

DSC234 Inside the Machine: AI Literacy Lesson Plan for February 2, 2026

Topic: The Use, and Abuse, of Human Labor in the Development and Training of AI - Image Classification and "The Mechanical Turk"

Materials Needed

Some Examples of "hidden pictures" puzzles (we used "Where's Waldo" books)

Between 8 and 12 images of creatures which have characteristics like cats (fossa, falanouc, mongoose, hyena, meerkat, genet, binturong are all possibilities) along with some which would be correctly classified as a "cat".

A collection of images hand drawn by children or adults (some cats, some dogs)

Learning Outcomes

The student will be able to state what the original Mechanical Turk was, and comment on its connection to fostering stereotypes and generalizations about people who have darker skin and wear turbans.

The student will participate in discussion concerning labor practices and lack of transparency within the corporations and entities involved in crowdsourcing AI models.

The student will be able to describe a process of image classification using a crowdsourced method - including how the method is employed, how workers are compensated (piece-rate pay), and analyze the accuracy of the results.

Lesson Details

1. (10 minutes) Start with a discussion (based on the assigned reading before class from Crawford) about the original Mechanical Turk. Make sure to include the following elements:

a. original was built and presented as a "machine" that would play chess with an unsuspecting person on the outside of the box.

b. a description: featured a life-sized, turbaned mannequin with brown skin dressed in robes—resembling a stereotypical "Ottoman sorcerer" —seated behind a large, dark-wood cabinet. Inside a real person would determine each chess move.

c. connection between it and what we now call "Artificial Intelligence" is that behind the scenes there are real people being exploited by companies to do the grunge work of what computer algorithms miss or get wrong. The concept of "No Harm" is a myth.

2. (10-15 minutes) Give students a printout of the following paragraph:

From the introduction of the paper

Yasmin, R., Hassan, M. M., Grassel, J. T., Bhogaraju, H., Escobedo, A. R., & Fuentes, O. (2022). Improving Crowdsourcing-Based Image Classification Through Expanded Input Elicitation and Machine Learning. *Frontiers in artificial intelligence*, 5, 848056.
<https://doi.org/10.3389/frai.2022.848056>

"A complementary approach for image classification that has received significant attention in various domain-specific applications is **crowdsourcing**. Its growth has been accompanied and propelled by the emergence of online crowd-sourcing platforms (e.g., [MTurk](#), [Prolific](#)), which are widely employed to recruit and compensate human participants to annotate and classify data that are difficult for machine-only approaches. In general, crowdsourcing works by utilizing the concept of the "wisdom of the crowd" (Surowiecki 2005), where the judgments or predictions of multiple participants are used to sift out the noise in individual predictions and better approximate a ground truth (Yi et al. 2012). Numerous studies over the last decade have established that, under the right circumstances and with the proper aggregation methods, the collective predictions of multiple non-experts are uncontroversially more accurate than those from almost any individual including well-informed experts. This concept of **using groups** to make collective decisions has been successfully applied to a number of visual tasks ranging from simple classification and annotation (Russakovsky et al. 2015) to complex real-world applications, including assessment of damages caused by natural disasters (Barrington et al. 2012) and segmentation of biomedical images for diagnostic purposes (Gurari et al. 2015)."

Ask students to identify areas in which crowdsourcing is being used to train AI. If possible, give them copies of the entire article (13 pages in all) and ask them what applications there are, as well as what assumptions are made. Students may be encouraged to do a search on job postings for this type of work.

3. (45 - 50 minutes) Experiment with Crowdsourcing

Go over the "Modified Majority Voting" Technique (Page 4) that we will now use to classify a set of images as either cat or not a cat. Display the images one at a time on the screen. Each student must answer the 3 questions without discussing them with anyone. A student volunteer then tallies the results on the board. Each image will then be labeled as "cat" or "not a cat". The instructor then reveals which ones the class correctly identified, and the accuracy rate discussed.

Note: When implemented on 2/3/2026 we only used 4 images (in the interest of time). Students were correct on 3 out of 4 of the images, until we employed the "100% confident rule". Then, the accuracy was 100%. In future, I will make sure to incorporate some hand-drawn pictures as well.

4. (10 - 15 minutes) Time permitting, run the following to illustrate piece-rate pay:

Divide the class into unequal sized groups - groups of 1, 2, and 3. Give each group a set of images with a hidden element. "Where's Waldo" for example. Offer a "bonus pay" of 5c per image to each group for correctly finding Waldo and placing a sticky on Waldo.

Post Lesson Reflection

What worked well: Students were highly engaged and energetic with the main experiment of crowdsourcing image classification. Everyone put "thumbs up" afterward when asked "did you learn something new today"?

Area for development: Students seemed bored and distracted when discussing the article at the beginning. It is difficult for some to engage in abstraction. This is why we spent only 10 minutes on this portion.

What I enjoyed as the instructor: I am beginning to develop a great rapport with this class. I enjoyed researching this topic and designed the activity without the help of AI. Had I used an AI tool, I would not have had the benefit of deeply diving into the topic and would not have been able to adequately determine beforehand if the lesson might accomplish what was intended.

Modified Majority Voting (MMV) for Binary Image Classification

To each participant:

Question 1: "Is this an image of a cat?" Answer: yes or no. Yes = 1, No = 0

Question 2: How confident are you in your response? Reply with a whole number between 0 and 100. 100% = absolutely certain. Be careful! If you say you are 100% certain and you end up being wrong, you lose all earnings \$\$.

Question 3: What do you predict a majority of other participants will say? Answer: yes = 1 or no = 2

Our Process of Crowdsourcing Image Classification

Step 1: Start with Majority Voting (MV)

The "image class" is either "1" if more participants said "TRUE or YES, this is a cat" or "0" if more participants said "FALSE or NO, this is not a cat"...

Step 2: In *case of a tie*, go to Confidence Weighted Majority Voting (CWMV)

If a participant is 100% certain it is a cat, their portion of the score would be 1. But if a participant is only 50% certain it is a cat, their score would be "weighted" accordingly, such as 0.5. Then do the same for all participants that said "no". Take the sums of all the weighted "yes" scores, compare with the sums of all the weighted "no" scores. The larger total is the deciding vote.

Step 3: If there is a participant who was 100% certain the image is a cat, and that participant has been evaluated as an "honest participant" then automatically make the image result = 1 (TRUE). Alternatively, if a participant is 100% certain the image is not a cat, despite the majority, then the image gets automatically assigned a score of 0 (FALSE) or not a cat.

In Machine Learning (computer) models, ***weights must be assigned at each hidden layer for image classification***. This is a basic overview of how those initial weights may be assigned before testing on a set of images.

Once all images have been given a score we check the group's accuracy. If the group correctly identified 9 out of 12 images the accuracy is 75%.

Compensation: You will each be compensated 10c for each correctly identified image. Anyone reporting 100% confidence resulting in an incorrect identification will receive nothing.

Notes About Crowdsourcing AI Training and Testing Within a Machine Learning Framework:

Binary Choice Elicitation

Workers are just asked to make a yes/no decision for each image, like a regular binary label. You then combine many people's yes/no answers (for example by majority vote or a machine-learning model) to decide the final label. For the machine learning model, a value that is the sum of all "zero counts" divided by the number of participants (like an average) is input into the model for calculations.

Example:

Show an image and ask: "Is there a cat in this picture? Yes or No." Each person clicks Yes or No; you aggregate all those clicks to label the image. Count how many "nos" and divide by the total number of participants. This (decimal number between 0 and 1) becomes the weight for the machine learning model.

Guess of Majority Elicitation

Workers still see the same image and possible labels, but **now you ask them what they think the majority of other people will answer (yes or no)**, not just what they personally think. You can then use how many people predict "most others will say yes" as an extra signal about the likely true label.

Example:

Show an image and ask: "What do you think most other workers will answer: Yes (cat) or No (no cat)?" Even if someone is unsure themselves, if they think "everyone will say Yes," that pushes the image toward the "cat" label. The count of all participants who said "No, I think a majority of others will say no it's not a cat" gets divided by the number of participants.

Confidence Elicitation

Workers give a normal yes/no answer and report how confident they are in that answer, usually on a 0 – 100 scale. The system can then weight high-confidence answers more heavily or use average confidence (for "yes" and "no" answers separately) as features in a classifier.

Example:

Ask: "Is there a cat in this picture? Yes or No. How confident are you in your answer, from 0 to 100?" Someone might say "Yes, 95% confident," while another says "No, 40% confident"; when aggregating, the 95%-confidence "Yes" is treated as more informative. For machine learning models, these "confidence values" are separated into "yes" or "no" first. Calculate an average confidence among the "yes" and an average confidence among the "no". Average confidence is

converted to a decimal value between 0 and 1. Both will be input into the machine learning model.

These methods were discussed in the research paper

[Enhancing Image Classification Capabilities of Crowdsourcing-Based Methods through Expanded Input Elicitation](#)

Official link to the above study can be [accessed here](#).

Some other sources of information regarding crowdsourcing-based image or audio classification methods and/or reliability of "belief elicitation" as a trustworthy means of gaining accurate data

[Mechanisms for belief elicitation without ground truth](#) by Niklas V Lehmann, May 8, 2025

[Eliciting Confidence for Improving Crowdsourced Audio Annotations](#) (ACM Digital Library)