Human-AI Collaboration for Conservation

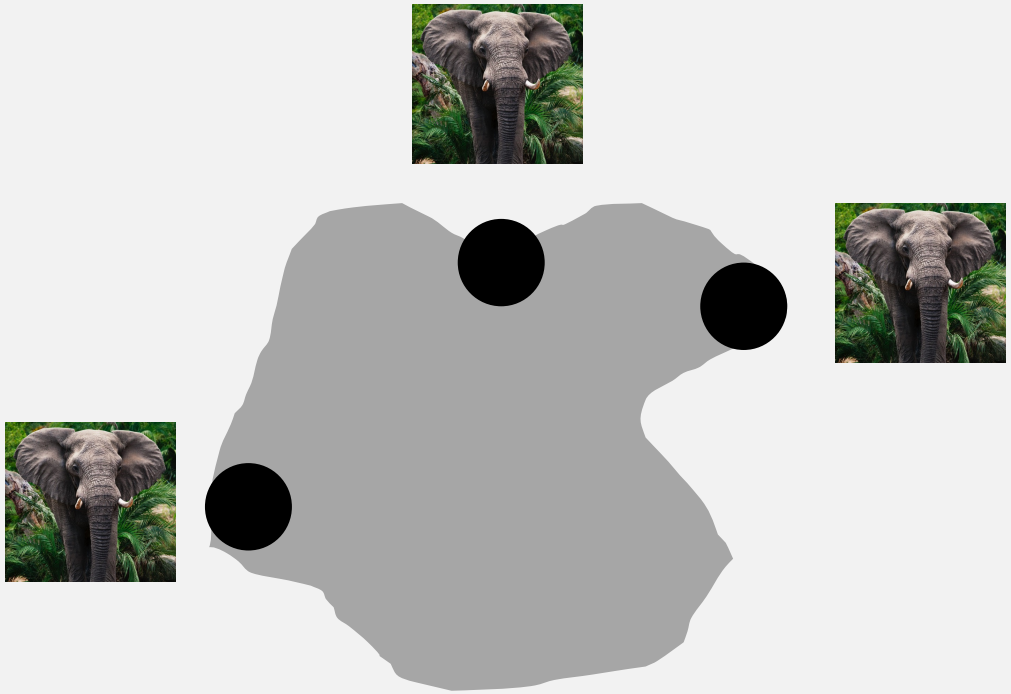
Elizabeth Bondi-Kelly

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

M

Photo Credit: Air Shepherd

AI in Conservation

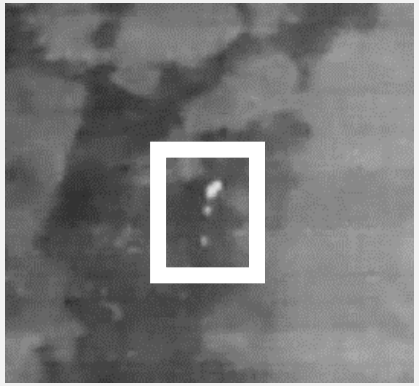


Park rangers

Drones

Targets

- Noisy
- Small
- Hard to collect

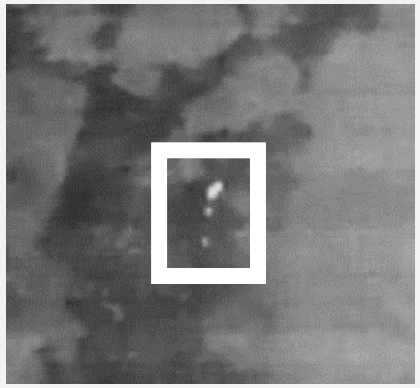


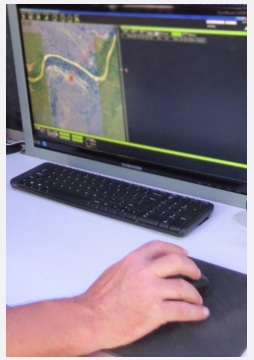
Humans involved in system

Photo Credit: Air Shepherd

AI in Conservation

- Noisy
- Small
- Hard to collect



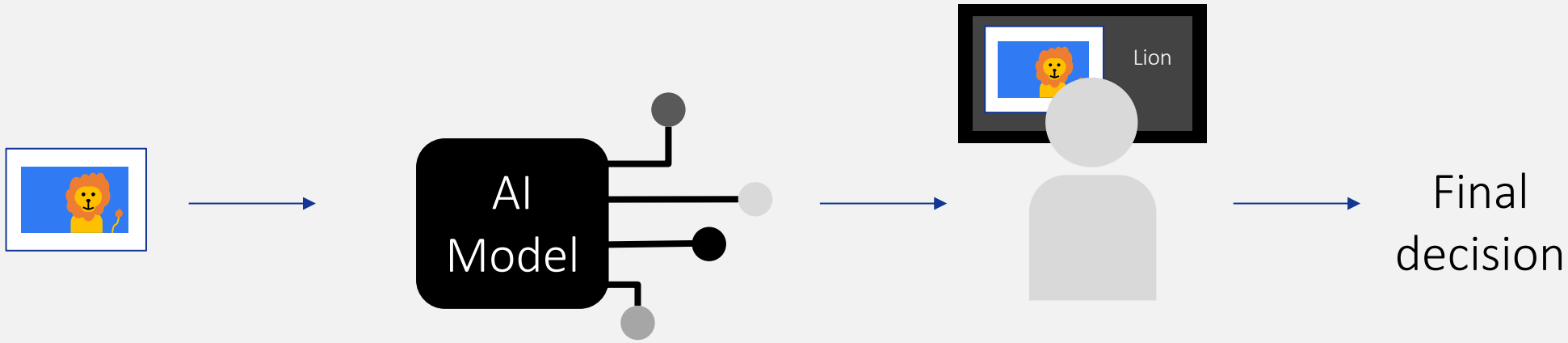


Humans involved in system

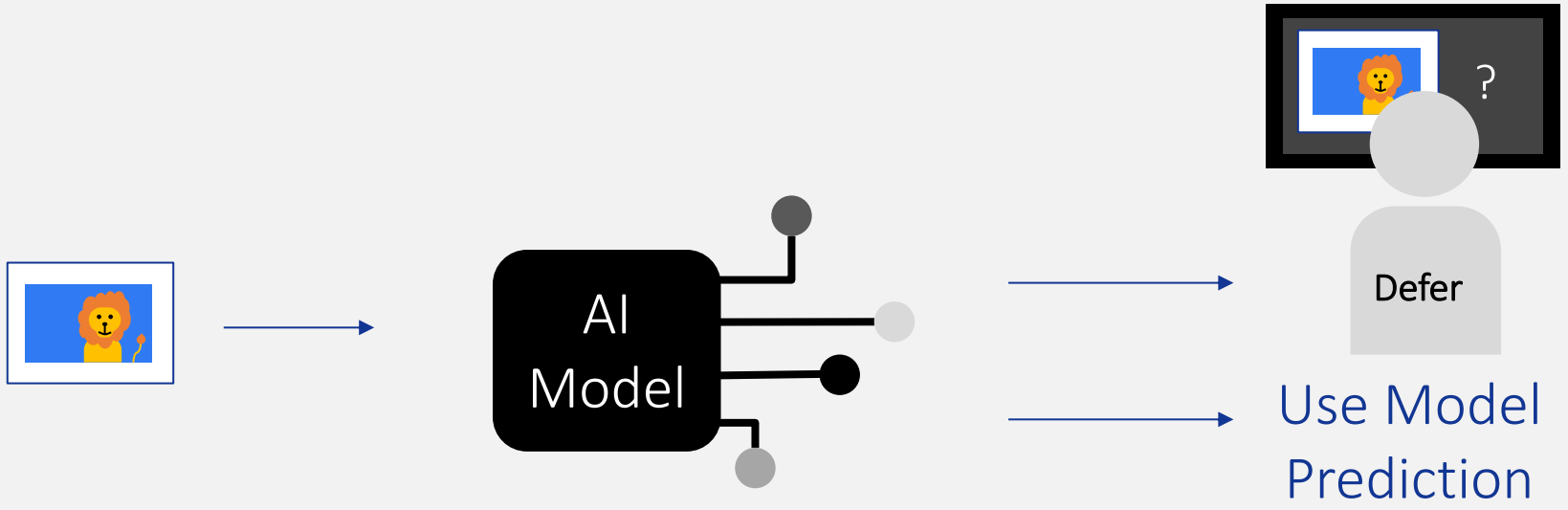
Photo Credit: Snapshot Serengeti

Example

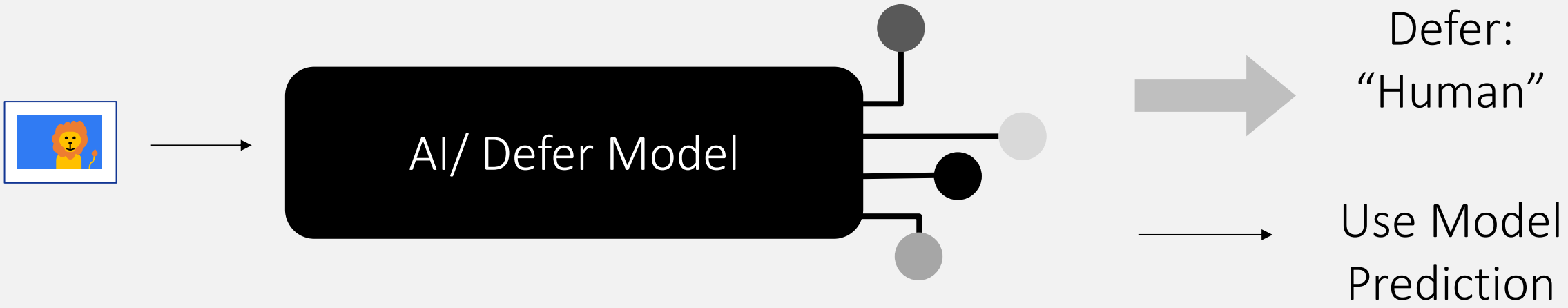
Human-AI Collaboration in Practice



Human-AI Collaboration in Practice



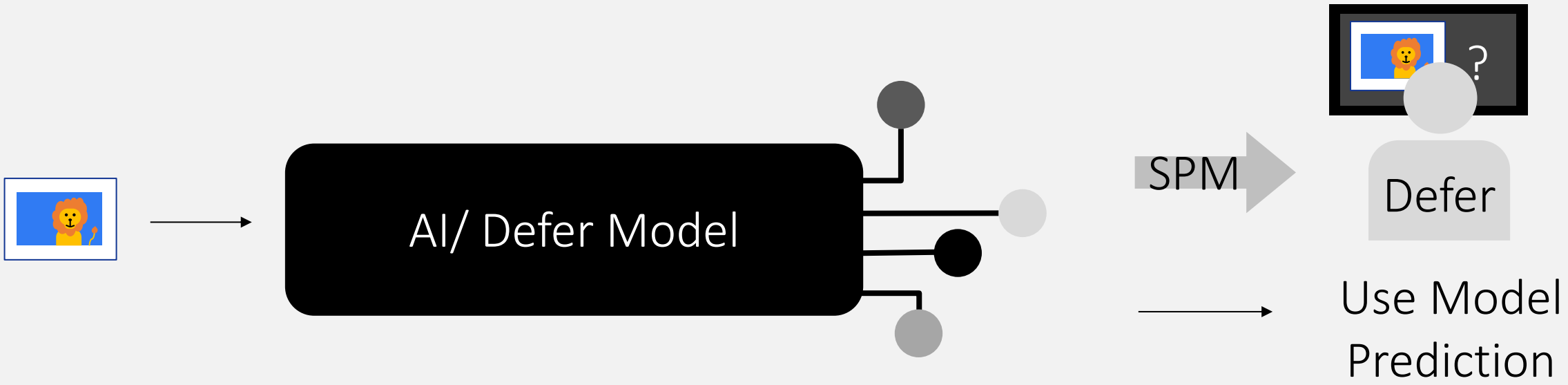
Prior Work on Selective Prediction



Humans may over-rely on AI model predictions and/or may distrust an AI model

- Dietvorst, Simmons, Massey 2015
- Rastogi, Zhang, Wei, Varshney, Dhurandhar, Tomsett 2020
- Gaube, Suresh, Raue, Merritt, Berkowitz, Lermer, Coughlin, Guttag, Colak, Ghassemi 2021
- …

Question



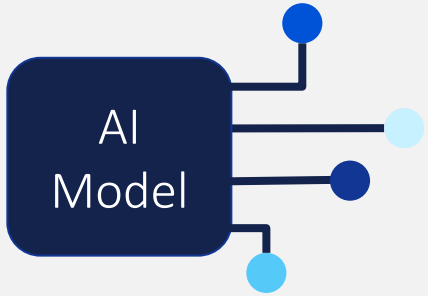
Does deferral still help in **human**-AI systems?

Deferral Model

Predictions

Ground truth

Historical human labels



Train deferral model

Photo Credit: Snapshot Serengeti

Training

Camera Trap Training

You are about to see camera trap
will be asked to label whether an i
images being harder to identify th

Camera traps are cameras triggere
motion. These can be used to captu
conservation.

They can be very ambiguous, sinc
portion of the animal while it mov
the camera, or it may even be can

Adding AI Assistance

We have also trained an AI model to assist humans in this task. It is correct 97% of the time.

The AI model is also equipped with deferral, so you may see a message like "AI model deferred." It means that the AI model is uncertain. On cases that are not deferred, it is correct 98% of the time.

The messages you will see communicating these results may vary. This includes cases where you may not see any message, as we are trying to collect human labels as well.

Please pay attention to the messages that are provided to you that give you context about the AI.

, the best label is animal present.

Photo Credit: Snapshot Serengeti

Selective Prediction Messaging

User would see one of the 4 conditions shown here:

Image 1

Image 40: AI model deferred.

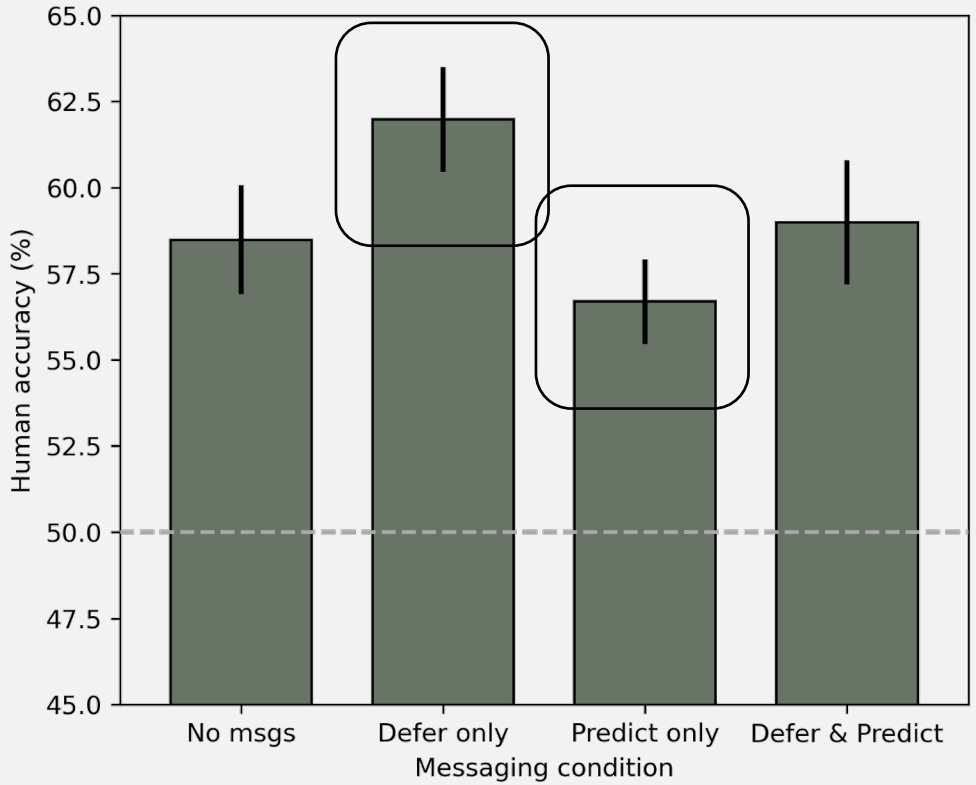
Image 6: AI model predicts no animal present.

Image 37: AI model deferred, but predicts no animal present.

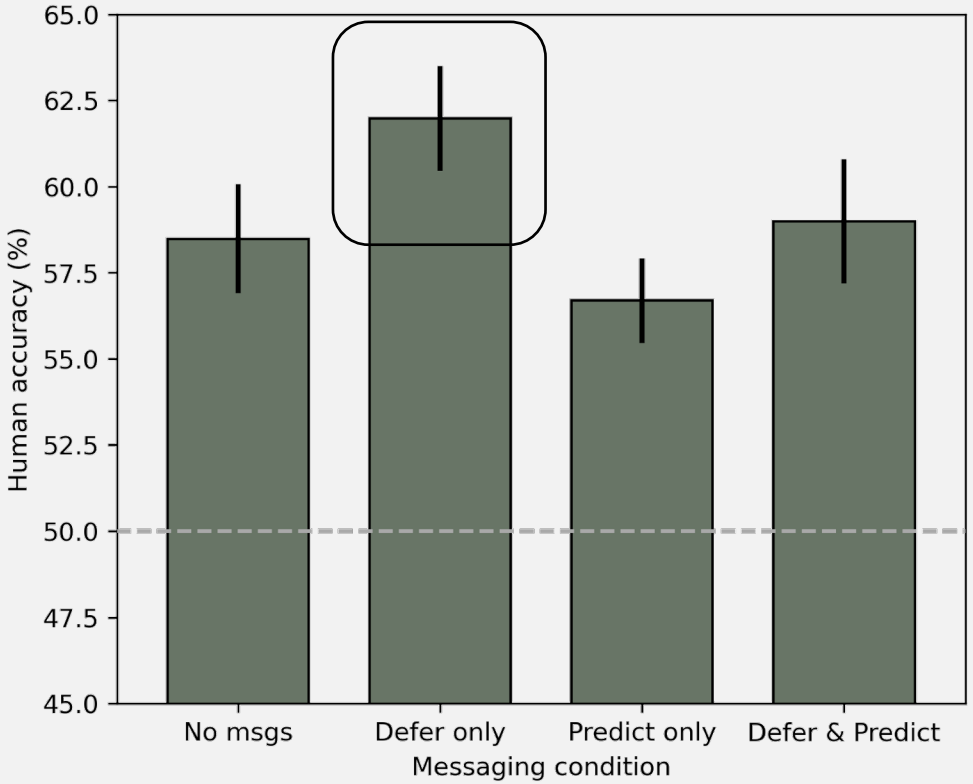
1 2 3 4 5

Definitely no animal present ○ ○ ○ ○ ◉ Definitely animal present

Results


65.0
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
Human accuracy (%)
No msgs
Defer only
Predict only
Defer & Predict
Messaging condition

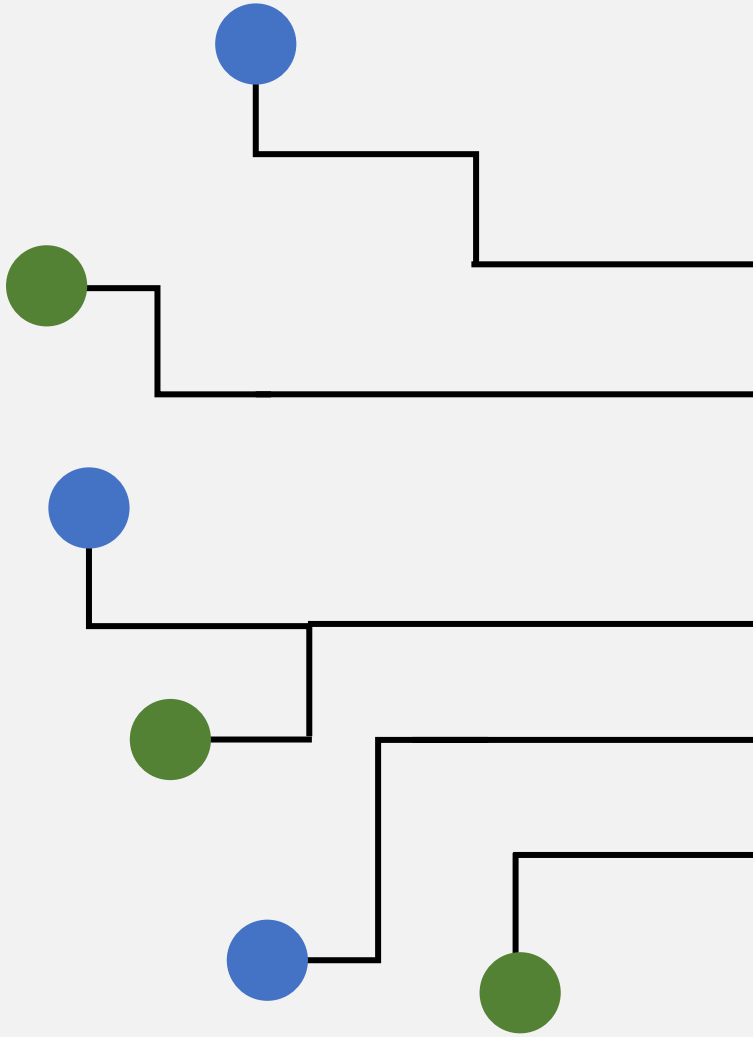
Results


65.0
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
Human accuracy (%)
No msgs
Defer only
Predict only
Defer & Predict
Messaging condition

THANK YOU!

Elizabeth Bondi-Kelly
Email: ecbk@mit.edu / ecbk@umich.edu
Twitter: @BondiElizabeth
https://sites.google.com/view/elizabethbondi/

www.try-ai.org
Twitter: @TryAIOrg



Experiment

- 198 Prolific participants recruited
- 80 deferred images (20 TN, 20 TP, 20 FP, 20 FN)
  - No repetition
  - No message that all are deferred
  - Randomly allocated across 4 conditions
- 4 separately seeded random allocations (about 50 participants each)

Results