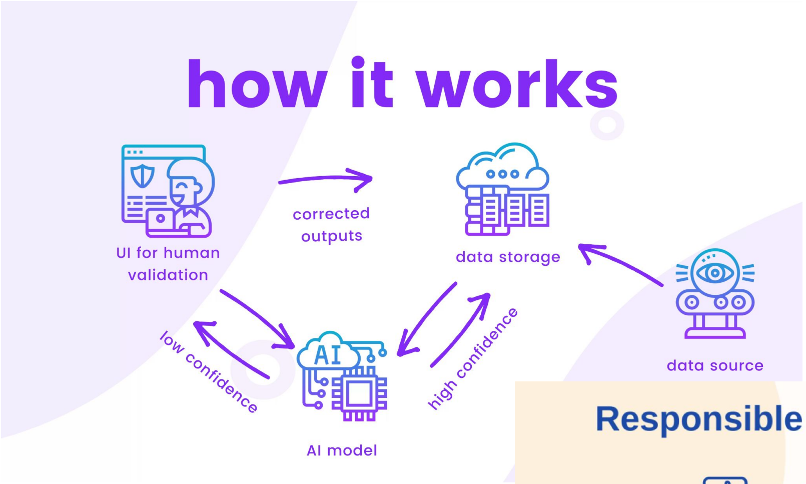
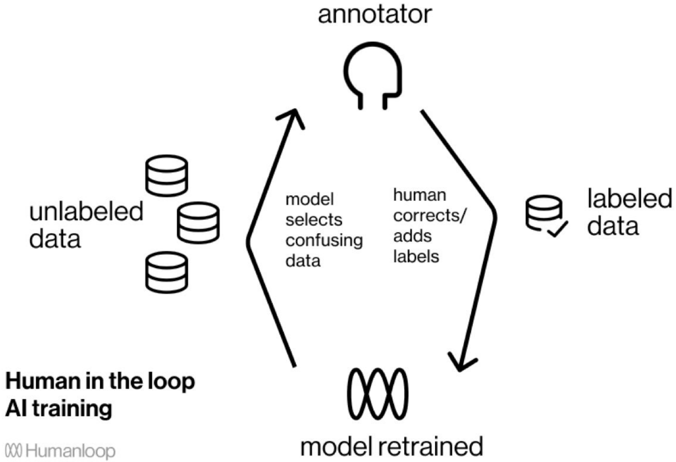
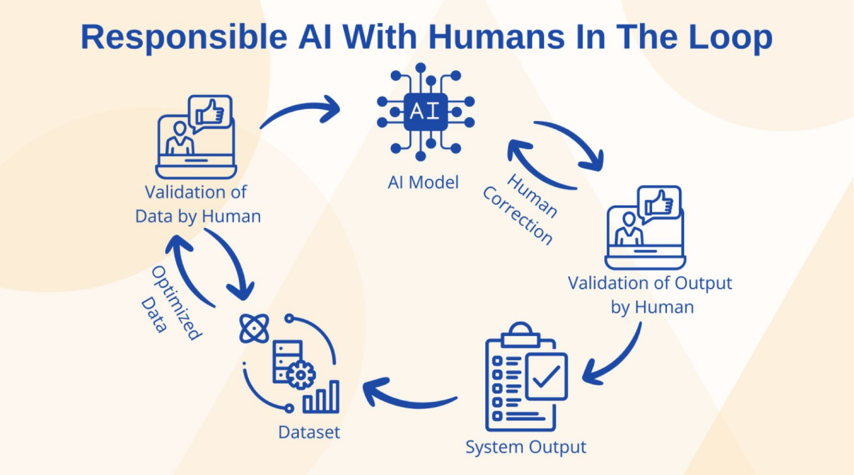
Lecture 8: Active Learning and Selective Prediction

Sara Beery | 4/8/25

Bringing humans “in the loop”





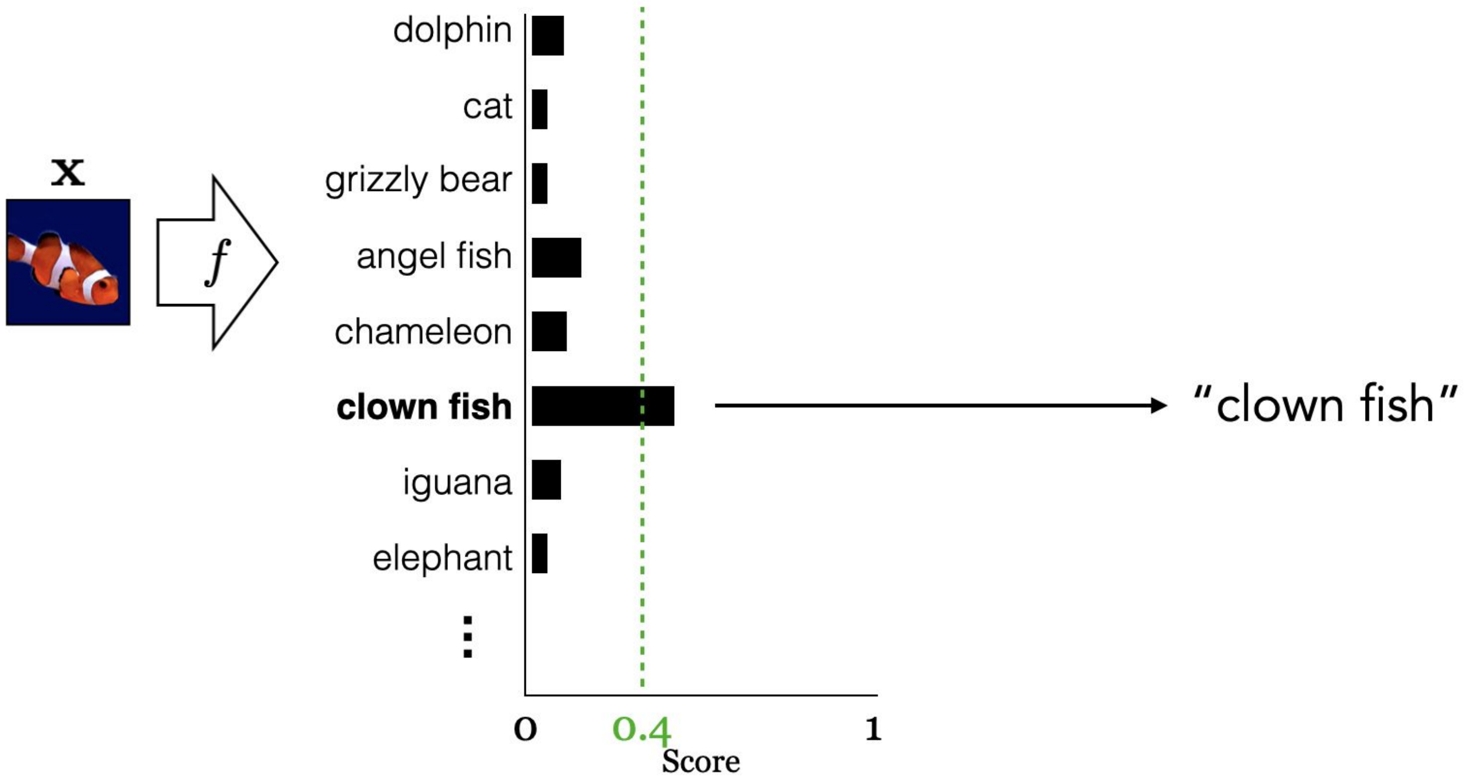


Diagrams courtesy of industry PR:
SalesForce
HumanLoop
Humans in the Loop

Prediction

$$\hat{\mathbf{y}}$$

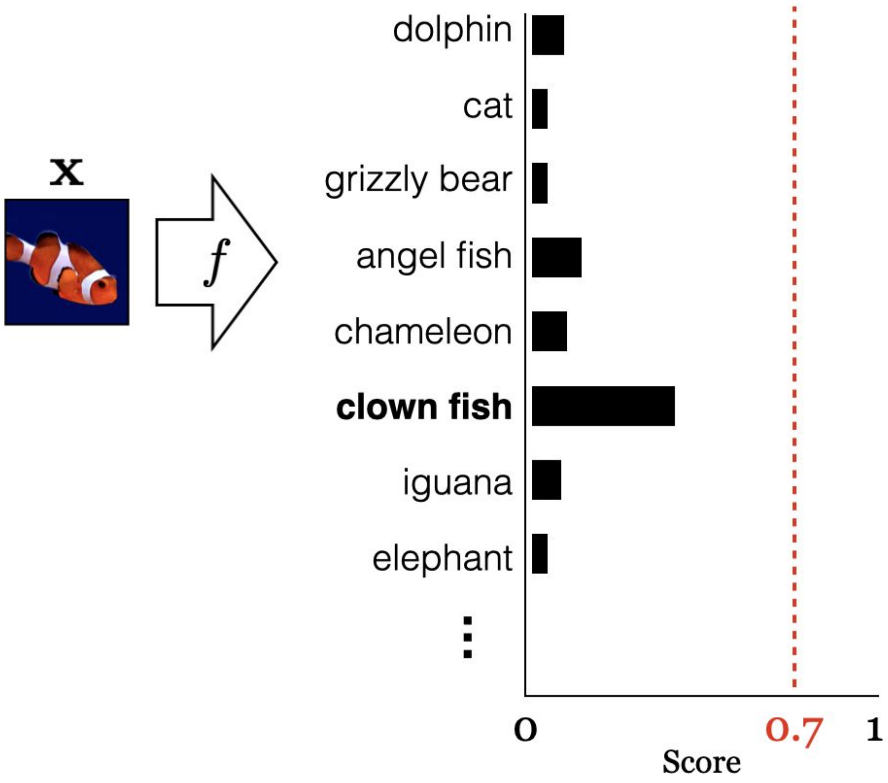
$$f_\theta : X \to \mathbb{R}^K$$



Prediction

$$\hat{\mathbf{y}}$$

$$f_\theta : X \to \mathbb{R}^K$$



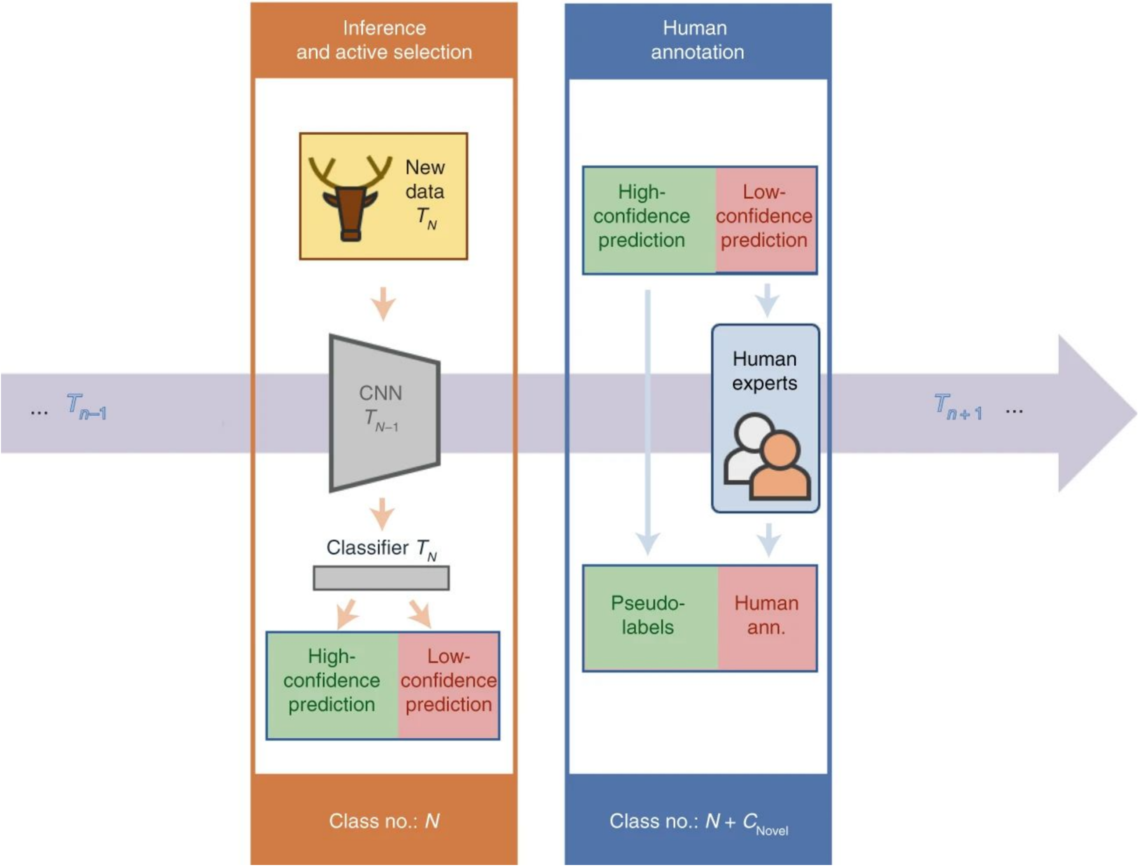
Selective prediction

“I don’t know”

Selective prediction gives an abstain option, it doesn’t force a decision but instead takes model confidence into consideration

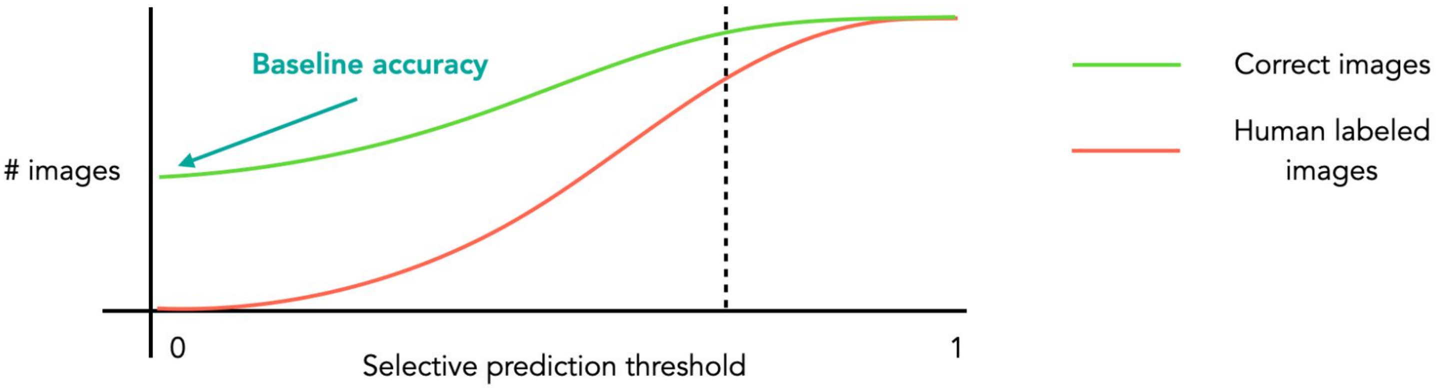
In practice, a human would then identify images that a model abstains

Selective prediction



https://www.nature.com/articles/s42256-021-00393-0

Accuracy vs human effort in selective prediction



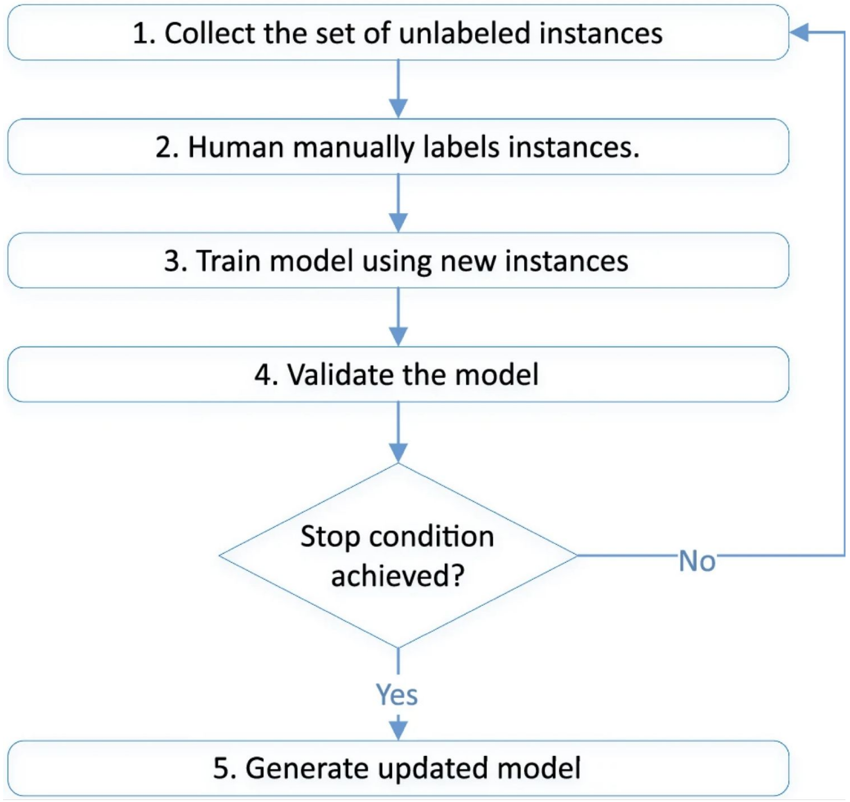
- Low thresholds mean the model is trusted more, thus less human effort needed to identify all the data but there is more possibility of error
- High thresholds mean the model is trusted less, thus humans ID more data but quality is easier to guarantee
- Threshold selection is an active area of research, calibrated models make this easier

Active learning

Learn to sample next data for human labeling automatically to optimize performance while minimizing human effort

Sampling criteria:

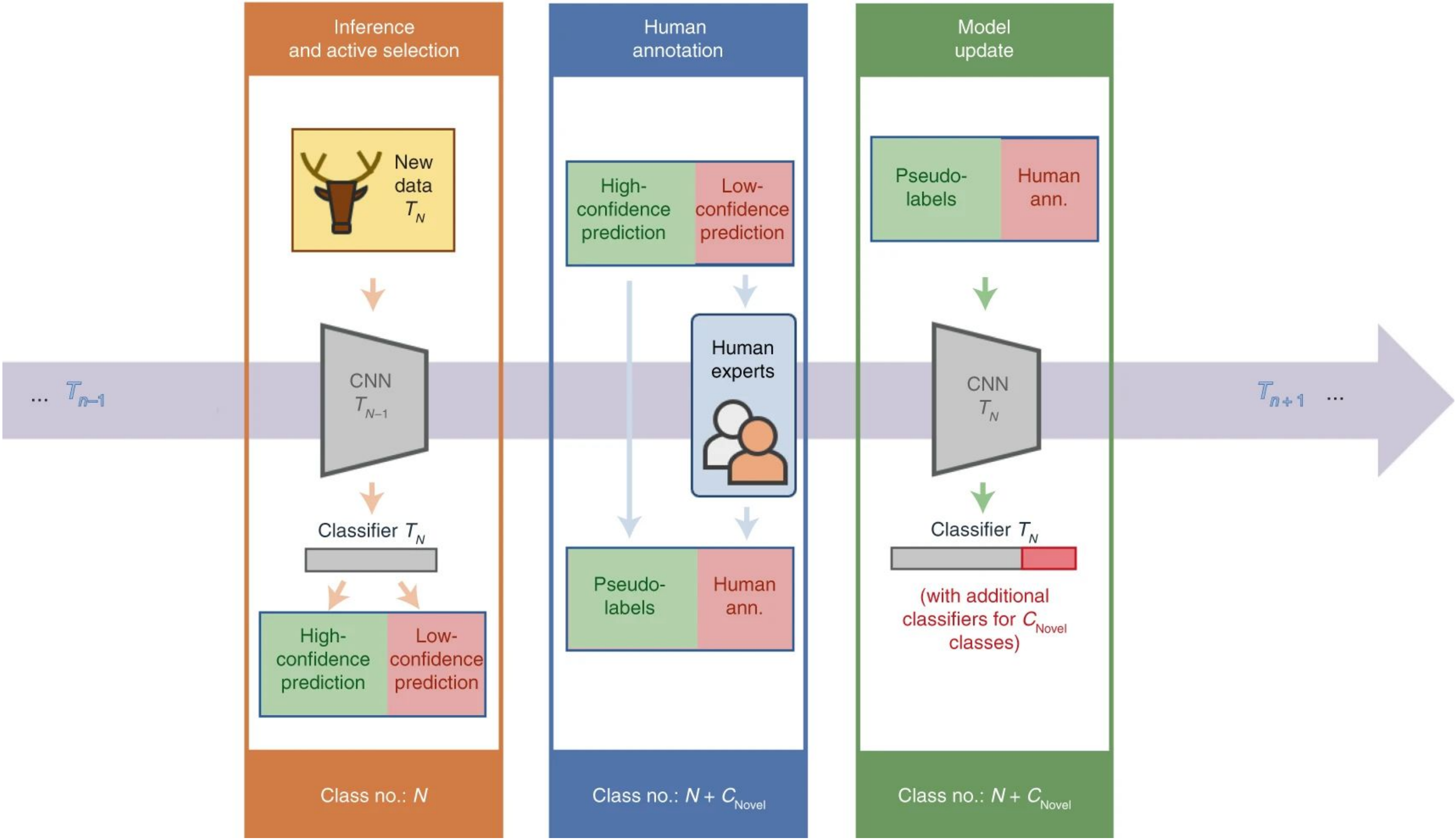
- Random
- Uncertainty (Exploit)
- Diversity (Explore)



Human-in-the-loop machine learning: a state of the art, Mosqueira-Rey et al., Artificial Intelligence Review 2022

Human-in-the-loop machine learning, Munro, Manning Publishing 2020

Active learning *via* selective prediction



https://www.nature.com/articles/s42256-021-00393-0

Active learning based on representations

One example:

- Use the MegaDetector to crop
- Cluster animals based on visual similarity in new cameras
- Humans ID examples from each cluster (active learning criteria)
- Gets same accuracy with **99.5% fewer labels**

A deep active learning system for species identification and counting in camera trap images, Norouzzadeh, Morris, Beery, et al., Methods in Ecology and Evolution 2021

Role of Human-AI Interaction in Selective Prediction

User would see one of the 4 conditions shown here:

Image 1

Image 40: AI model deferred.

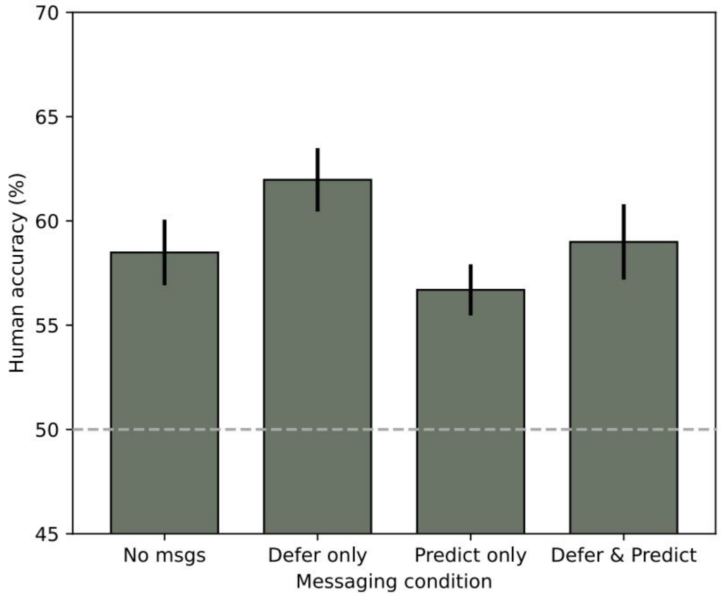
Image 6: AI model predicts no animal present.

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

1 2 3 4 5

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

Human accuracy decreases when model results are presented



https://ojs.aaai.org/index.php/AAAI/article/view/20465

Role of Human-AI Interaction in Selective Prediction

User would see one of the 4 conditions shown here:

Image 1

Image 40: AI model deferred.

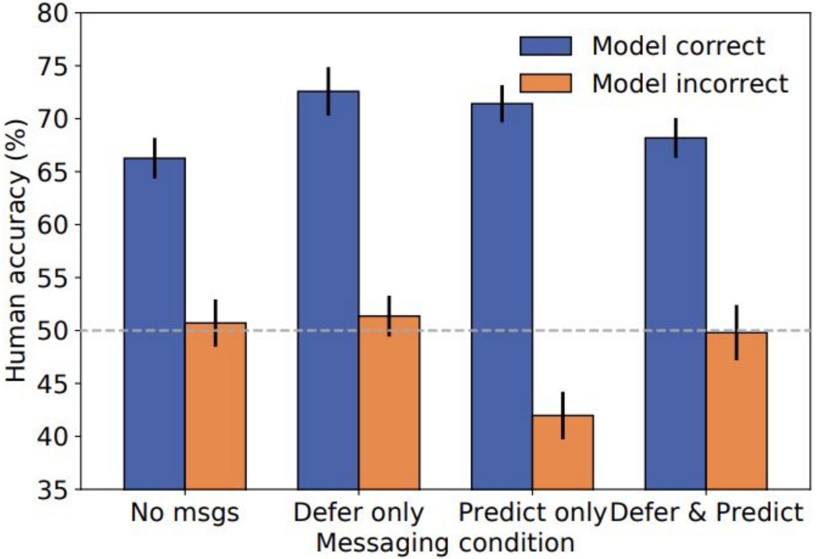
Image 6: AI model predicts no animal present.

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

1 2 3 4 5

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

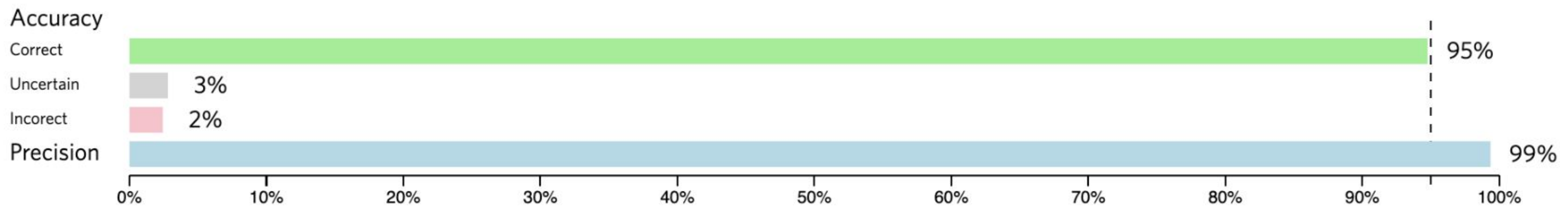
Human accuracy decreases when model results are presented



https://ojs.aaai.org/index.php/AAAI/article/view/20465

Confirmation bias

For the Research Grade subset, 95% were Correct, 3% were Uncertain and 2% were Incorrect. The average Precision was 99%.



I had actually (not long ago) studied the question of subspecies of Apis mellifera in Africa and therefore knew, that bees from NE Namibia, SW Zambia and the Zambezi valley can't be identified to a subspecies, this area is a zone of introgression between A. m. scutellata and A. m. adansonii.
(My "wisdom" comes from a PHD thesis available for download here: Radloff, S. 1996. Multivariate analysis of selected honeybee populations in Africa https://commons.ru.ac.za/vital/access/manager/Repository/vital:5734/SOURCEPDF?site_name=Rhodes+University)
Obviously none of the other identifiers was aware of this. And this is when the confirmation bias sets in - you just agree without actually considering that you do not know how to identify this taxon.

Posted by traianbertau about 1 year ago