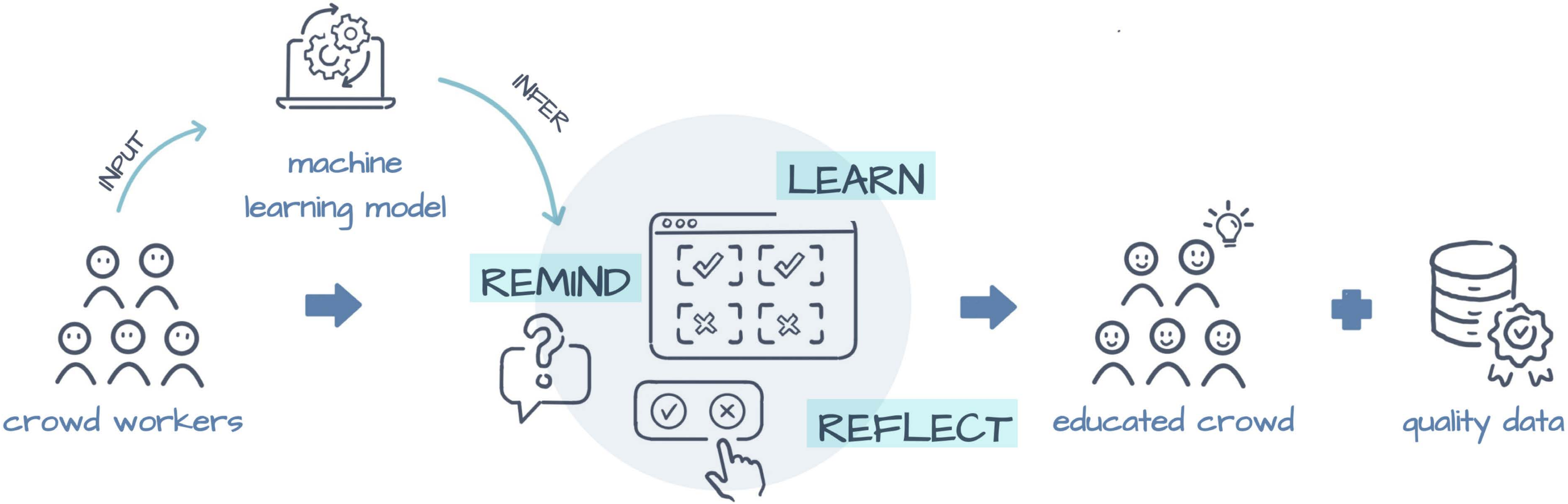
Learning Facilitation Methods

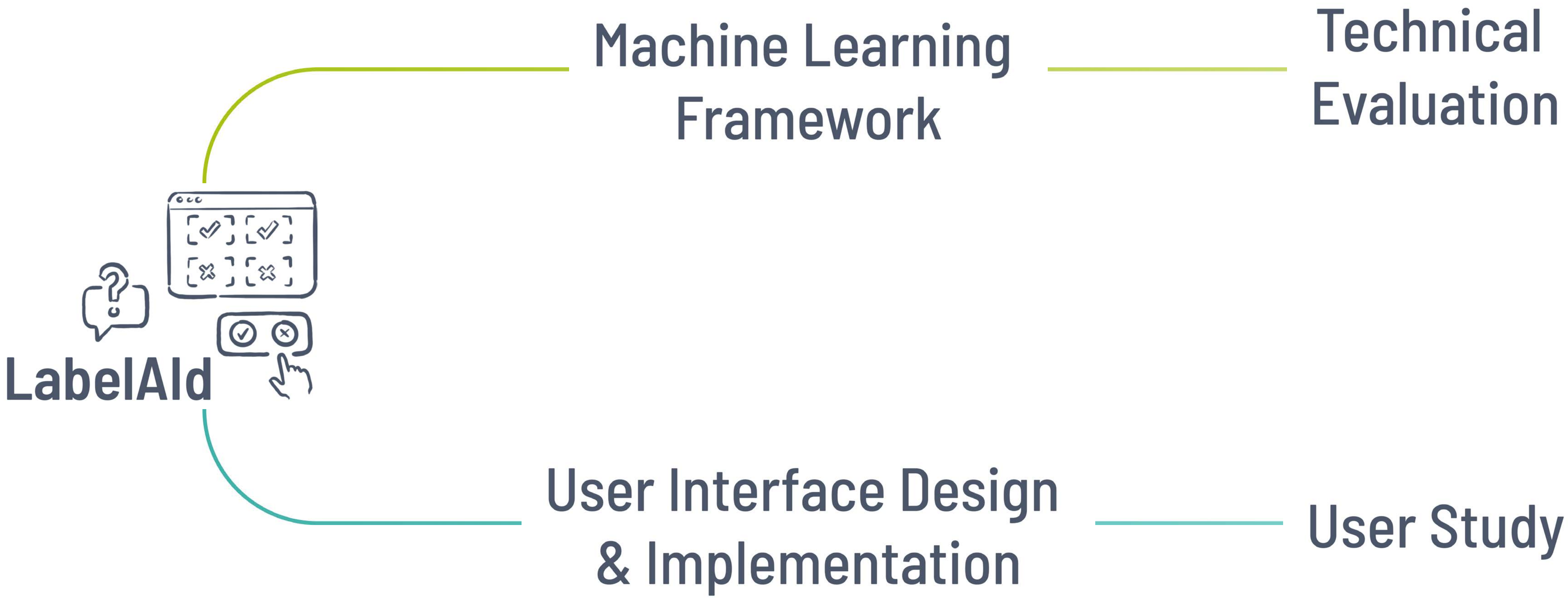
- Peer review
- Expert feedback
- Self-assessment

Limitation

- Requires additional review commitments
- Impacts scalability

LabelAId


INPUT
machine
learning model
INFER
LEARN
REMIND
REFLECT
crowd workers
educated crowd
quality data


LabelAId
Machine Learning Framework
Technical Evaluation
User Interface Design & Implementation
User Study

Machine Learning Framework

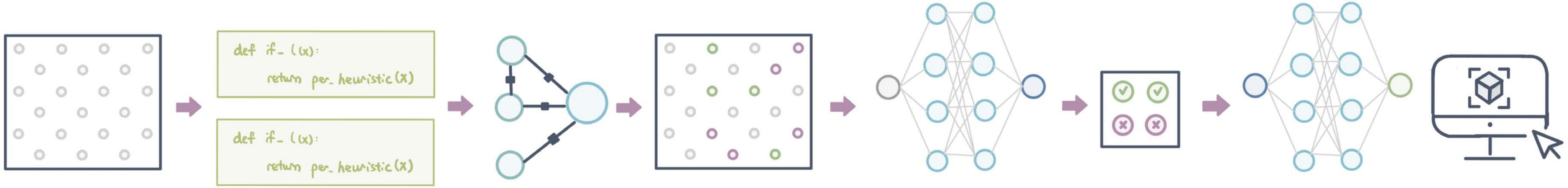
Technical Evaluation

LabelAId

User Interface Design & Implementation

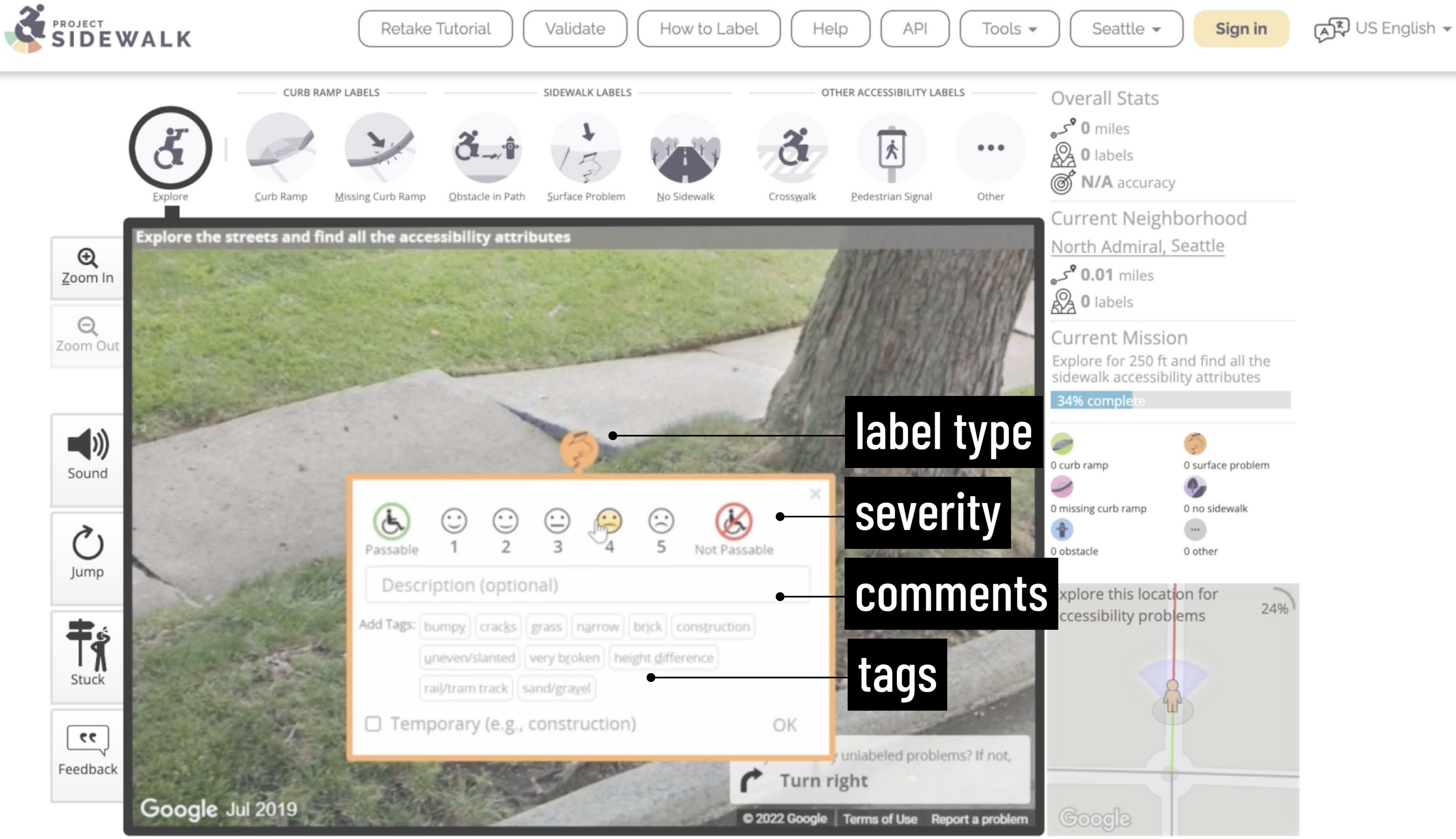
User Study

LabelAId Pipeline



PROJECT SIDEWALK

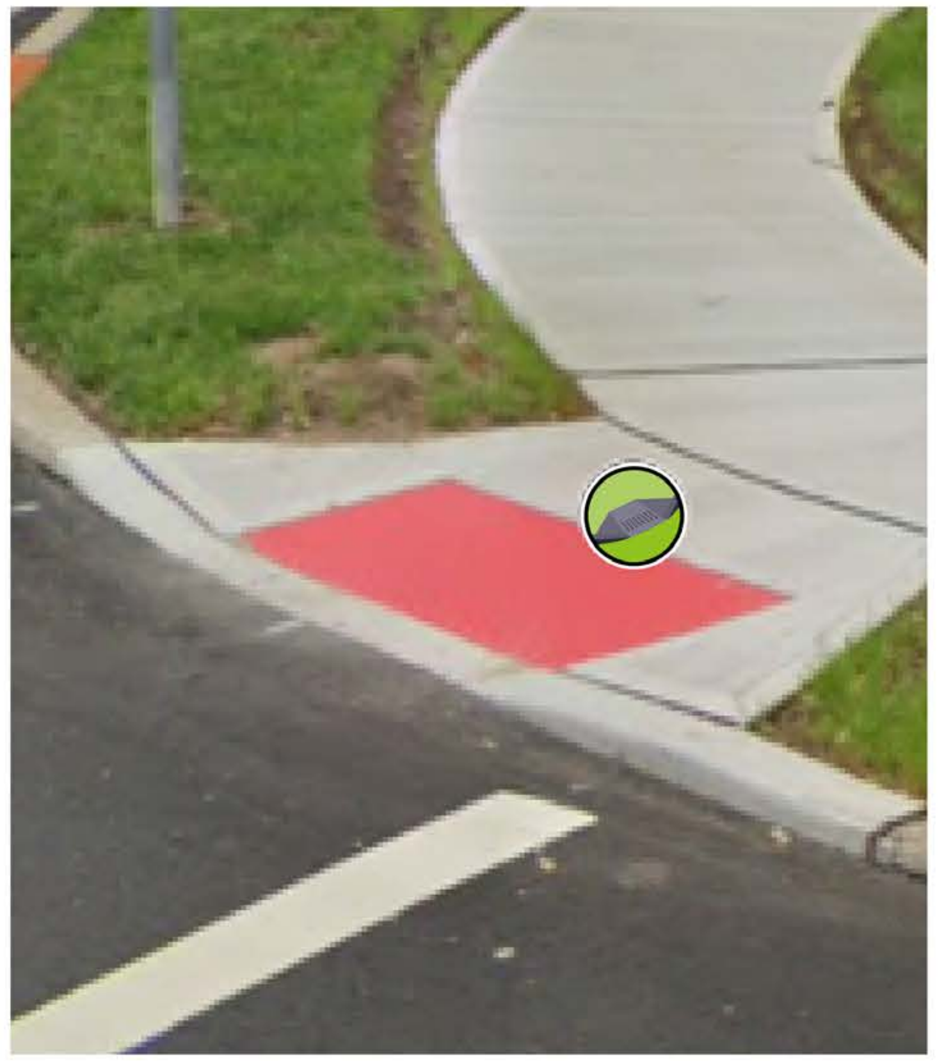
projectsidewalk.org


PROJECT SIDEWALK
Retake Tutorial
Validate
How to Label
Help
API
Tools
Seattle
Sign in
US English
CURB RAMP LABELS
SIDEWALK LABELS
OTHER ACCESSIBILITY LABELS
Explore
Curb Ramp
Missing Curb Ramp
Obstacle in Path
Surface Problem
No Sidewalk
Crosswalk
Pedestrian Signal
Other
Overall Stats
0 miles
0 labels
N/A accuracy
Current Neighborhood
North Admiral, Seattle
0.01 miles
0 labels
Current Mission
Explore for 250 ft and find all the sidewalk accessibility attributes
34% complete
0 curb ramp
0 surface problem
0 missing curb ramp
0 no sidewalk
0 obstacle
0 other
xplore this location for
ccessibility problems
24%
Zoom In
Zoom Out
Sound
Jump
Stuck
Feedback
Explore the streets and find all the accessibility attributes
Passable
1
2
3
4
5
Not Passable
Description (optional)
Add Tags:
bumpy
cracks
grass
narrow
brick
construction
uneven/slanted
very broken
height difference
rail/tram track
sand/gravel
Temporary (e.g., construction)
OK
unlabeled problems? If not,
Turn right
Google Jul 2019
© 2022 Google
Terms of Use
Report a problem
label type
severity
comments
tags

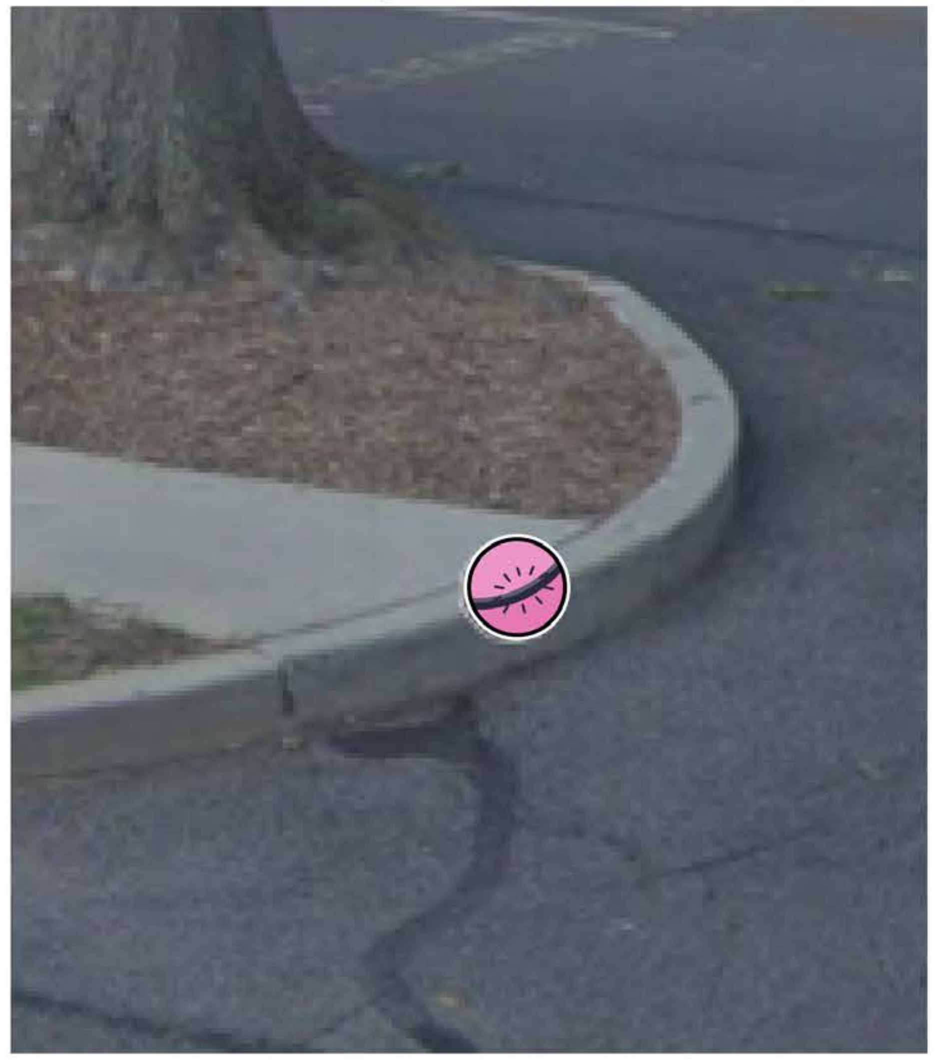
Project Sidewalk

5 Major Label Types

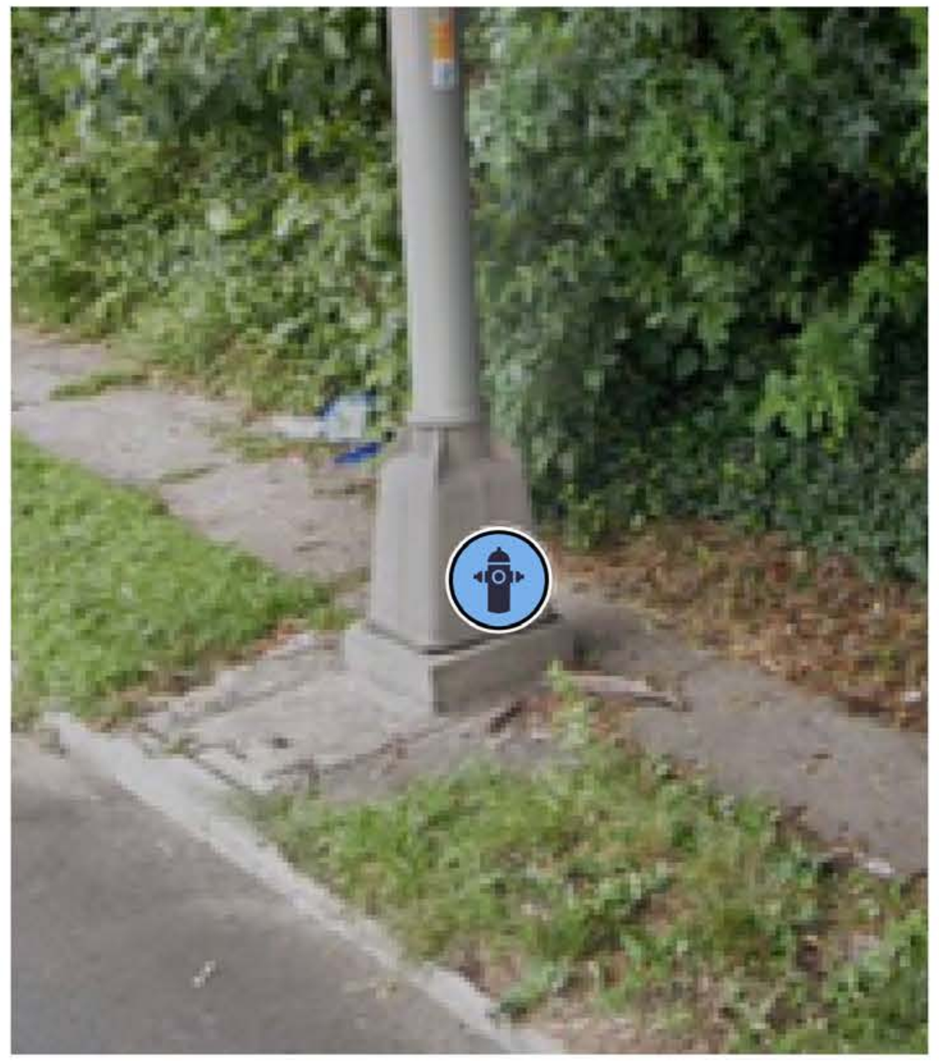
Curb Ramps



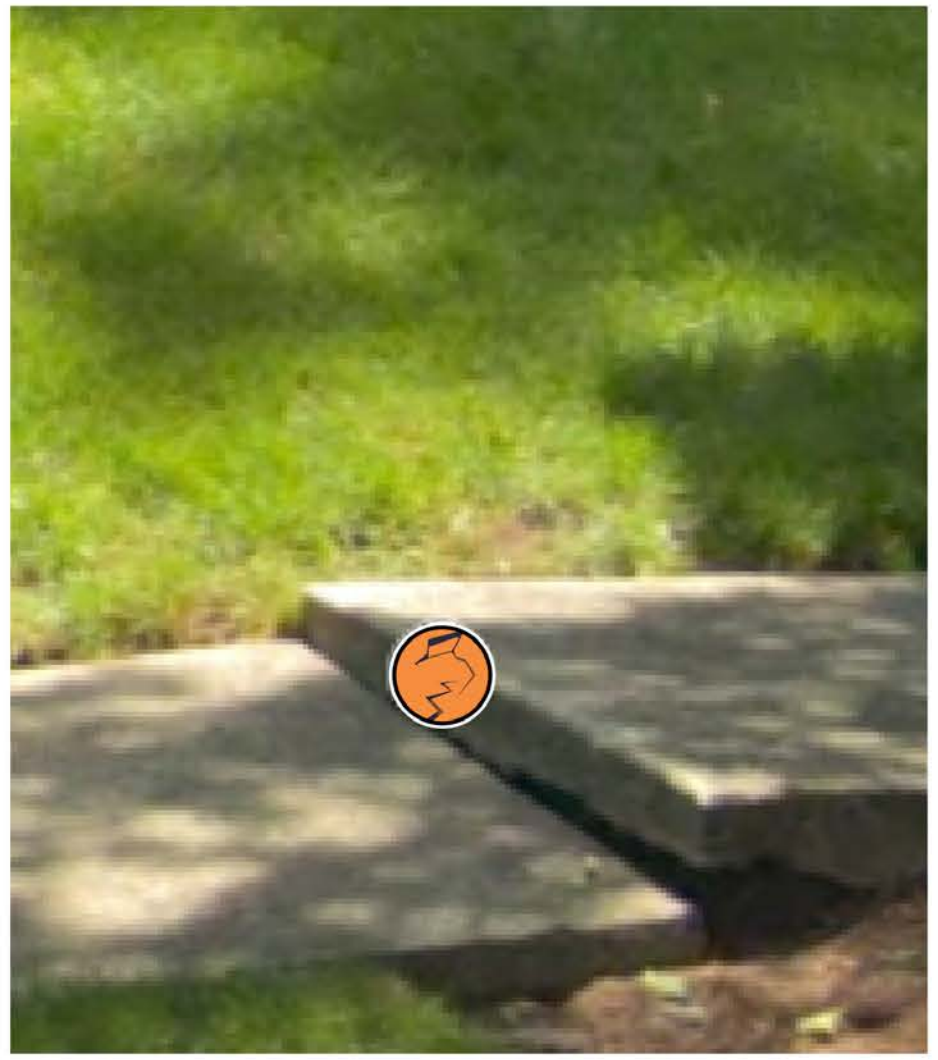
Missing Curb Ramps



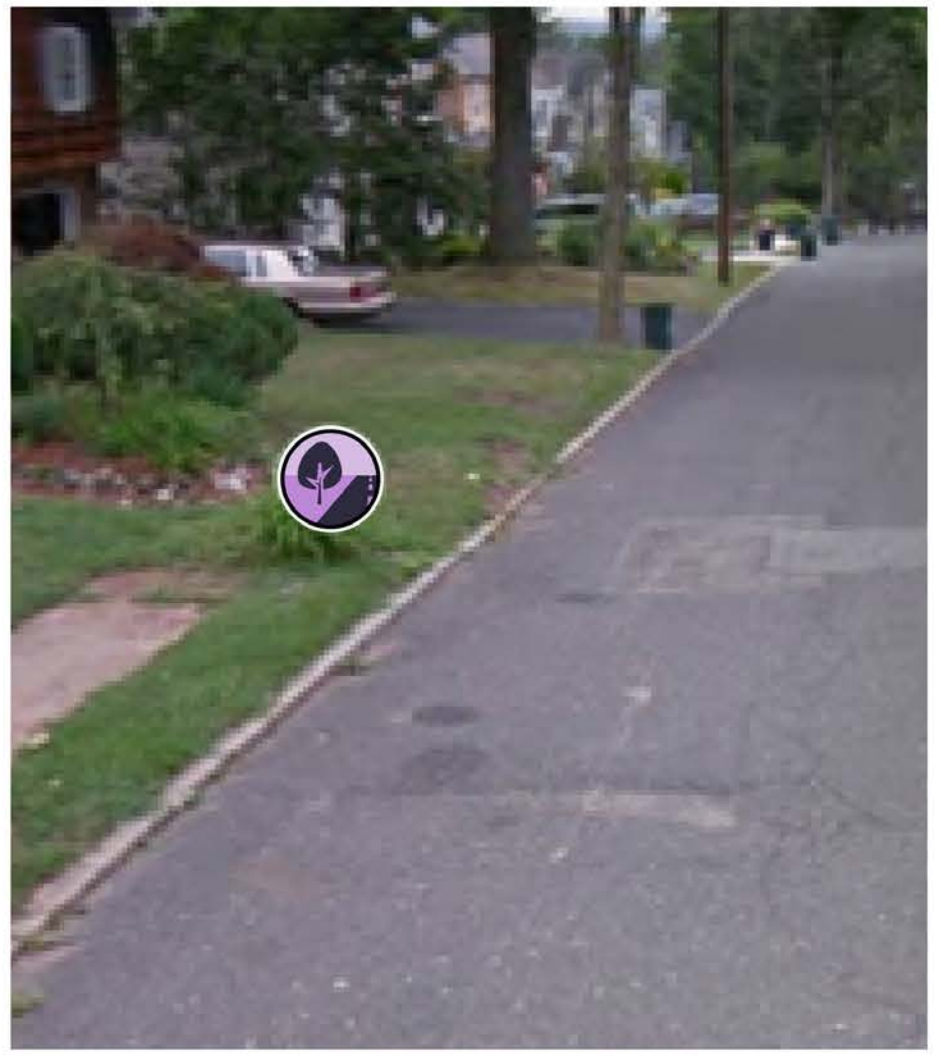
Obstacles



Surface Problems

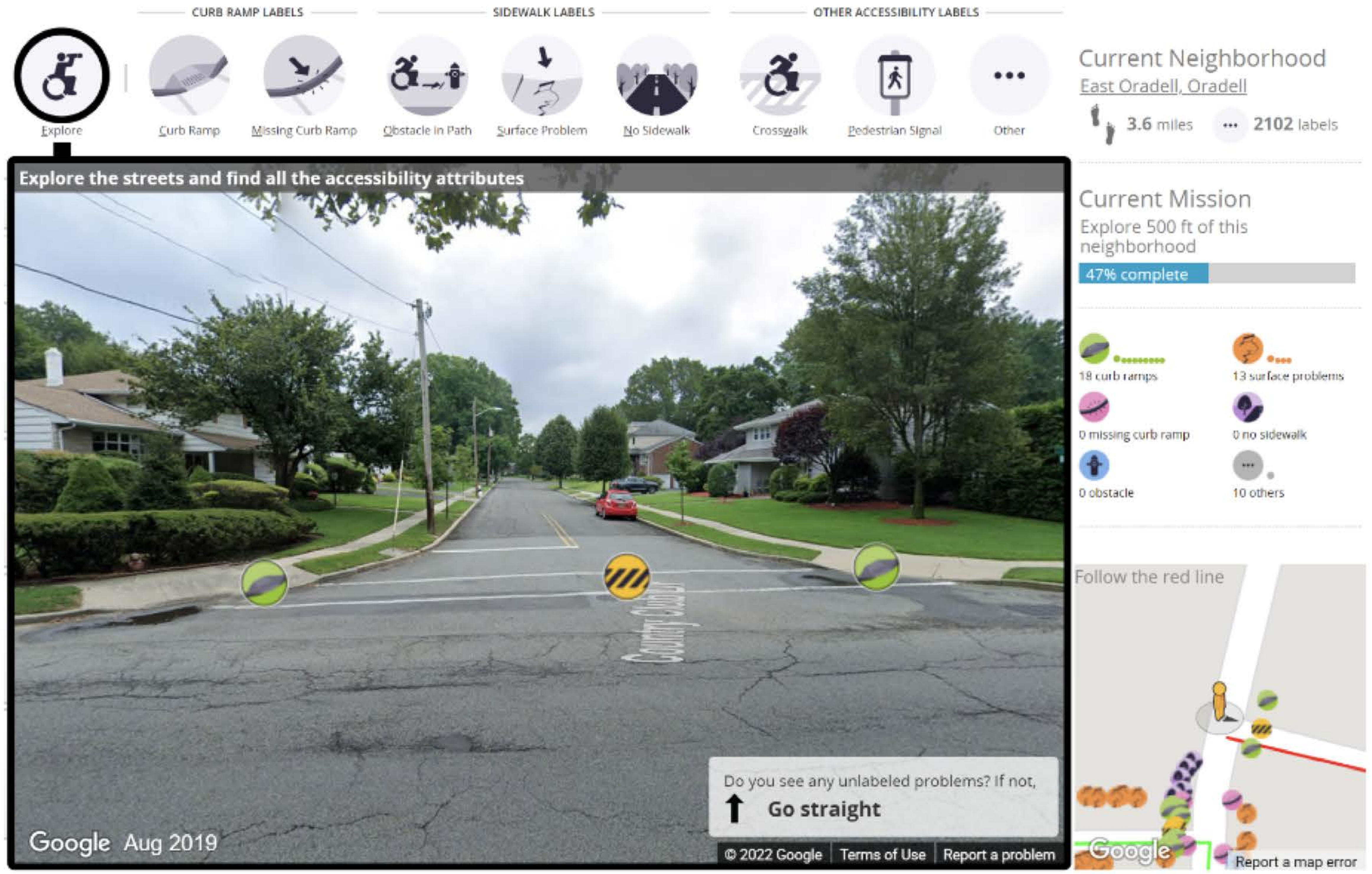


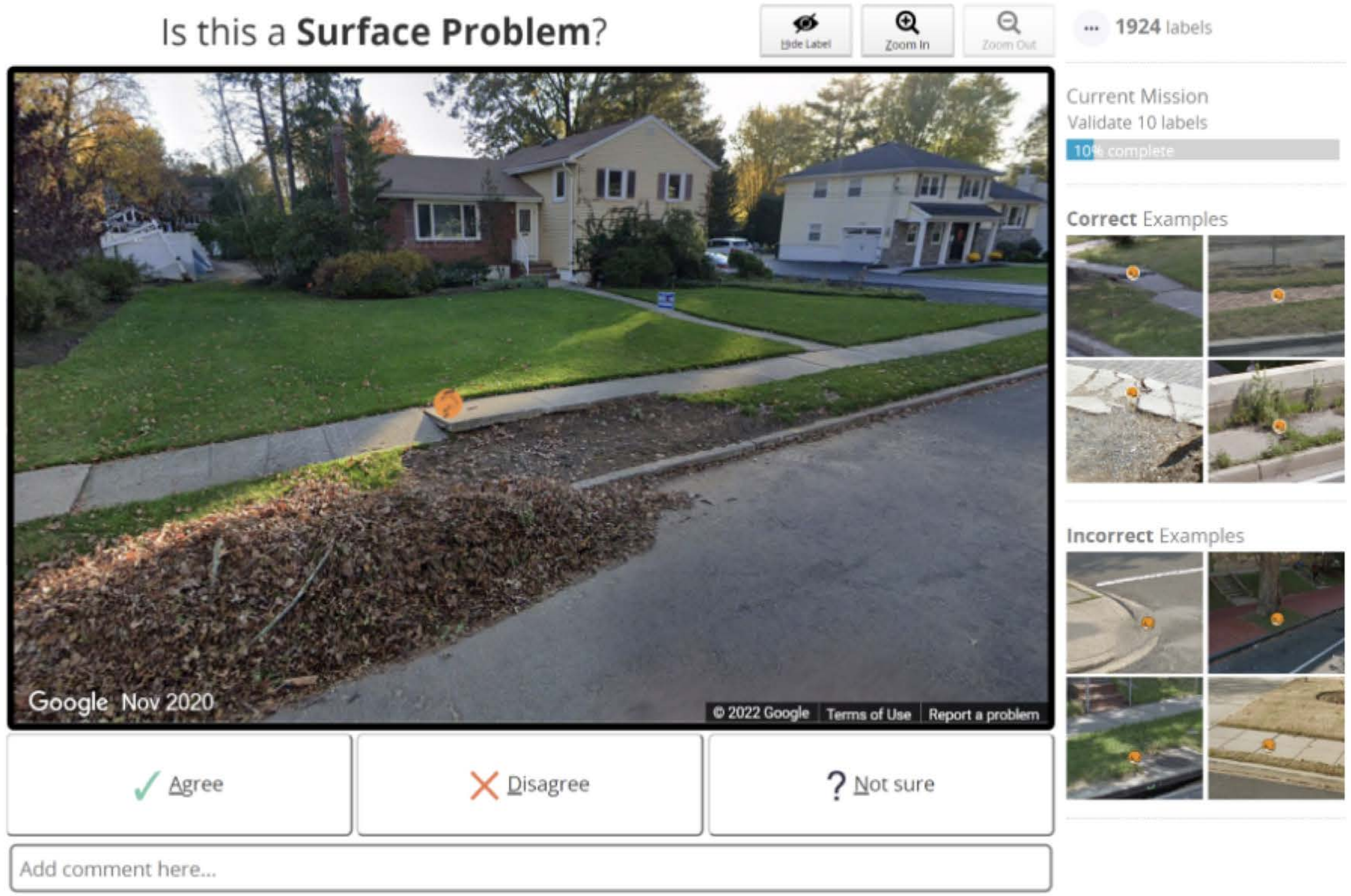
Missing Sidewalk



Project Sidewalk

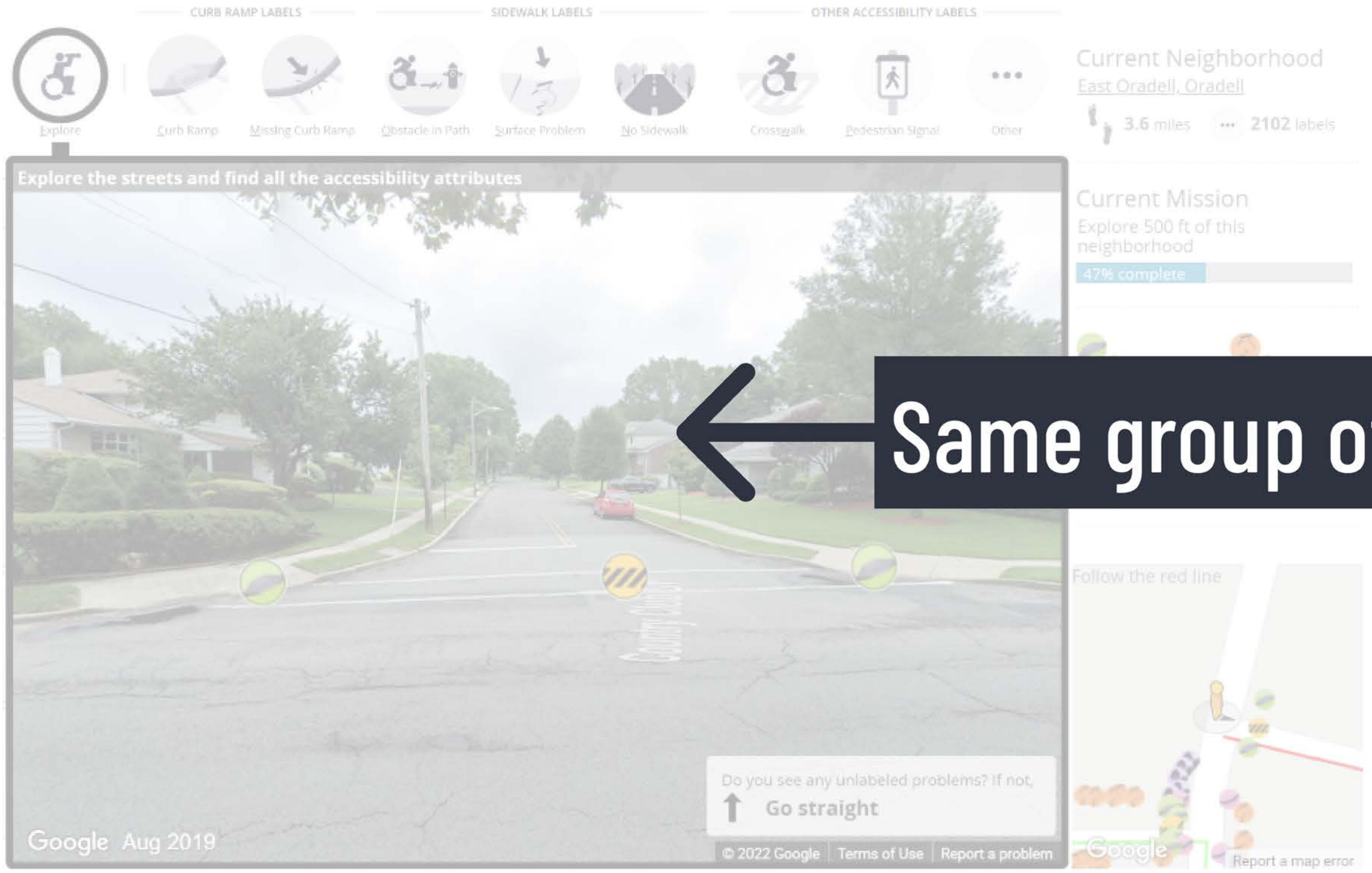
Project Sidewalk Quality Control



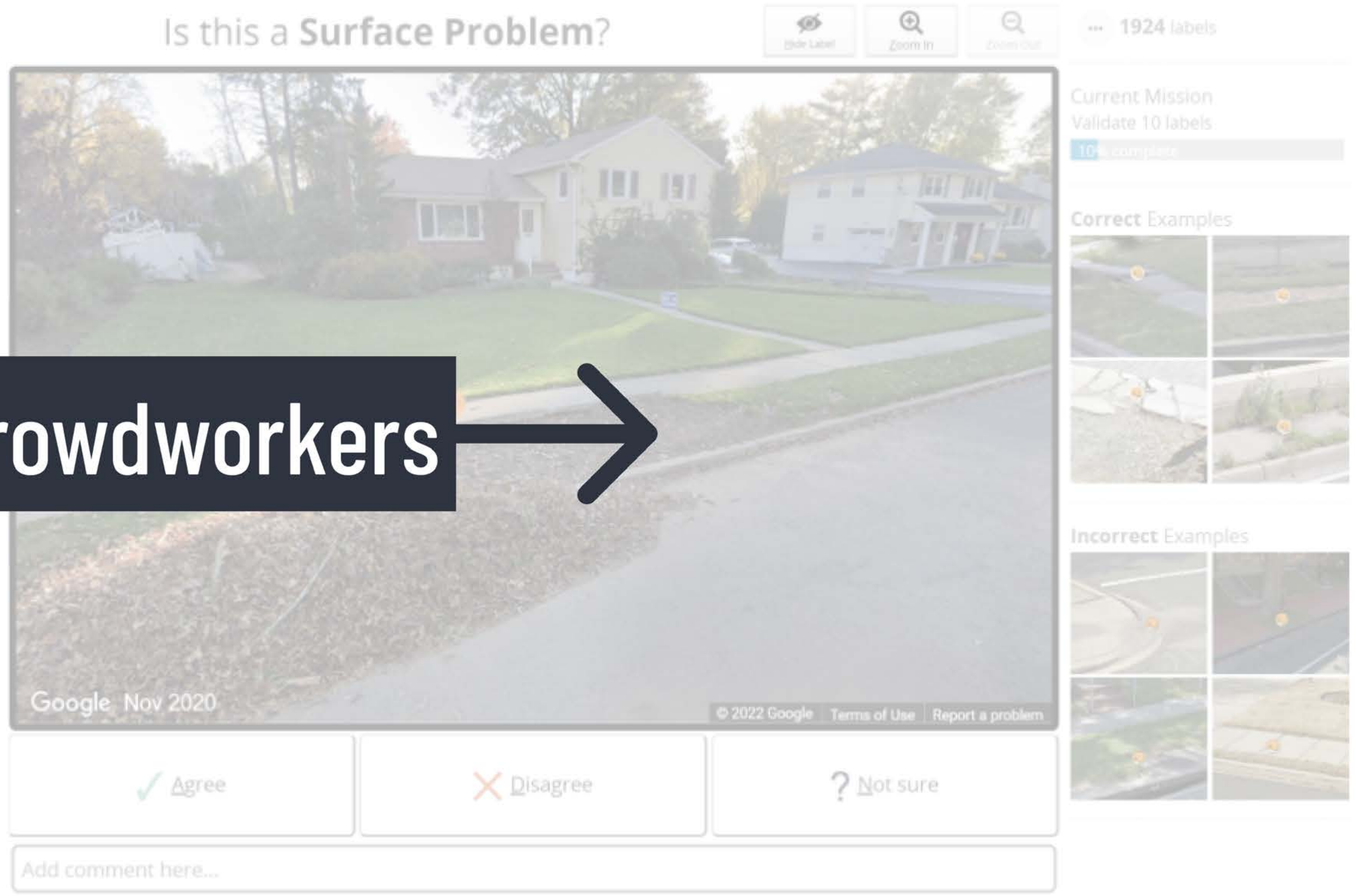


1 Find, label, and assess sidewalks

2 Validate & correct other labels



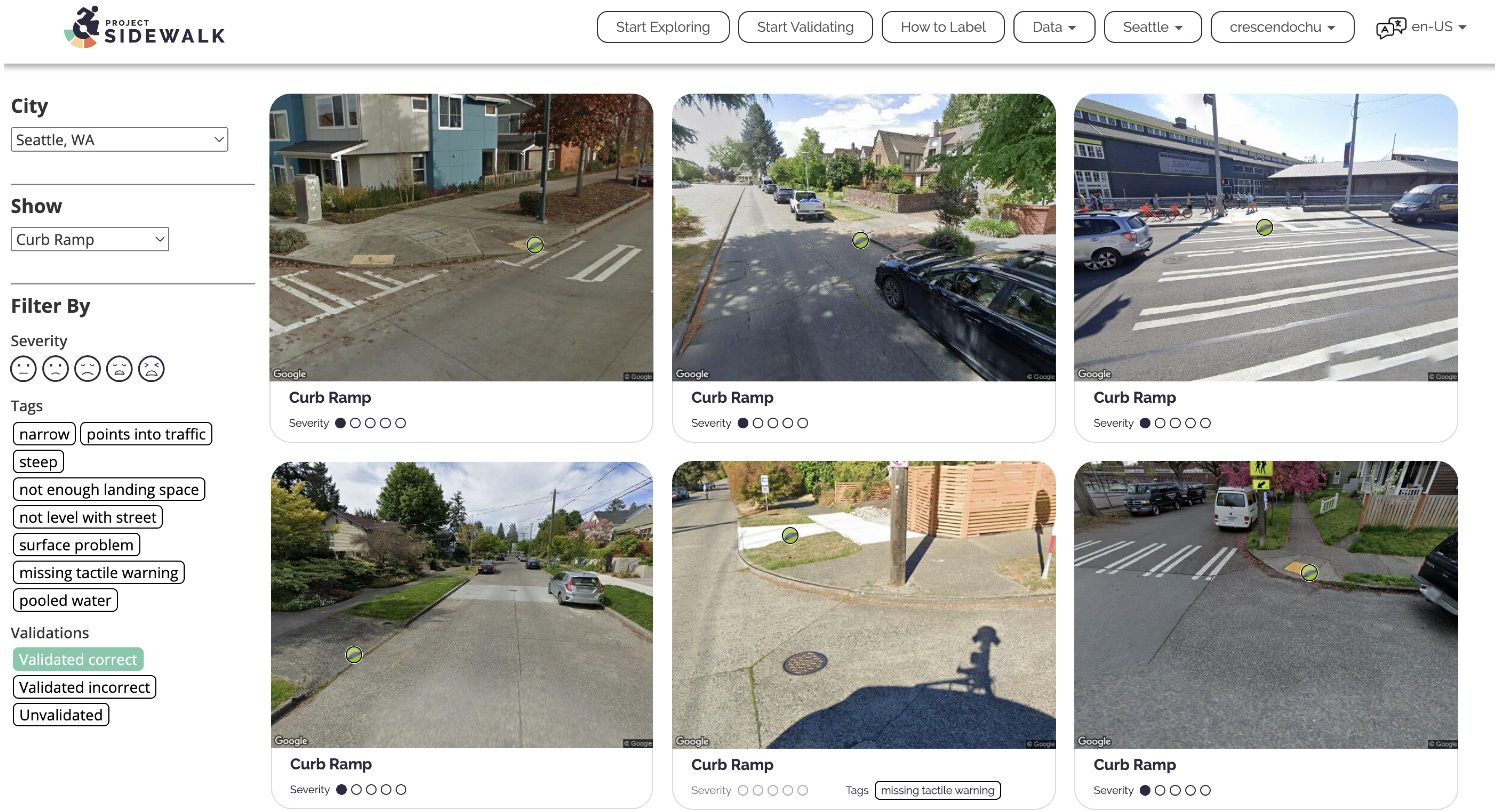
Same group of crowdworkers



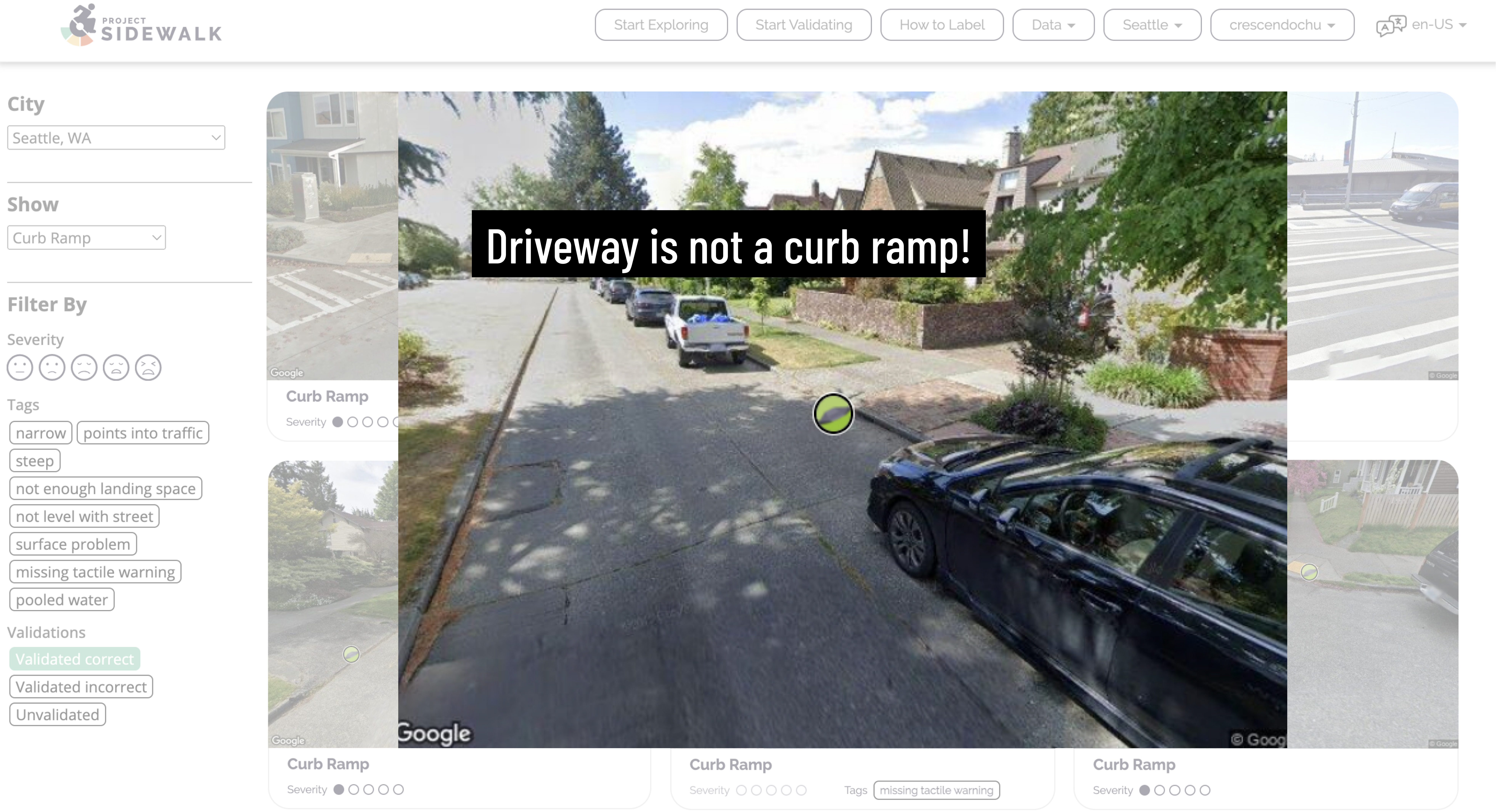
1 Find, label, and assess sidewalks

2 Validate & correct other labels

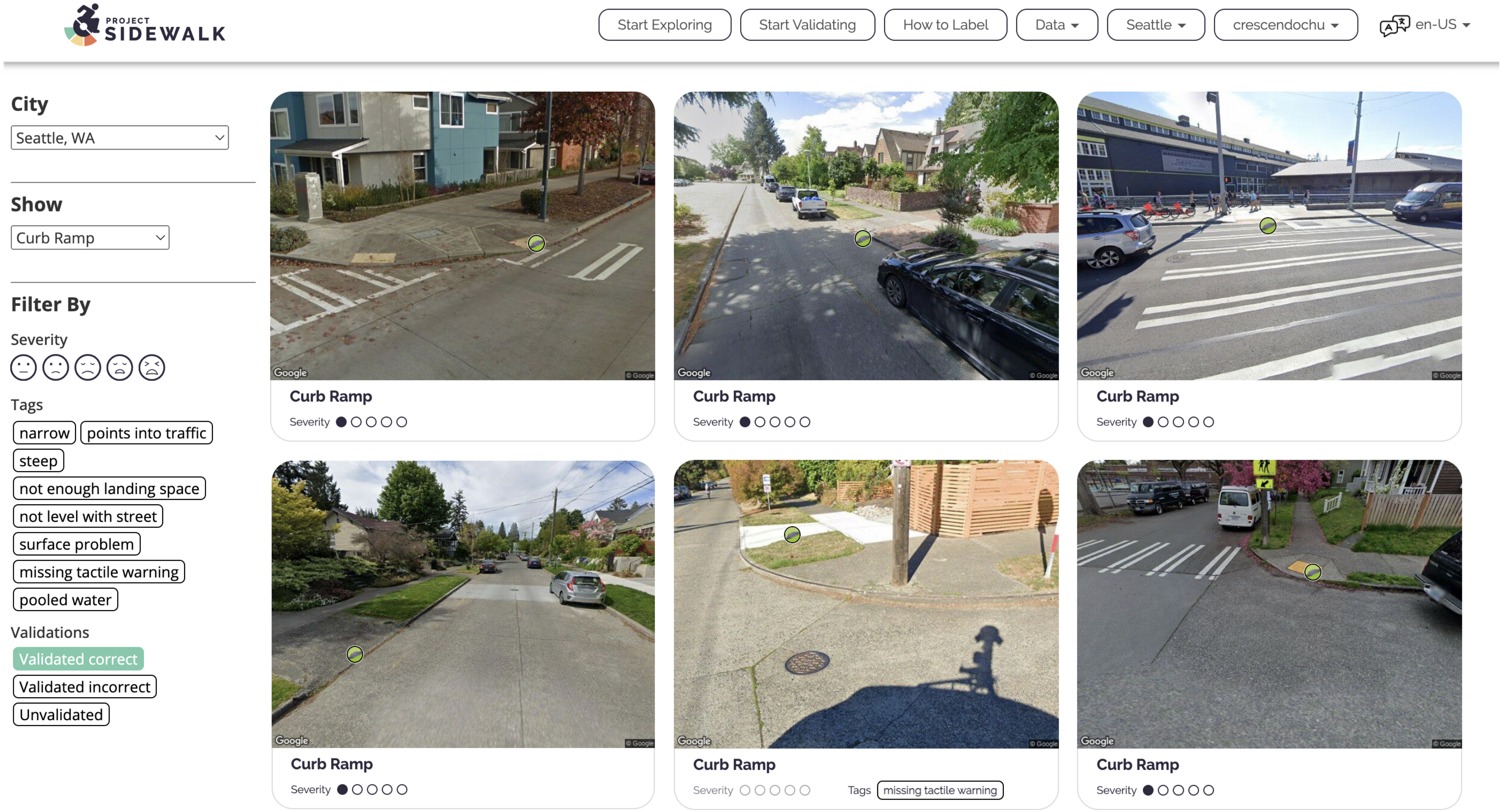
Project Sidewalk Quality Control



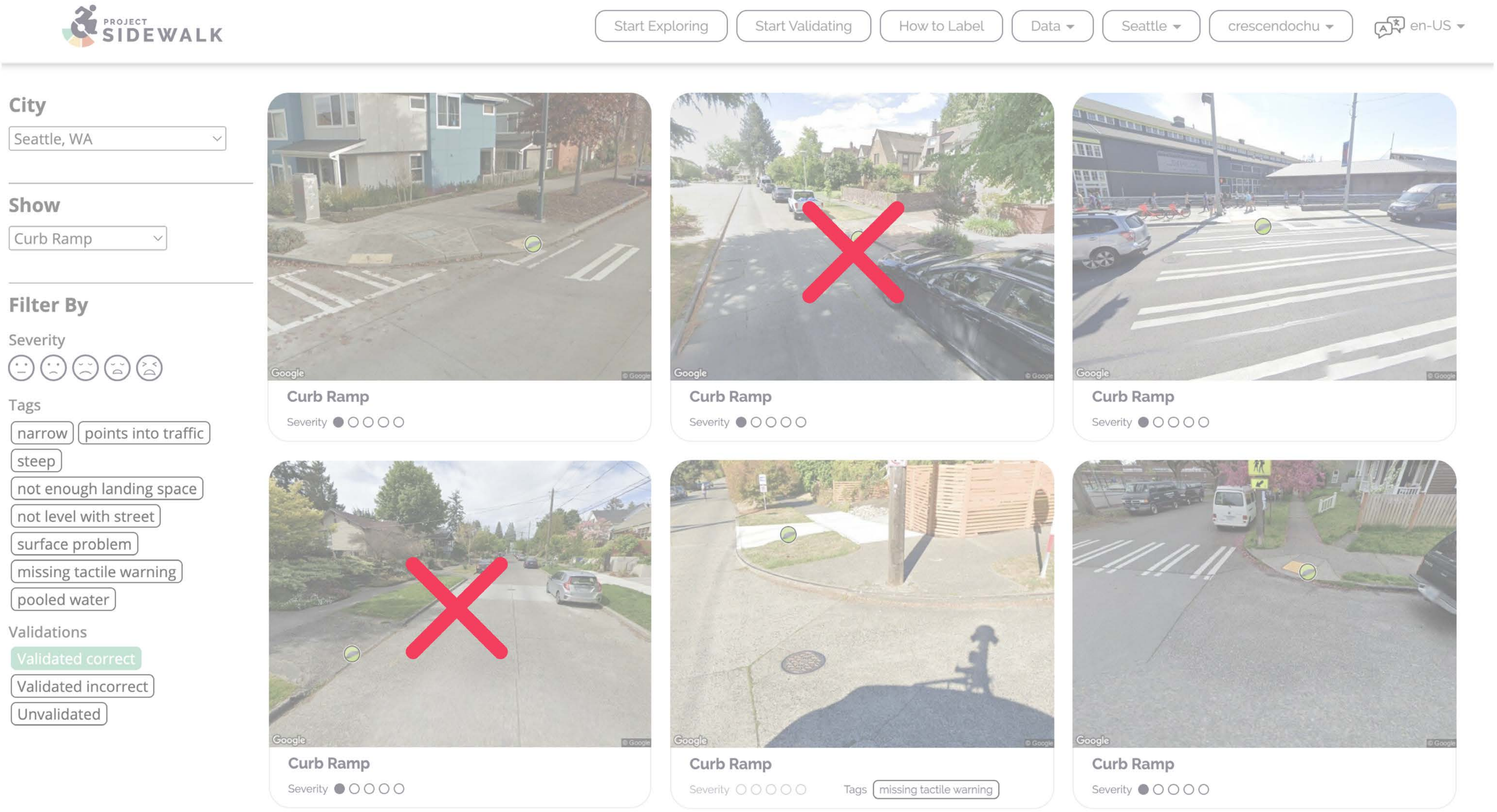
Project Sidewalk Quality Control



Project Sidewalk Quality Control



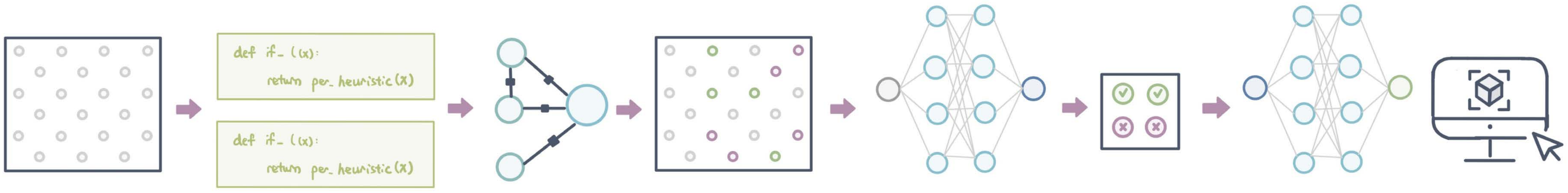
Project Sidewalk Quality Control



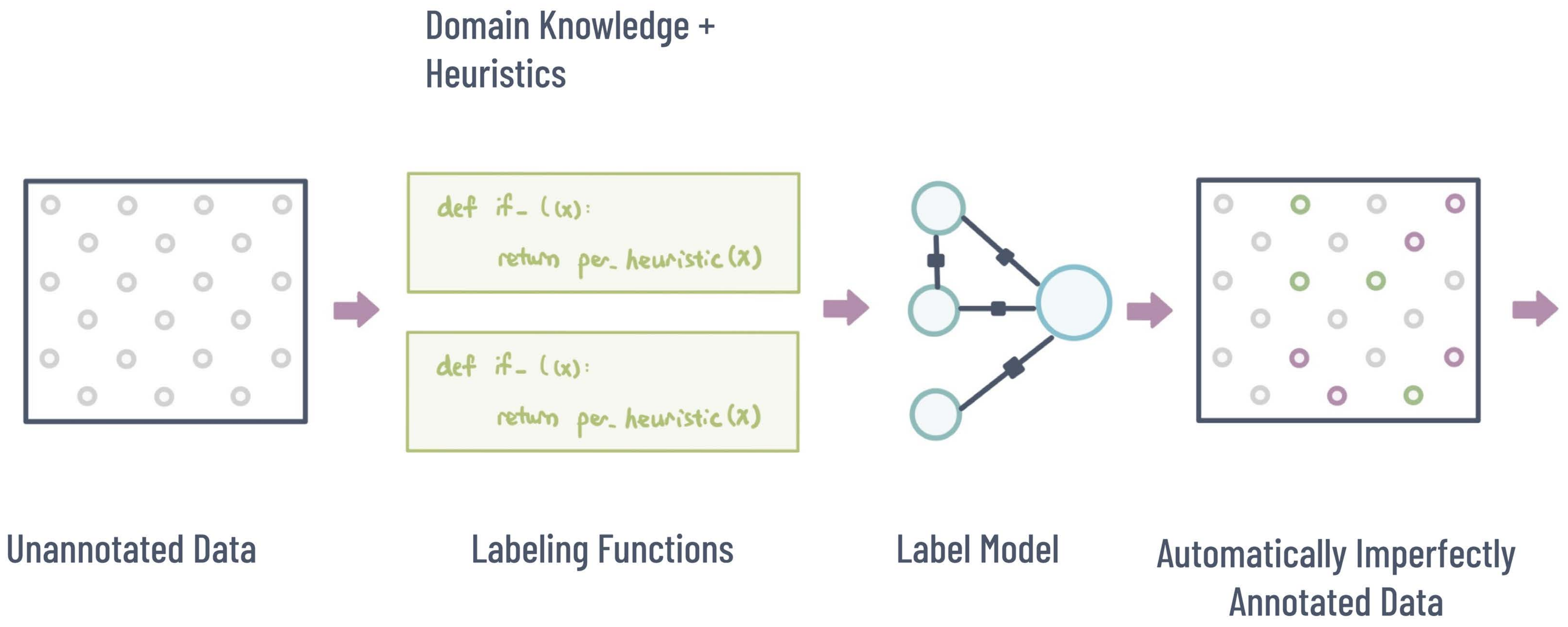
Project Sidewalk Quality Control

Data noise at neighborhood scale

Adapting LabelAId to Project Sidewalk

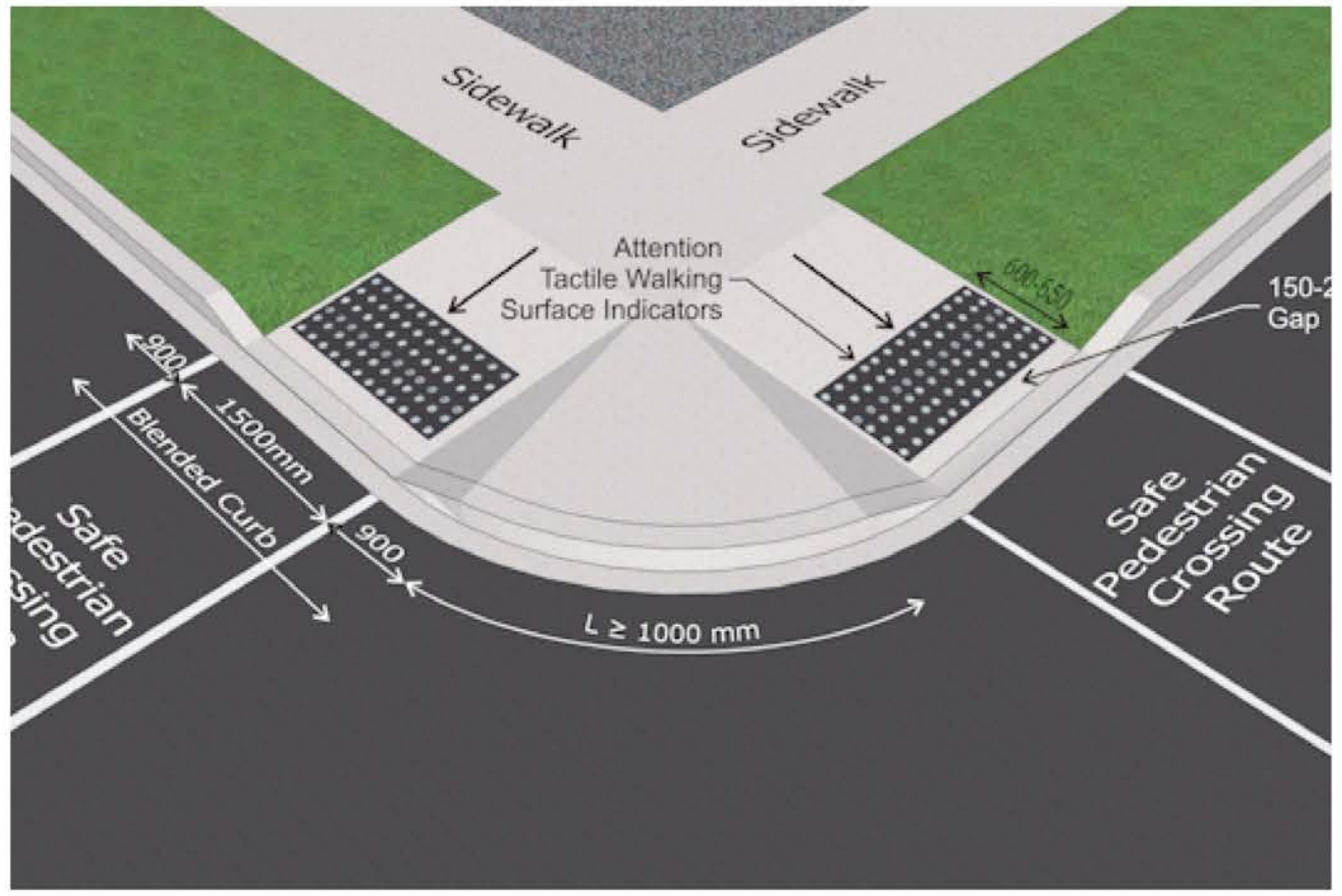


Programmatic Weak Supervision



Labeling Function Example

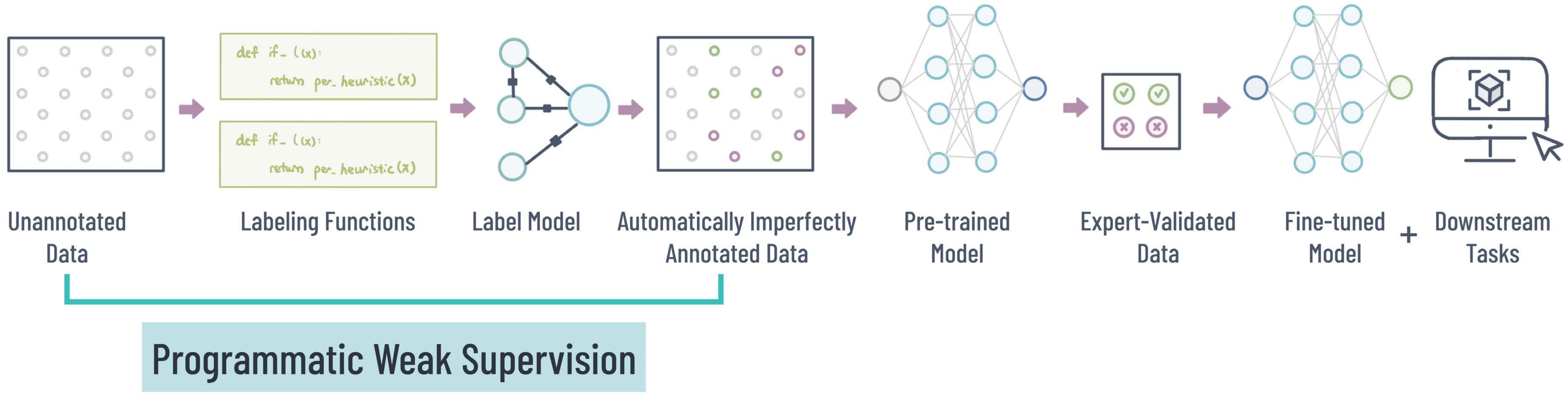
Domain Knowledge



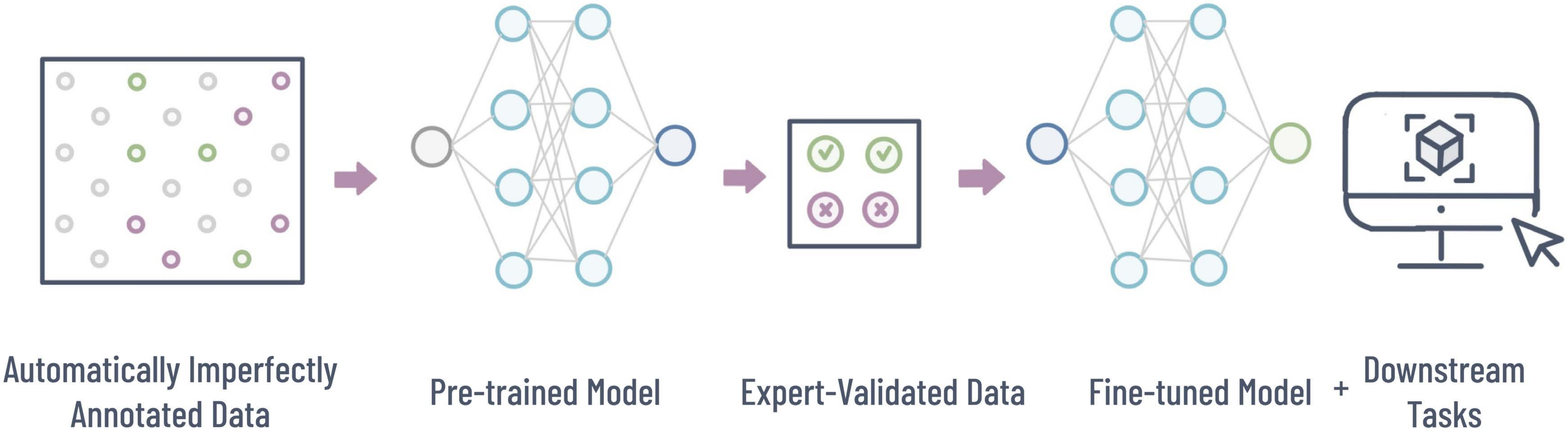
Set Of Labeling Functions

Planning Guidelines	Crowdsource Nature	User Behaviors	Label Characteristics
Distance to urban infrastructure	In proximity with other user's labels	Optional inputs & labeling zoom level	Severity rating

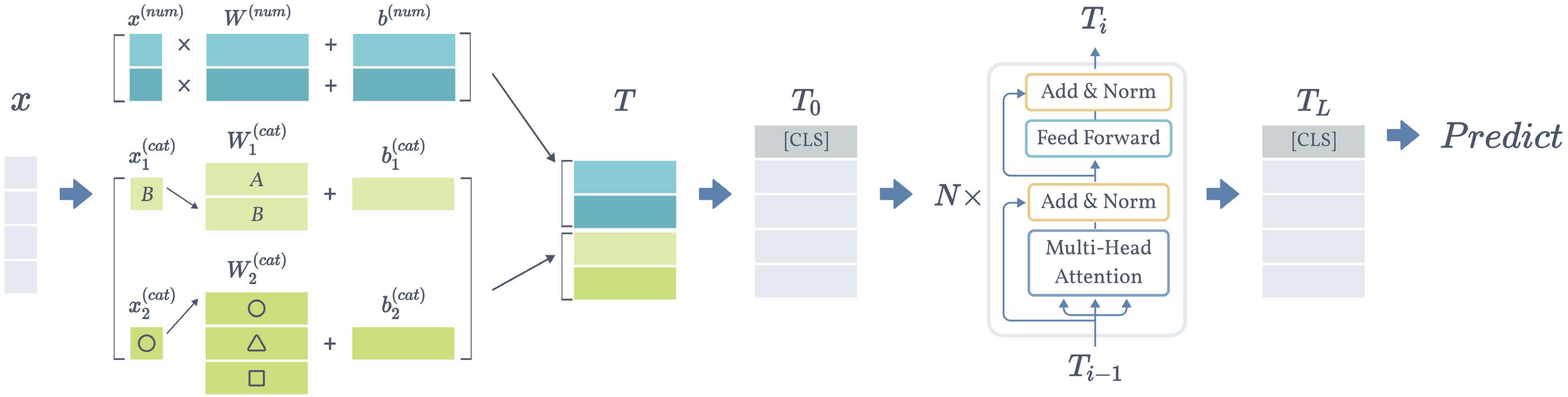
LabelAId Pipeline

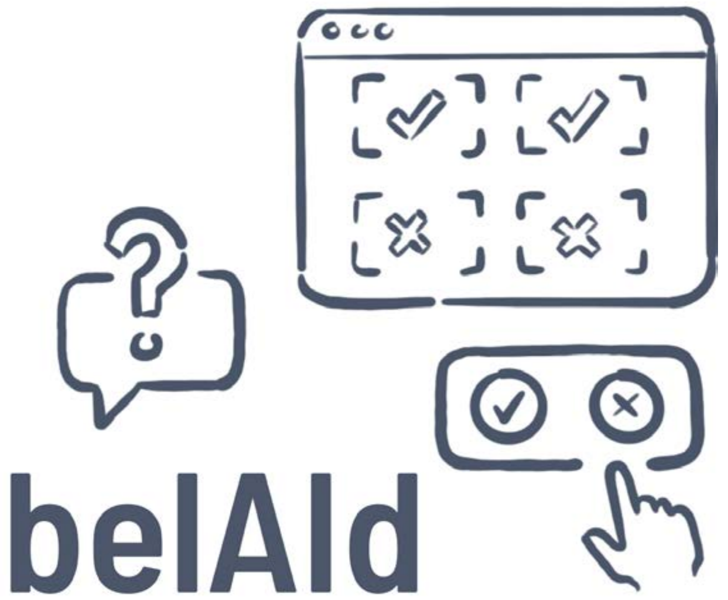


Pre-Training & Fine-Tuning



FT-Transformer-Based Model Architecture





LabelAId

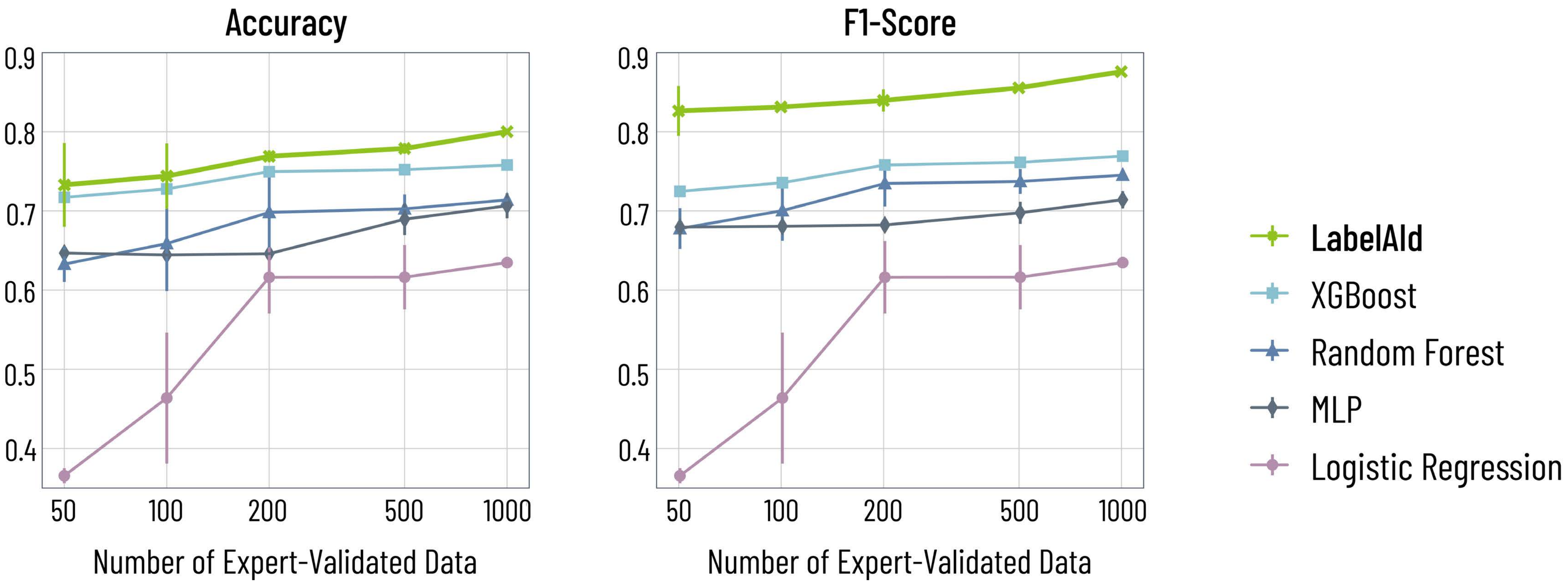
Machine Learning Framework

Technical Evaluation

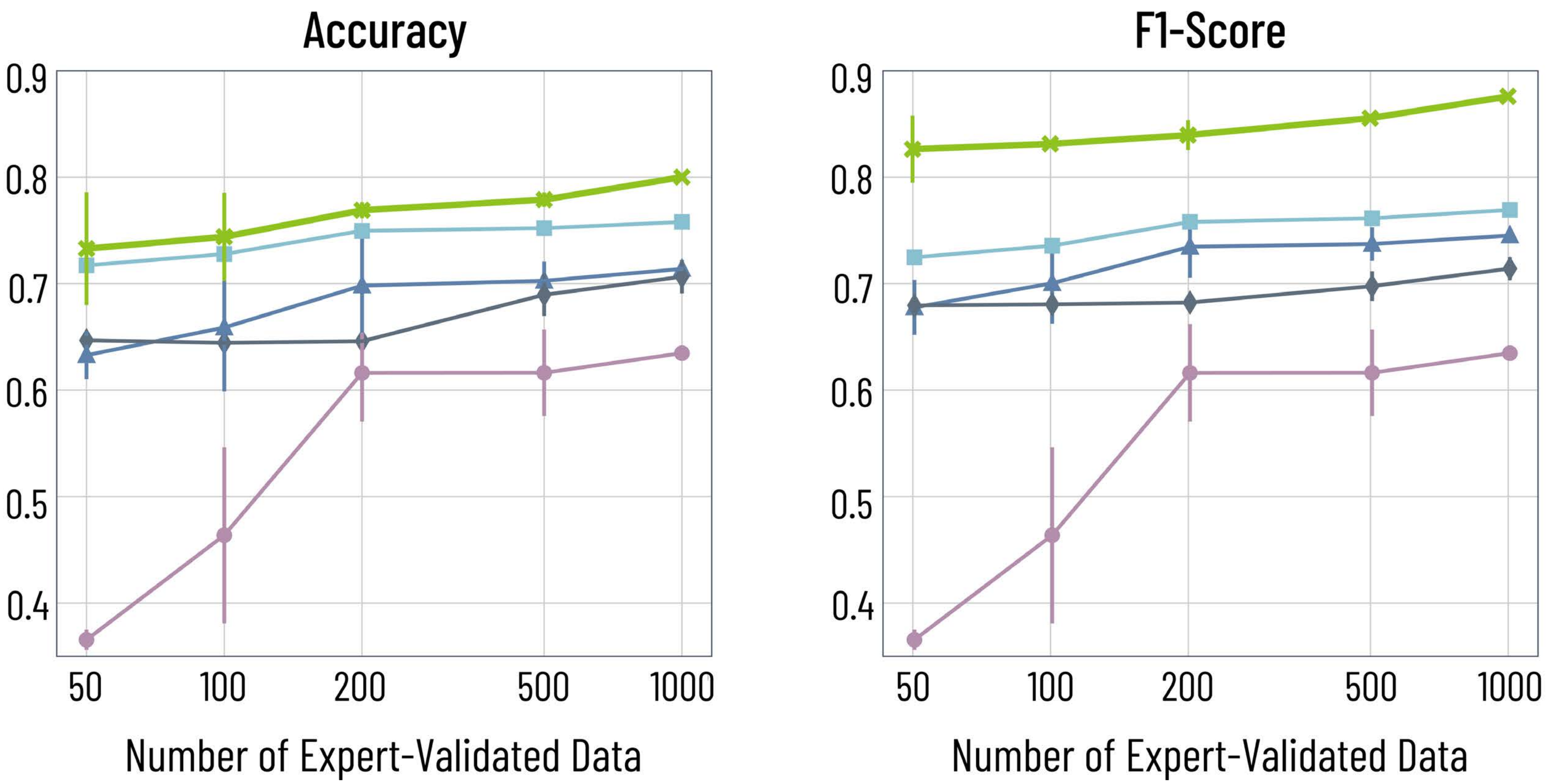
User Interface Design & Implementation

User Study

Technical Evaluation



Technical Evaluation



LabelAId improves accuracy by up to **37%** with just **50** downstream samples

LabelAId

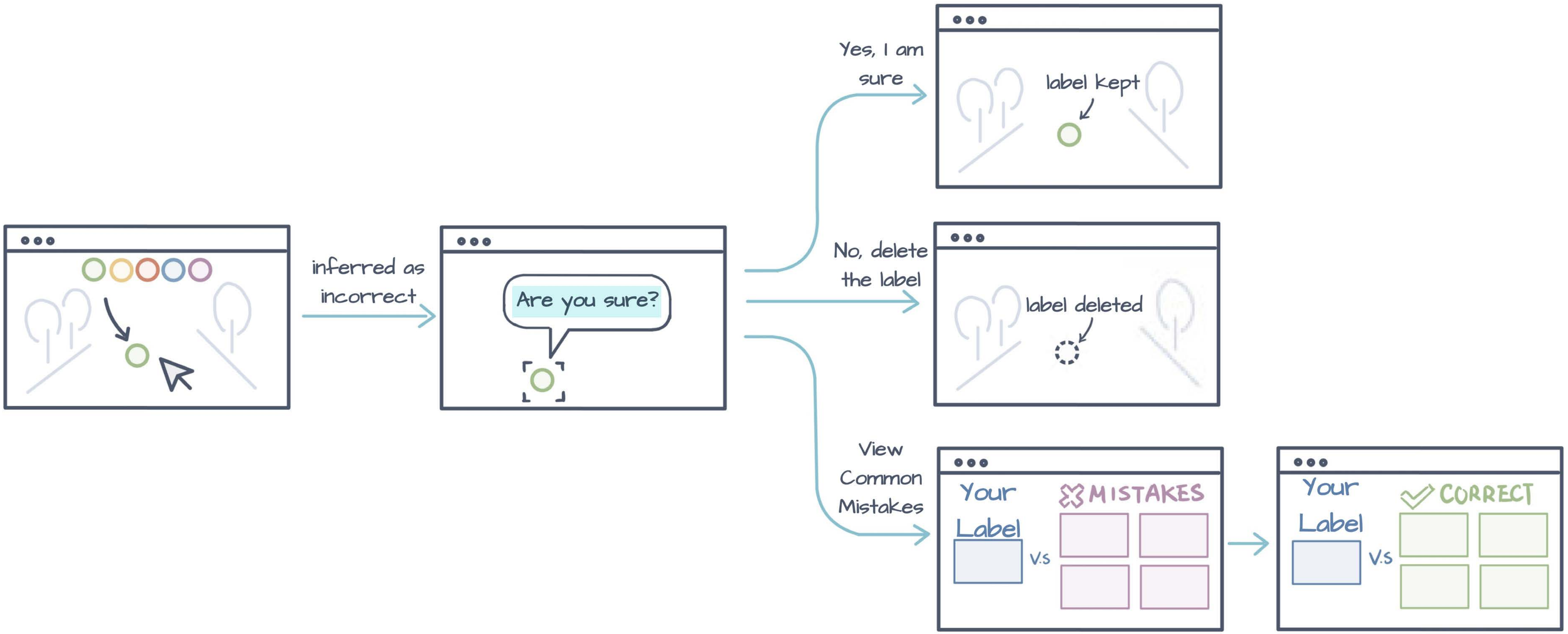
Machine Learning Framework

Technical Evaluation

User Interface Design & Implementation

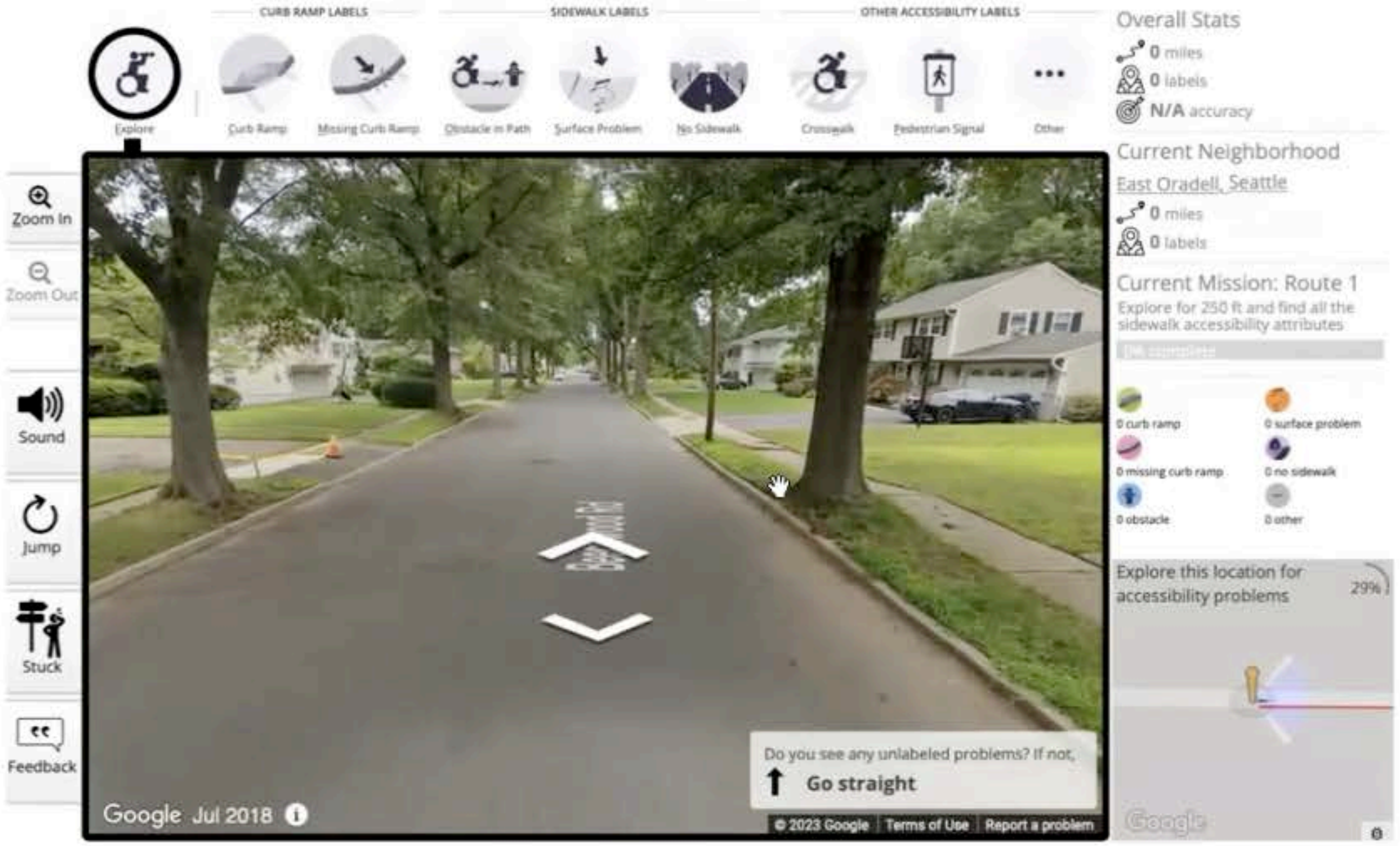
User Study

User Flow

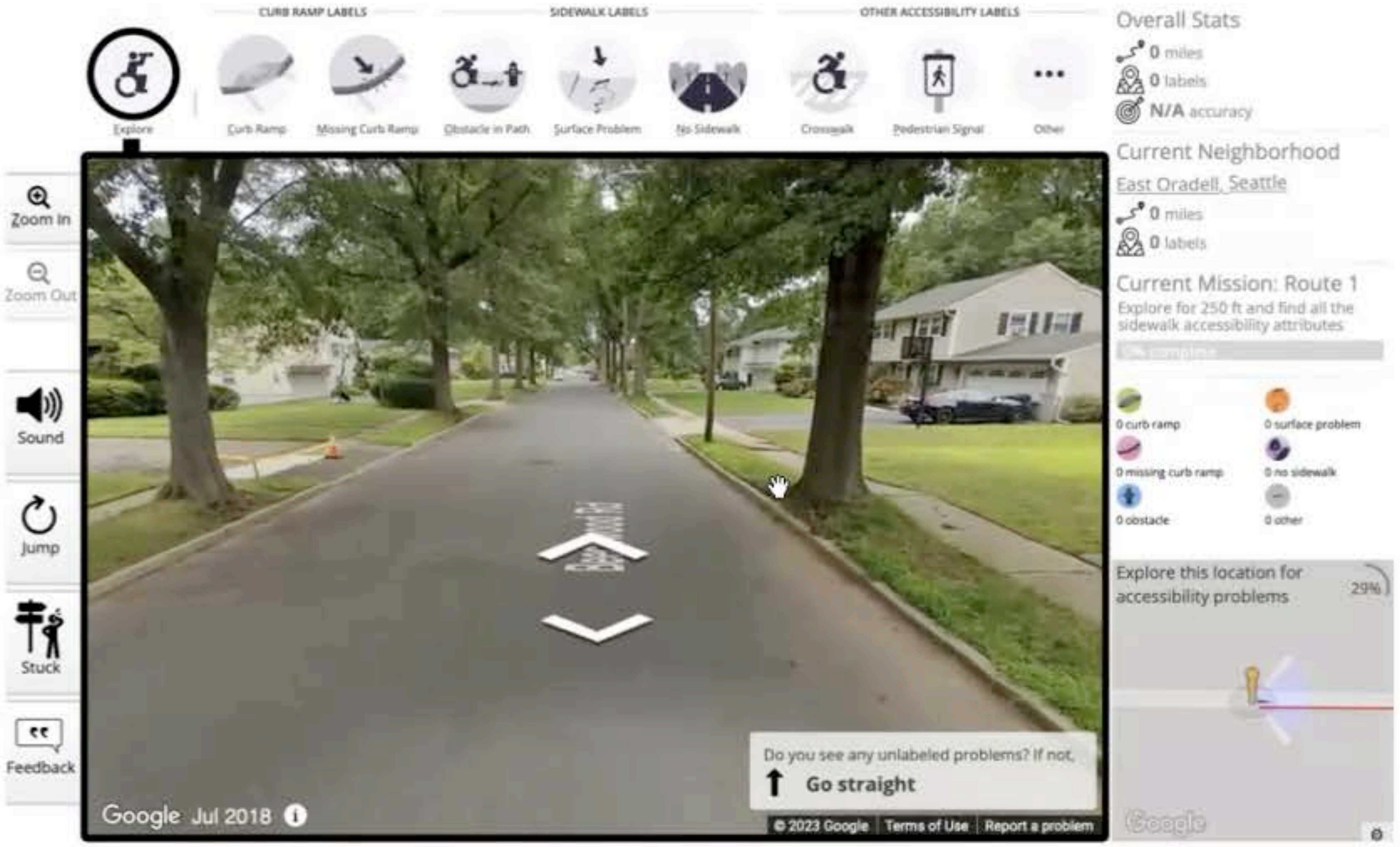


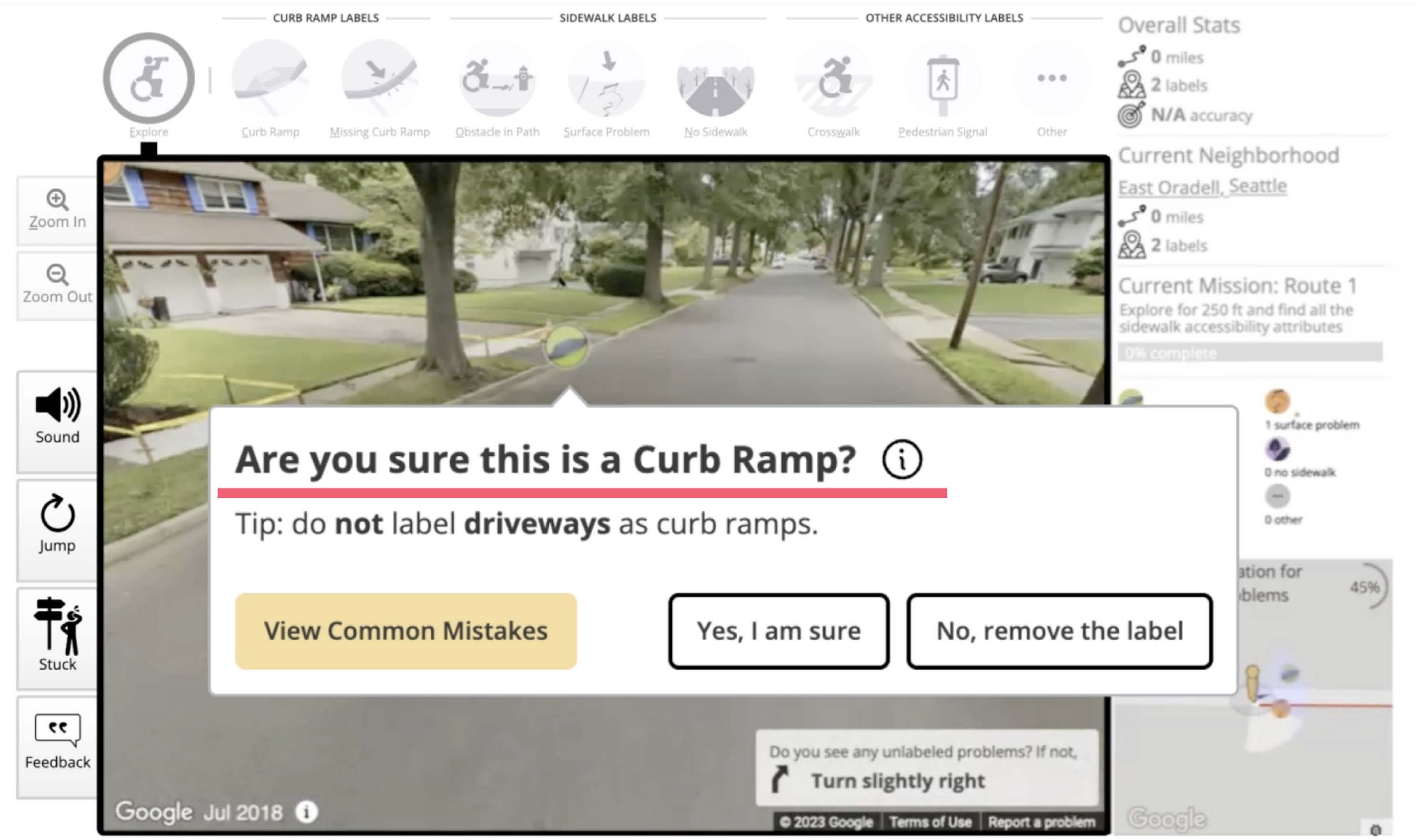
inferred as incorrect
Are you sure?
Yes, I am sure
label kept
No, delete the label
label deleted
View Common Mistakes
Your Label
V.s
MISTAKES
Your Label
V.s
CORRECT

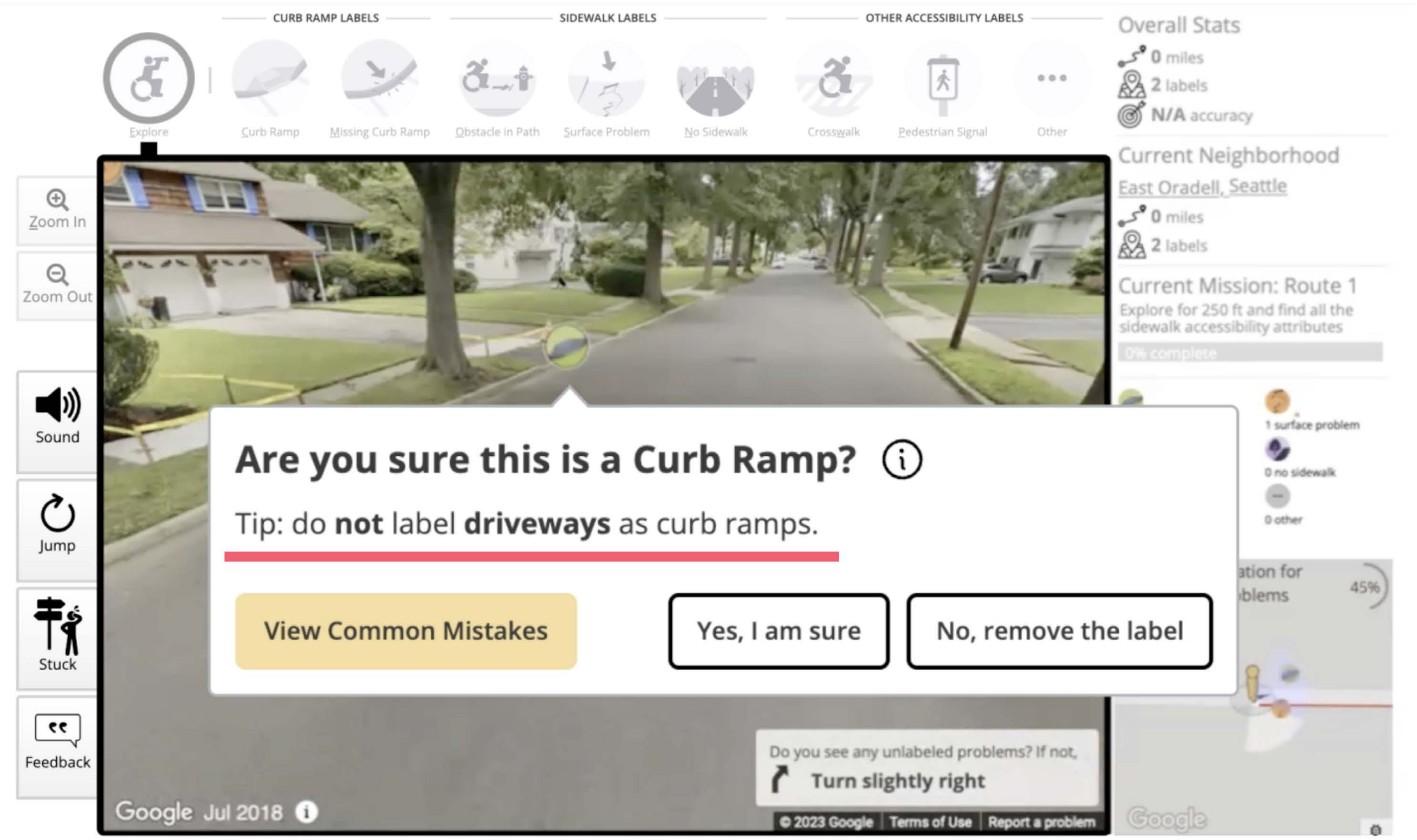
Original Project Sidewalk

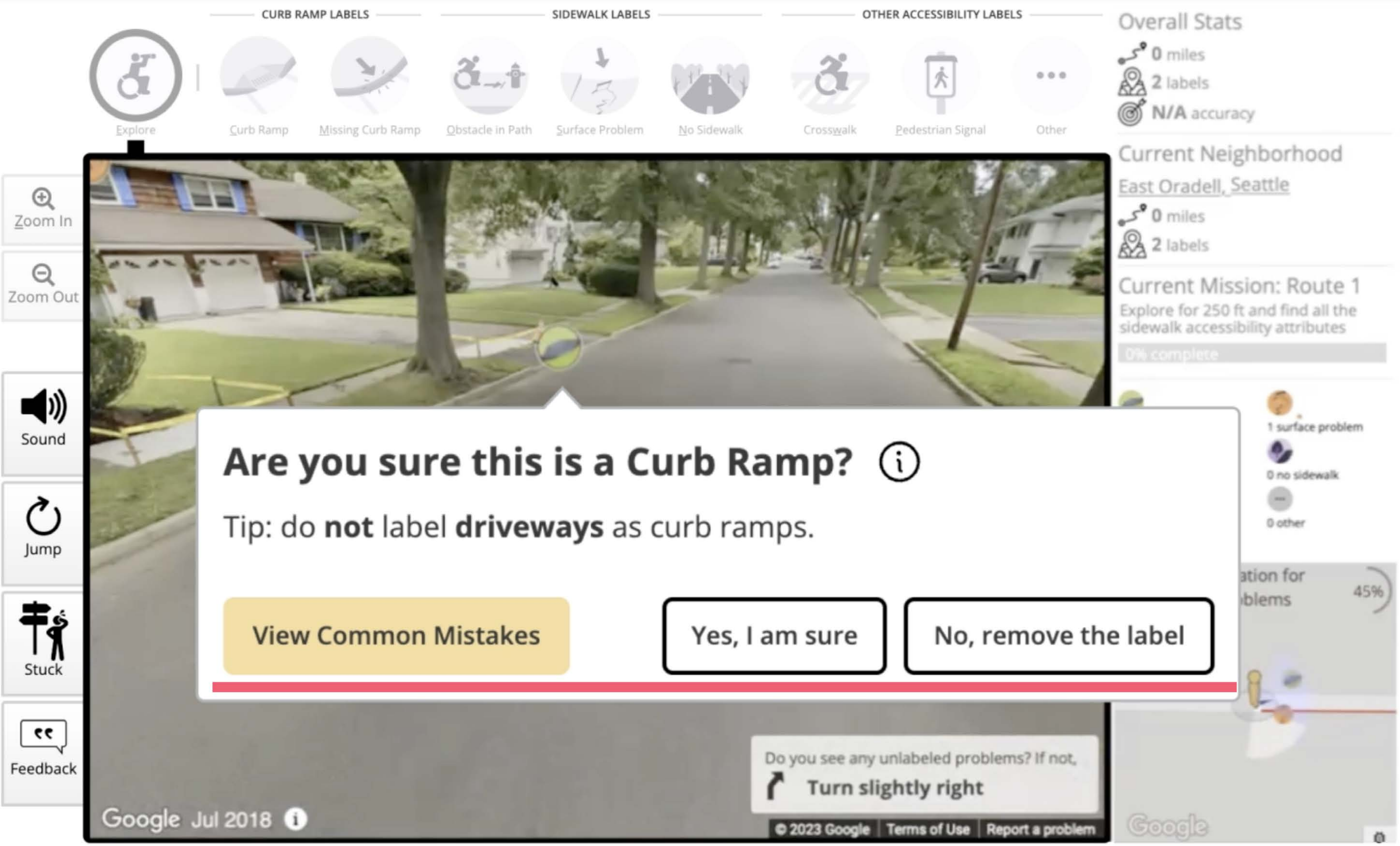


Project Sidewalk with LabelAId




CURB RAMP LABELS
SIDEWALK LABELS
OTHER ACCESSIBILITY LABELS
Explore
Curb Ramp
Missing Curb Ramp
Obstacle in Path
Surface Problem
No Sidewalk
Crosswalk
Pedestrian Signal
Other
Overall Stats
0 miles
2 labels
N/A accuracy
Current Neighborhood
East Oradell, Seattle
0 miles
2 labels
Current Mission: Route 1
Explore for 250 ft and find all the sidewalk accessibility attributes
Zoom In
Zoom Out
Sound
Jump
Stuck
Feedback
Are you sure this is a Curb Ramp?
Tip: do not label driveways as curb ramps.
View Common Mistakes
Yes, I am sure
No, remove the label
1 surface problem
0 no sidewalk
0 other
45%
Do you see any unlabeled problems? If not,
Turn slightly right
Google Jul 2018
© 2023 Google
Terms of Use
Report a problem


CURB RAMP LABELS
SIDEWALK LABELS
OTHER ACCESSIBILITY LABELS
Explore
Curb Ramp
Missing Curb Ramp
Obstacle in Path
Surface Problem
No Sidewalk
Crosswalk
Pedestrian Signal
Other
Overall Stats
0 miles
2 labels
N/A accuracy
Current Neighborhood
East Oradell, Seattle
0 miles
2 labels
Current Mission: Route 1
Explore for 250 ft and find all the sidewalk accessibility attributes
Zoom In
Zoom Out
Sound
Jump
Stuck
Feedback
Are you sure this is a Curb Ramp?
Tip: do **not** label **driveways** as curb ramps.
View Common Mistakes
Yes, I am sure
No, remove the label
1 surface problem
0 no sidewalk
0 other
45%
Do you see any unlabeled problems? If not,
Turn slightly right
Google Jul 2018
© 2023 Google | Terms of Use | Report a problem


CURB RAMP LABELS
SIDEWALK LABELS
OTHER ACCESSIBILITY LABELS
Explore
Curb Ramp
Missing Curb Ramp
Obstacle in Path
Surface Problem
No Sidewalk
Crosswalk
Pedestrian Signal
Other
Overall Stats
0 miles
2 labels
N/A accuracy
Current Neighborhood
East Oradell, Seattle
0 miles
2 labels
Current Mission: Route 1
Explore for 250 ft and find all the sidewalk accessibility attributes
Zoom In
Zoom Out
Sound
Jump
Stuck
Feedback
Are you sure this is a Curb Ramp?
Tip: do not label driveways as curb ramps.
View Common Mistakes
Yes, I am sure
No, remove the label
1 surface problem
0 no sidewalk
0 other
45%
Do you see any unlabeled problems? If not,
Turn slightly right
Google Jul 2018
© 2023 Google Terms of Use Report a problem

CURB RAMP LABELS SIDEWALK LABELS OTHER ACCESSIBILITY LABELS Overall Stats

Zoom In

Zoom Out

Sound

Jump

Stuck

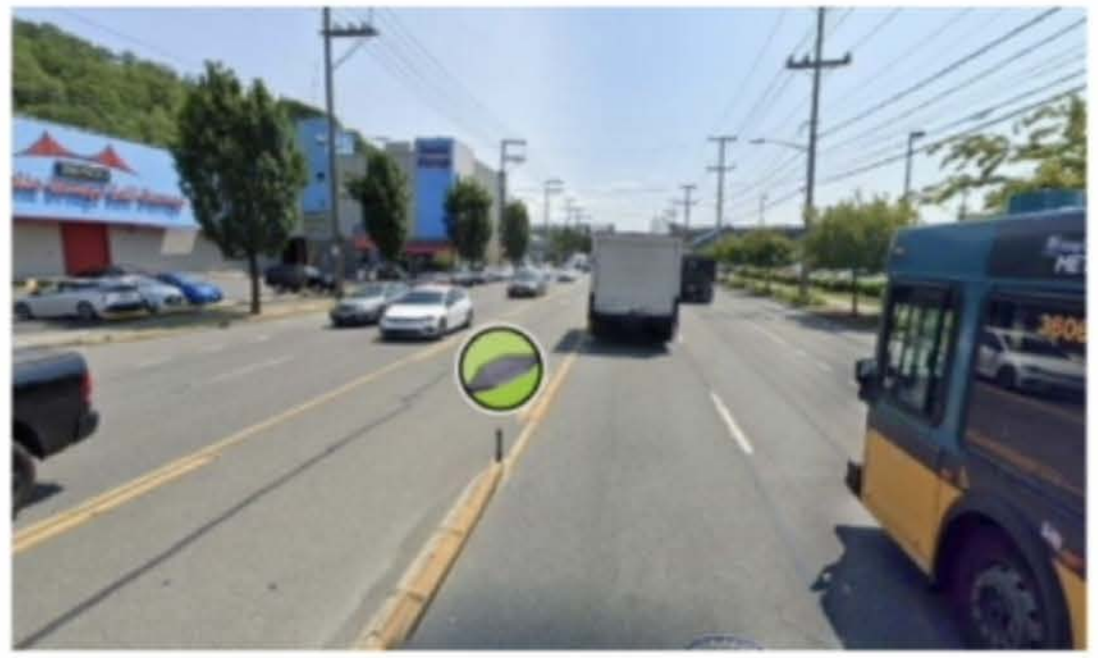
Feedback

Your Curb Ramp Label

Common Mistakes

Driveways. Driveways are **not** curb ramps. They are designed for vehicles and not pedestrians.

Curb Missing. When a curb ramp is missing, use the Missing Curb Ramp label instead.



Not on pedestrian route. Curb ramps are not needed at paths not intended for pedestrians.

Back to Labeling

See common mistakes 1 / 2 See correct examples

Version 7.15.6 | Last Updated: August 31, 2023

CURB RAMP LABELS SIDEWALK LABELS OTHER ACCESSIBILITY LABELS Overall Stats

Zoom In

Zoom Out

Sound

Jump

Stuck

Feedback

Your Curb Ramp Label

Correct Examples

This is a good **curb ramp**. it's wide, has a yellow tactile warning, and is not too steep.

This is an OK **curb ramp**. It's missing a tactile warning strip and is angled into the street.

Label flat **curb ramps** with tactile warning strips.

Some corners have very wide **curb ramps** to support travel in both directions.

Back to Labeling

See common mistakes < 2/2 > See correct examples

Version 7.15.6 | Last Updated: August 31, 2023

LabelAId

Machine Learning Framework

Technical Evaluation

User Interface Design & Implementation

User Study

Between-Subjects User Study

Participants & Task

- 34 participants
- Randomly assigned to 2 conditions
- 8 labeling routes

Measures

- Task performance
- Labeling confidence
- Knowledge gain
- User preference

Approach

- Data logging
- Quiz
- Questionnaire
- Interview

Can LabelAId improve **labeling performance**?

Findings - Task Performance

- LabelAId improves precision by up to 19% without compromising labeling time

Label Type	Control	Intervention	U	p-value
Overall	0.699 (±0.199)	0.891 (±0.053)	50.0	0.001 **
Curb Ramp	0.686 (±0.346)	0.956 (±0.067)	70.0	0.038 *
No Curb Ramp	0.802 (±0.164)	0.918 (±0.091)	80.5	0.025 *
Obstacle	0.7610 (±0.126)	0.812 (±0.111)	85.5	0.183
Surface Problem	0.812 (±0.230)	0.894 (±0.116)	100.0	0.423
No Sidewalk	0.842 (±0.267)	0.867 (±0.208)	66.5	0.480

precision low — precision high

How did the **perception of labeling ability** and **knowledge gained** differ between the two groups?

Findings - Self Efficacy & Learning Gains

- Higher self-efficacy in intervention group

- Similar objective learning gains

- Higher subjective learning gains in intervention group

How did participants **perceive LabelAId**?

Findings - Perceived Usefulness

- LabelAId was helpful (82.35%) & likable (64.7%)

> There were times when **I was not sure** if I should label it, and the system popped-up for me and said 'Are you sure about this?' **I found that really helpful.**

- Intervention Group Participant

Discussion

- Can AI-assistance replicate human feedback?
- How to design interactions with imperfect ML models?
- Cognitive forcing function reduces over-reliance on AI

Growing concern about AI-based assistance:

over-reliance on AI, reduced human cognitive engagement

Discussion

Obstacles high false positive rate 36.2%

Users rejected 83 % incorrect suggestions

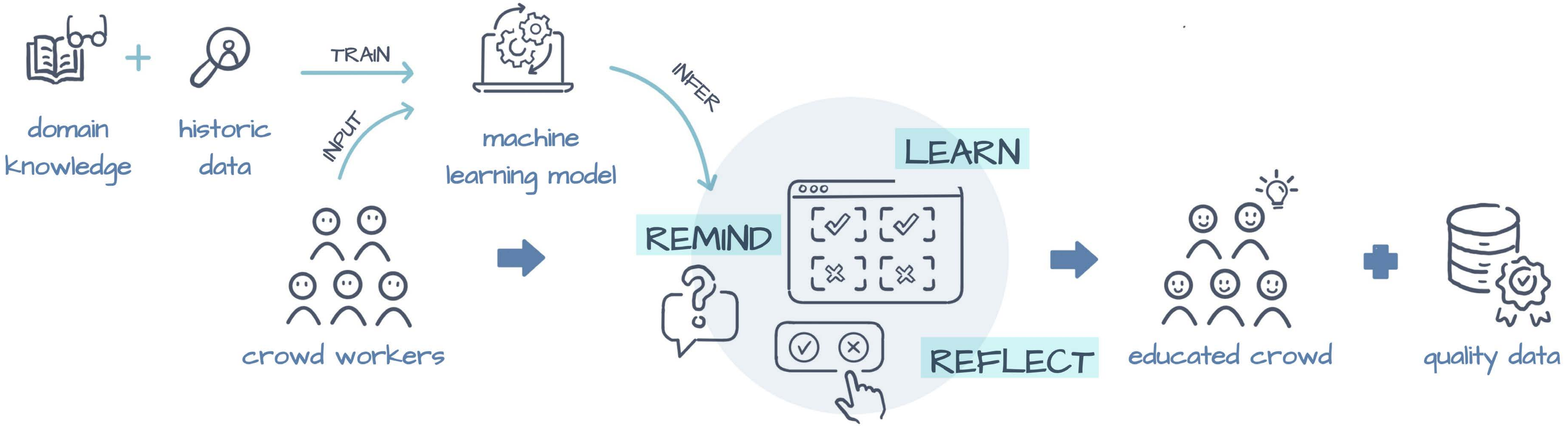
Users decide before seeing AI's suggestion

Cognitive forcing functions

elicit analytical thinking at decision making time

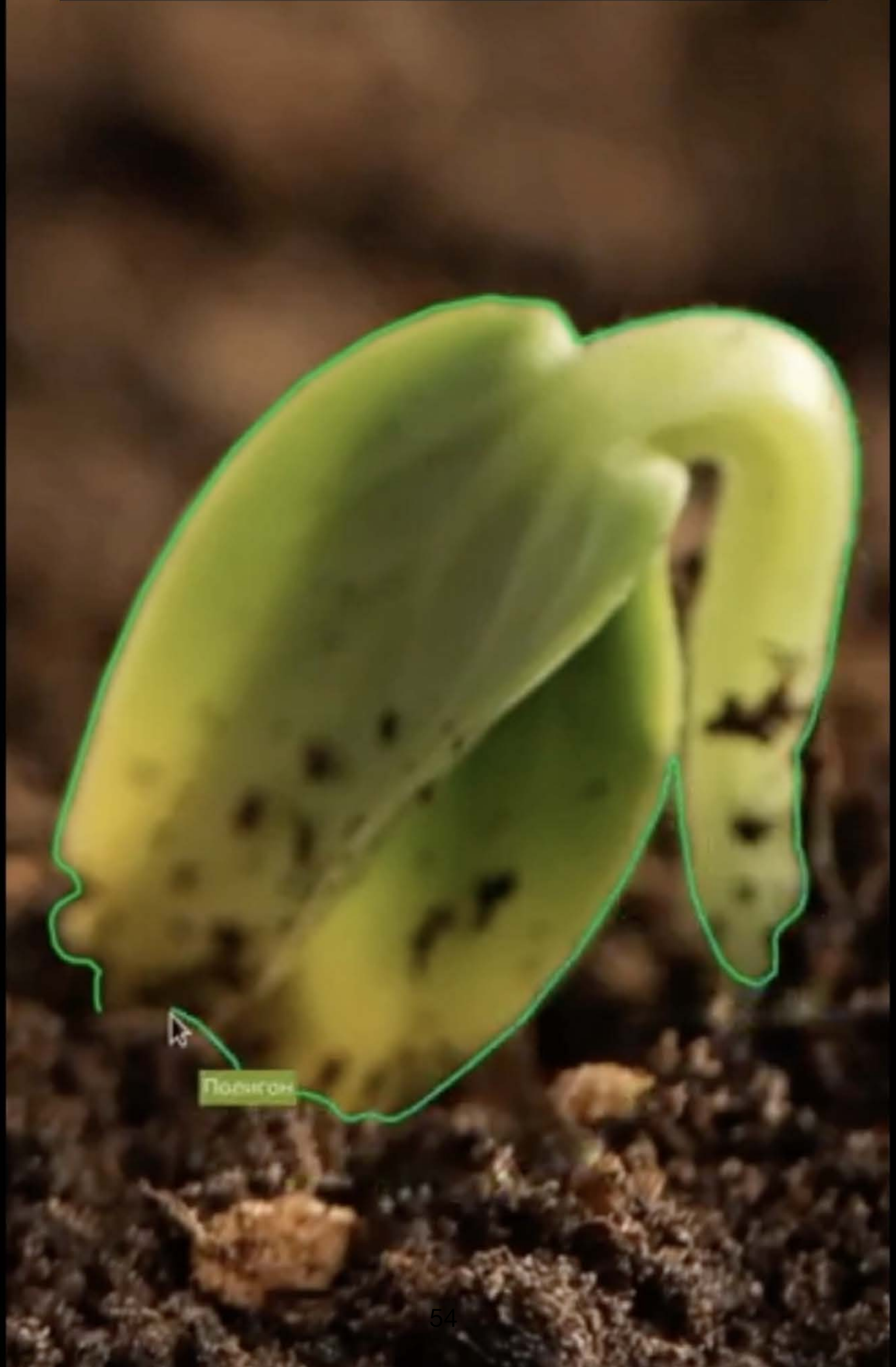
effectively reduces reliance on AI

LabelAId Just-in-time AI Interventions

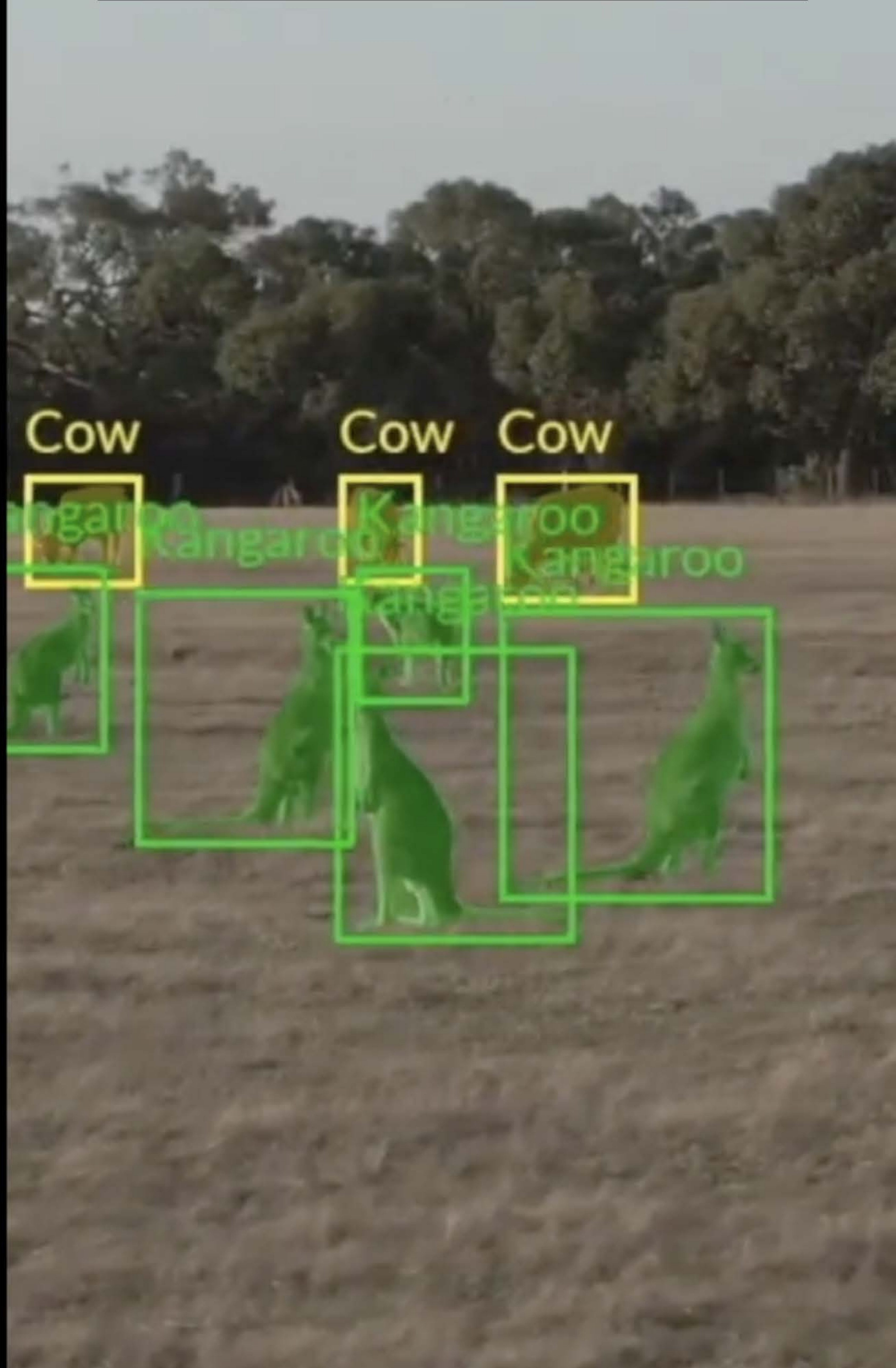


Medical Image Labeling

Agriculture Recognition



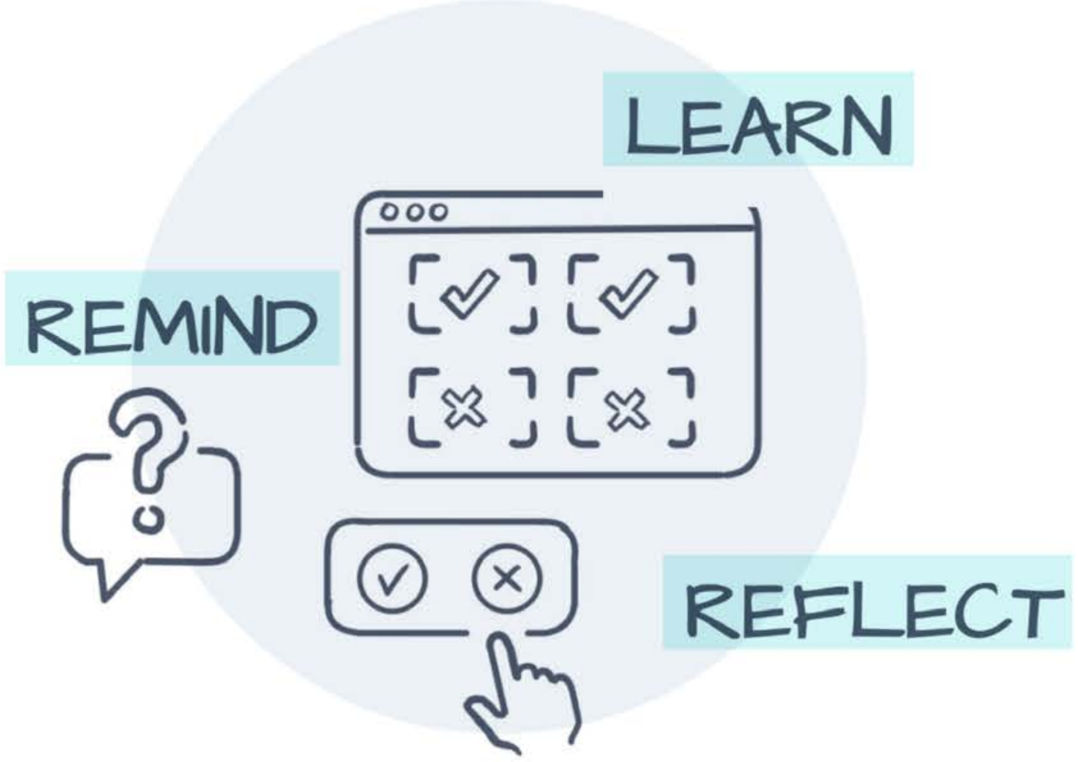
Wildlife Classification



LabelAId

Just-in-time AI Interventions

for Improving Human Labeling Quality and Domain Knowledge



A novel pipeline that facilitates the training of AI-based inference models for detecting labeling mistakes

A human-AI collaborative system designed to create teachable moments in crowdsourcing workflows

Contacts

chuchuli@cs.uw.edu
zzhihan@cs.uw.edu


Chu Li


Zhihan Zhang


Michael Saugstad


Esteban Safranchik


Minchu Kulkarni


Xiaoyu Huang


Shwetak Patel


Vikram Iyer


Tim Althoff


Jon E. Froehlich

MAKEABILITY LAB
DESIGN USE BUILD
CREATE
Center for Research and Education on Accessible Technology and Experiences
PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING
Berkeley UNIVERSITY OF CALIFORNIA