
Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning

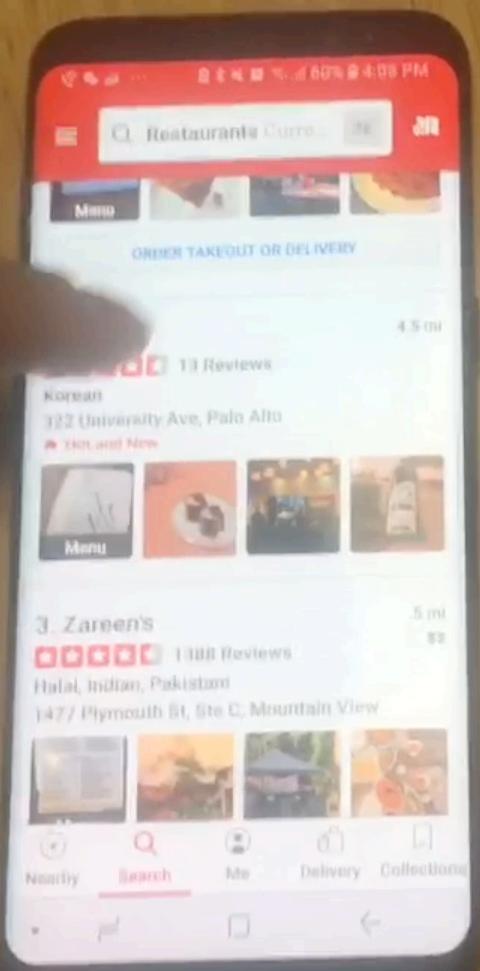
Amanda Swearngin, University of Washington

Yang Li, Google

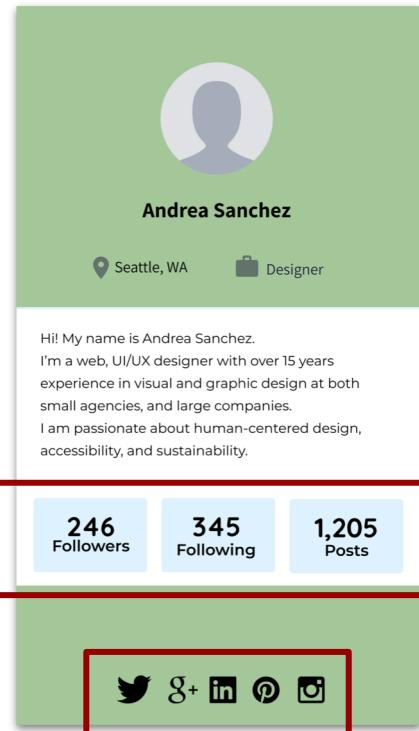


PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING





Signifiers



Color (e.g., A Link)

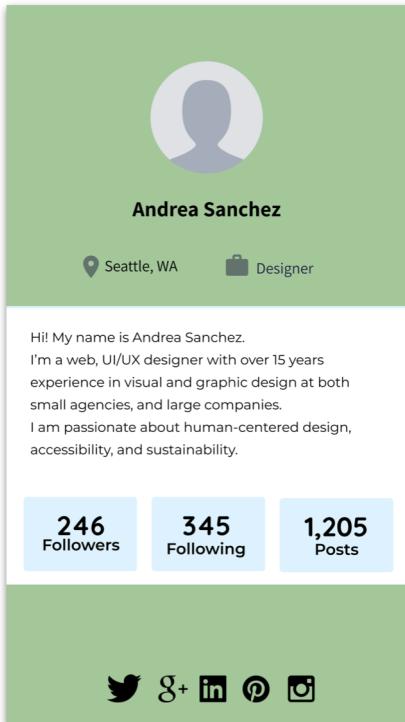
Shape of a button

Location on screen

A combination

...

What if a designer uses the wrong signifier?



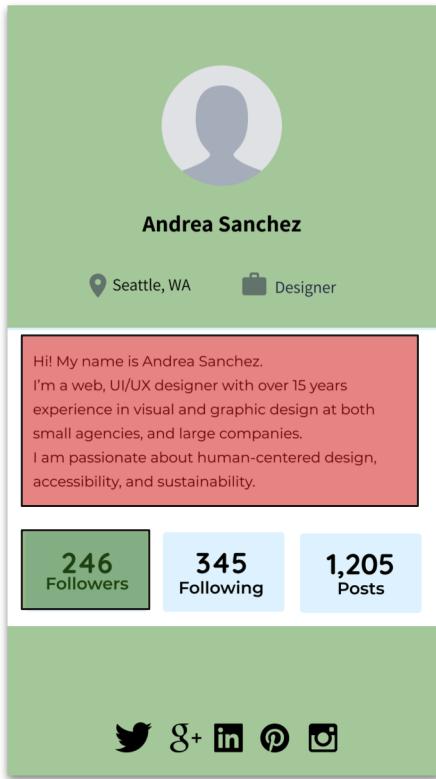
No Signifier

- Lack of discoverability

False Signifier

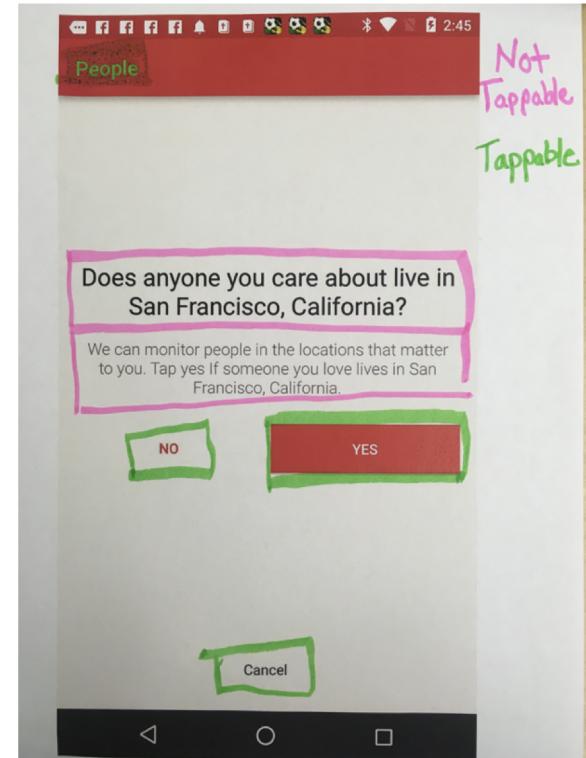
- Tap with no response -> frustration

A Tappability Study



Tappable

Not Tappable



Challenges

For Designers

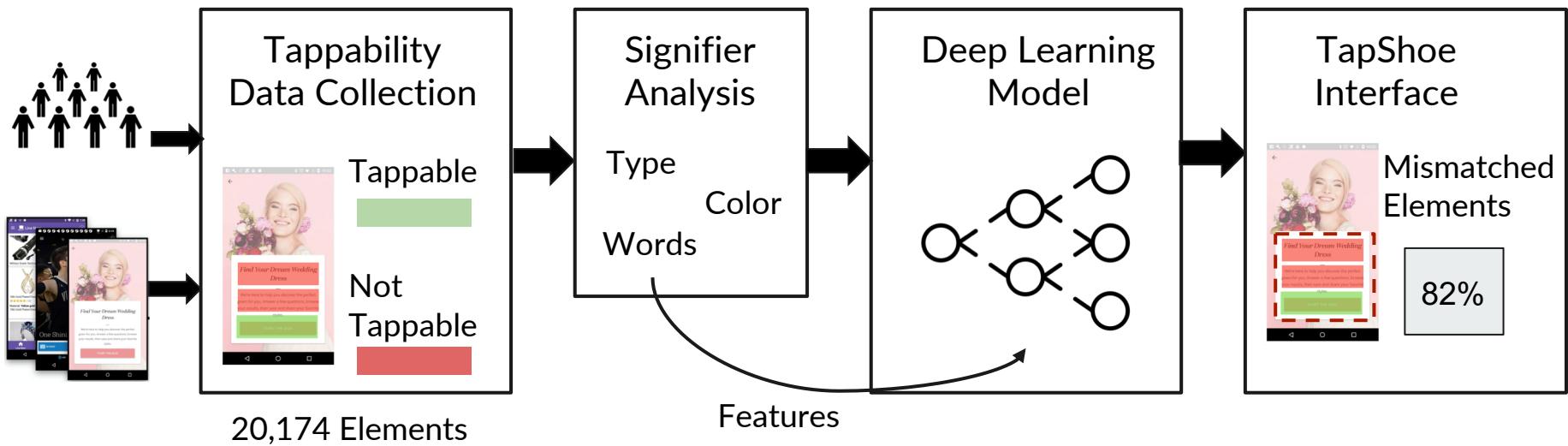
- Expensive
- Time consuming



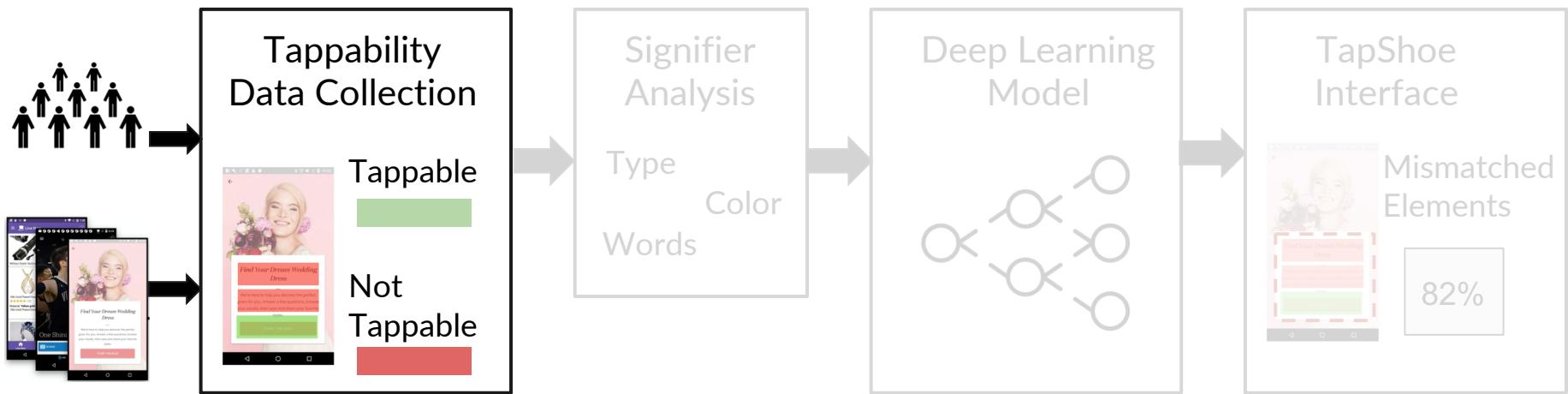
For design & research community

- No consistent understanding of signifiers at a large scale.
- Diverse tappability data needed to build automated approaches.

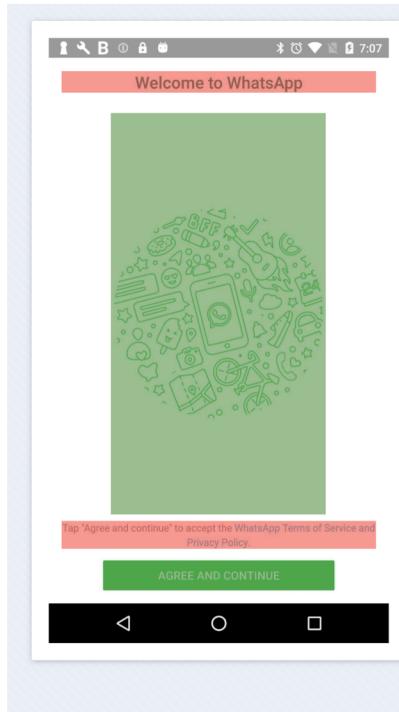
Our Approach



Talk Overview



Crowdsourcing Tappability Data



The screenshot shows the WhatsApp welcome screen. At the top, there's a red bar with the text "Welcome to WhatsApp". Below it is a large green circular graphic containing various icons related to messaging and communication. At the bottom, there's a red footer bar with the text "Tap 'Agree and continue' to accept the WhatsApp Terms of Service and Privacy Policy." and a green "AGREE AND CONTINUE" button. Below the footer is a black navigation bar with three white icons: a left arrow, a circle, and a square.

For the screenshot on the left, indicate whether each element is **tappable** or not **tappable**.

Tappable means that when you tap on it, an action will happen.

Not tappable means that when you tap on it, no action will happen.

Not tappable

Tappable

To submit, you need to select each element. You have **0** targets left.

Submit

Tappability Data

3,470 screens

743 workers

20,174 elements



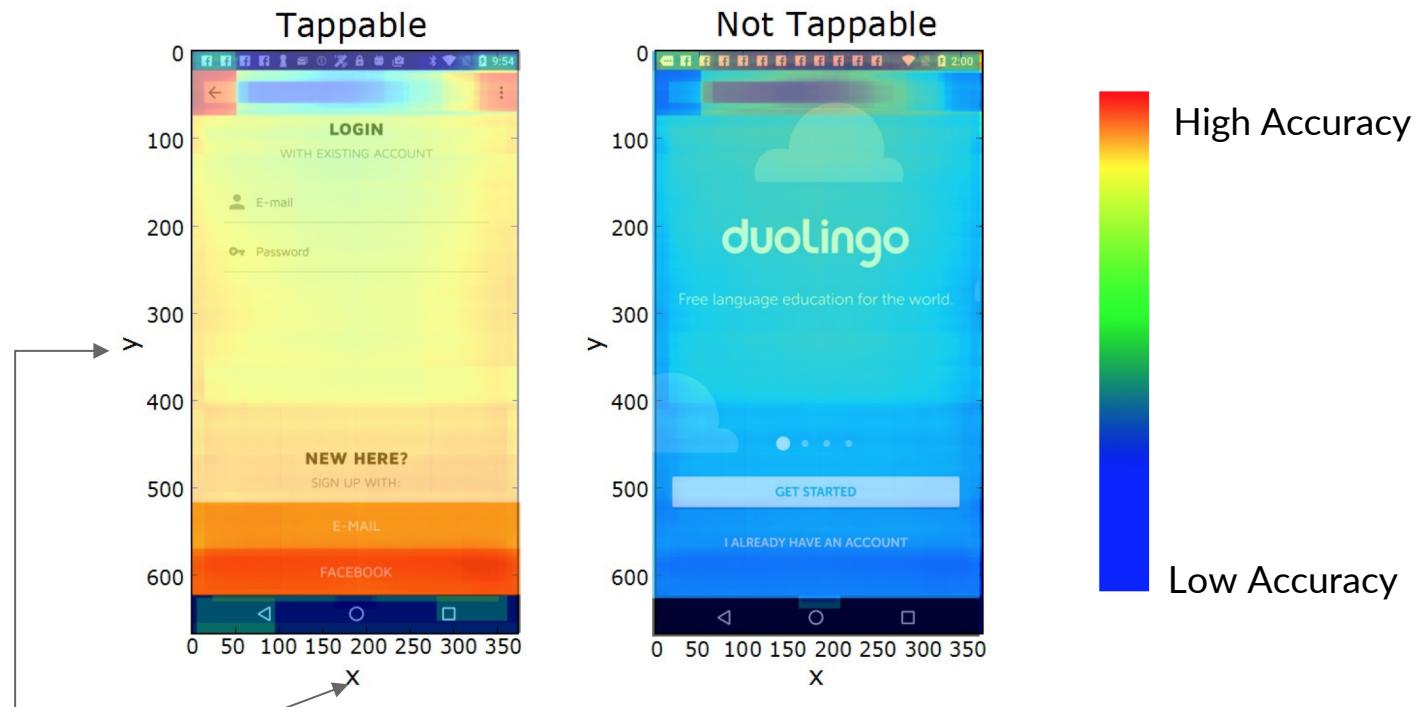
Accuracy of the Worker Labels

	# Labels Collected	Precision	Recall
Tappable	14,301	89.99%	79.67%
Not Tappable	5,873	61.31%	78.43%

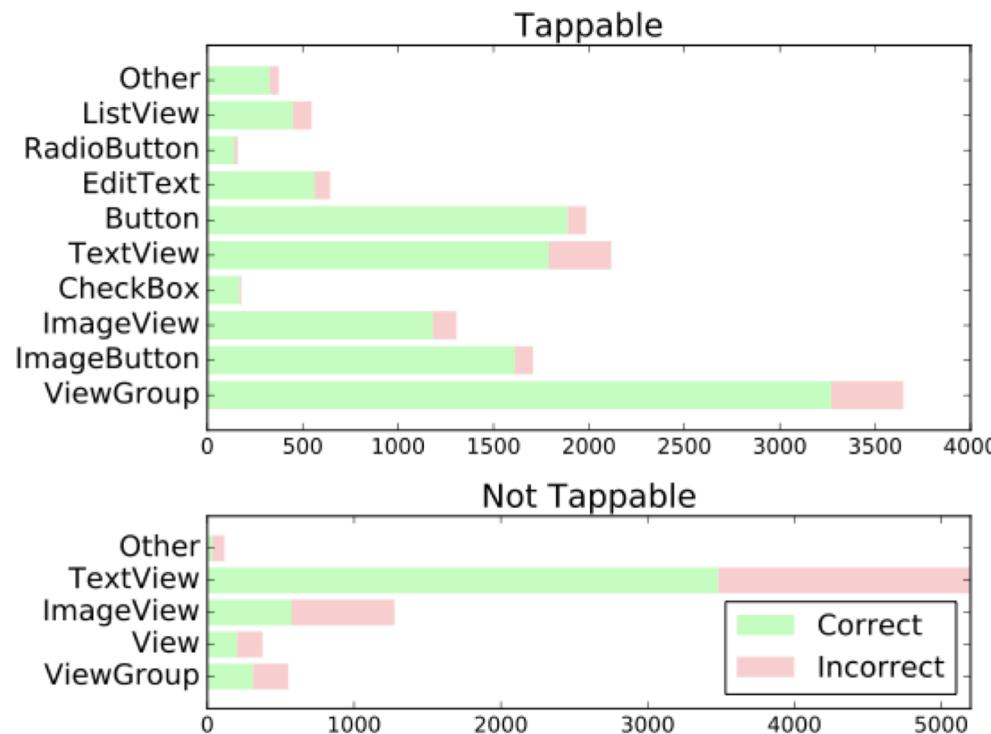
Talk Overview



How does location indicate tappability?



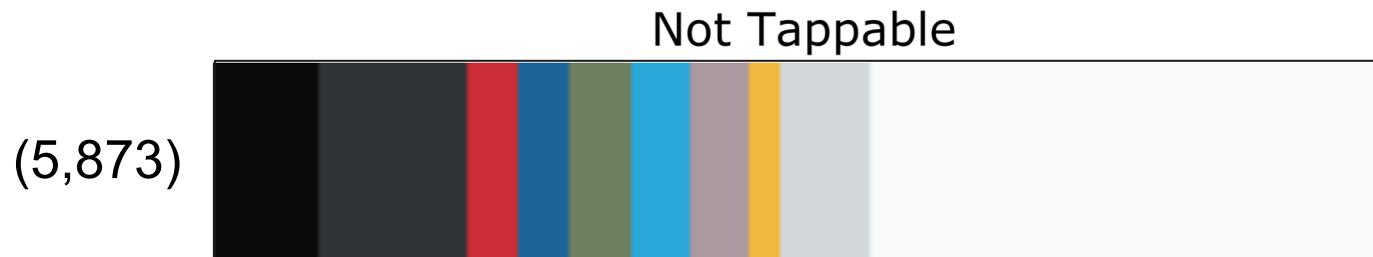
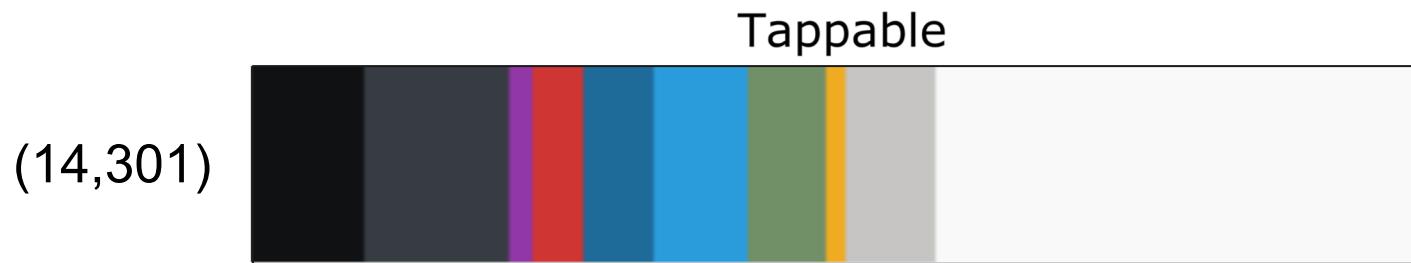
How does element type indicate tappability?



Tappable elements more correct, especially for common tappable elements (e.g., buttons, checkboxes)

Not tappable elements most common types have more flexibility in design -> more ambiguity.

What colors are more common in tappable elements?



Do tappable elements have fewer words, and more actionable keywords?

Not tappable elements had **1.84** more words per element, on average.

Top 5 Tappable Keywords ← TF-IDF Analysis

1. Submit
2. Close
3. Brown
4. Grace
5. Beauty

Talk Overview

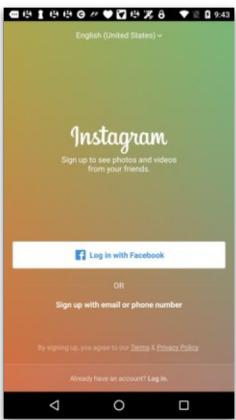


Tappability Model

Features

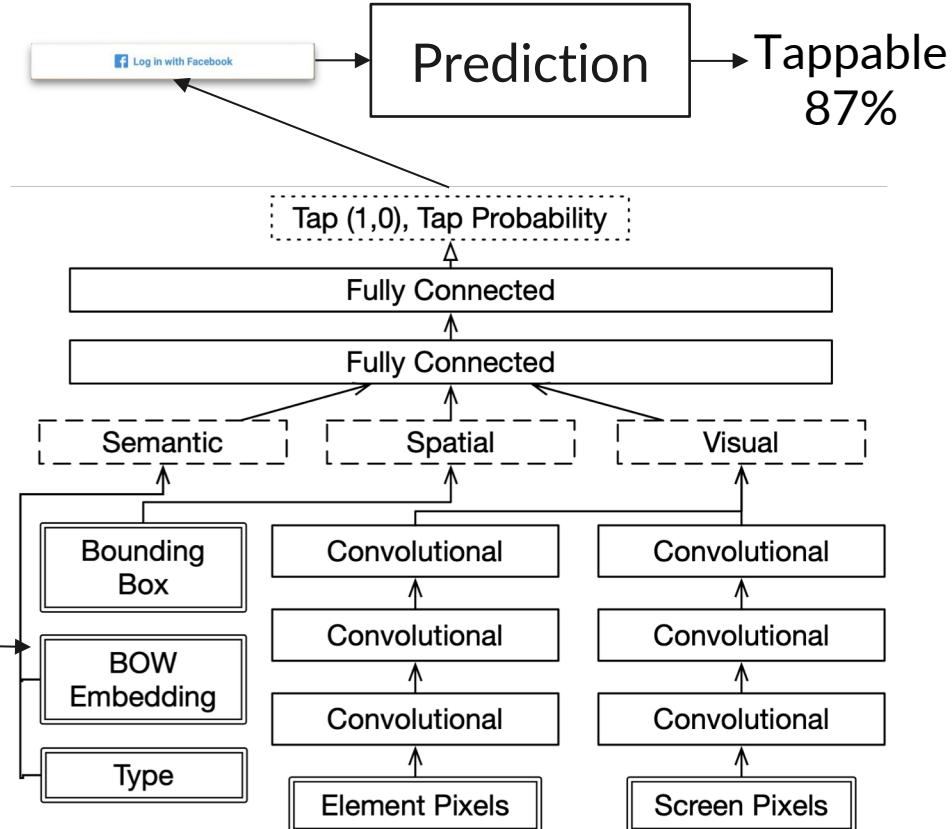
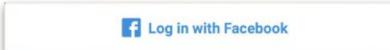
Bounding Box

{ x: 22, y: 30, width: 120, height: 40}



Type: Button

"Log in with
Facebook",
Number of Words: 4



How well can we predict tappability?

Original
Dataset

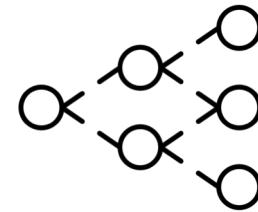
Tappable	Not Tappable
P: 90.2% (SD: 0.3)	P: 70% (SD: 2.0)
R: 87.0% (SD: 1.6)	R: 78% (SD: 3.0)

Balanced
Dataset

Tappable	Not Tappable
P: 82% (SD: 0.3)	P: 81% (SD: 2.0)
R: 84% (SD: 1.6)	R: 86% (SD: 3.0)

How can we improve the model's accuracy?

Add more features, improve model



Are human labels inconsistent?



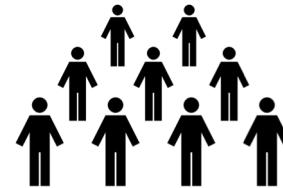
How consistent are the tappability labels?

290 workers

2,000 unique elements

334 screenshots

Each element labeled 5 times



Tappable

Not Tappable

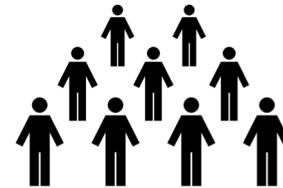
How consistent are the tappability labels?

Results

58% elements labeled the same among all 5 workers.

Agreement Score¹: 0.834

Fleiss' Kappa²: 0.520 (Moderate)

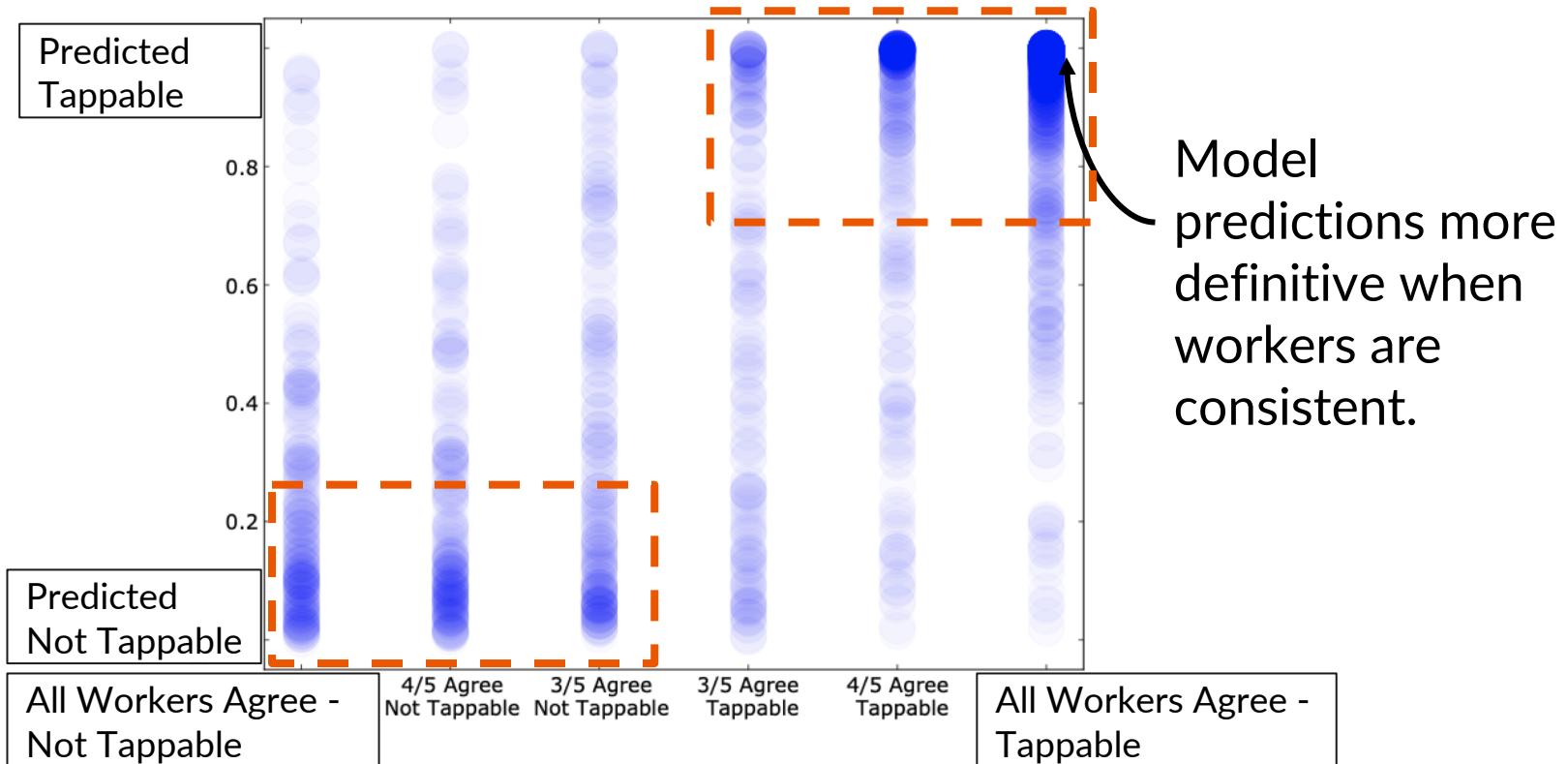


Tappable

Not Tappable

1. Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. "Maximizing the Guessability of Symbolic Input", CHI 2005
2. Joseph L Fleiss, "Measuring Nominal Scale Agreement Among Many Raters", Psychological Bulletin, 1971

Do the model results reflect consistency?



Talk Overview



TapShoe Interface

The screenshot shows the TapShoe interface with a teal header bar. On the left, there's a user profile card for "Andrea Sanchez" with a placeholder profile picture. The profile includes the name "Andrea Sanchez", location "Seattle, WA", and occupation "Designer". Below the profile is a bio: "Hi! My name is Andrea Sanchez. I'm a web, UI/UX designer with over 15 years experience in visual and graphic design at both small agencies, and large companies. I am passionate about human-centered design, accessibility, and sustainability." To the right of the bio are three stats boxes: "246 Followers", "345 Following", and "1,205 Posts". A callout box at the bottom right of the stats says: "User Tappable: Tappable In Code: Not Tappable Probability: 59%". On the right side of the interface, under "Tappability Results", it states: "TapShoe found 4 mismatched elements. Click an element to the left to see further details." At the top of the main content area, there are "Display Options" and a "Sensitivity" slider set to 0. There are also two checked checkboxes: "Tappable -> Not Tappable" and "Not Tappable -> Tappable".

User Tappable:
Tappable
In Code: Not Tappable
Probability: 59%

This Element

This target is Not Tappable in the view hierarchy but there is a 59% chance users will think it is Tappable.

Designer Interviews



Informal interviews with 7 professional designers

Demonstrated them TapShoe interface and model

Questions:

- How do you see the TapShoe interface fitting into your design process?
- How can you envision using the models predictions, beyond the TapShoe interface?

How can we help designers understand tappability?

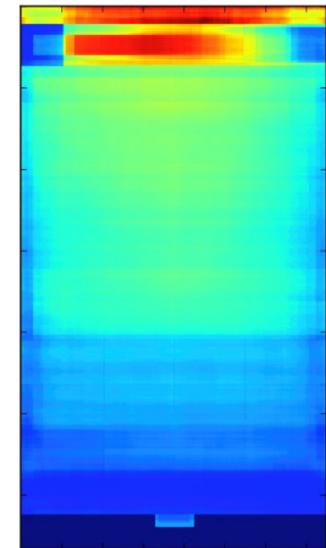
TapShoe interface - Provide recommendations for a fix

Spatial visualization of tappability (i.e., Heatmap)

Tool to explore small variations, and discover new signifiers.

Train on existing datasets or platforms

Predictions on early stage mockups (i.e., Sketch documents)



Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning

Key Takeaways:

- People have low accuracy in distinguishing tappable from not tappable elements.
- We can build models that use visual, spatial, and semantic features to predict human tappability perception.
- This can help designers understand and improve the usability of their interfaces.

Amanda Sweeny*
Paul G. Allen School
University of Washington
amaswea@cs.washington.edu

Yang Li
Google Research
yangli@acm.org



* This work was completed while the first author was an intern at Google.