

Welcome!

CATS4ML:

Crowdsourcing Adverse Test Sets for ML

cats4ml.humancomputation.com

Google Research



data is the compass for AI - AI advances where there is data

data quality must be addressed in AI practices especially in the way we evaluate AI

improving evaluation of AI must consider ways to measure variance and capture bias to bring us one step closer to data excellence

to address bias in AI evaluation we propose a novel method for crowdsourcing adverse test sets for ML models (CATS4ML)

TAKE HOME MESSAGE

**FAIR &
UNBIASED
MACHINE LEARNING**

**RIGOROUS
EVALUATION
& ML**



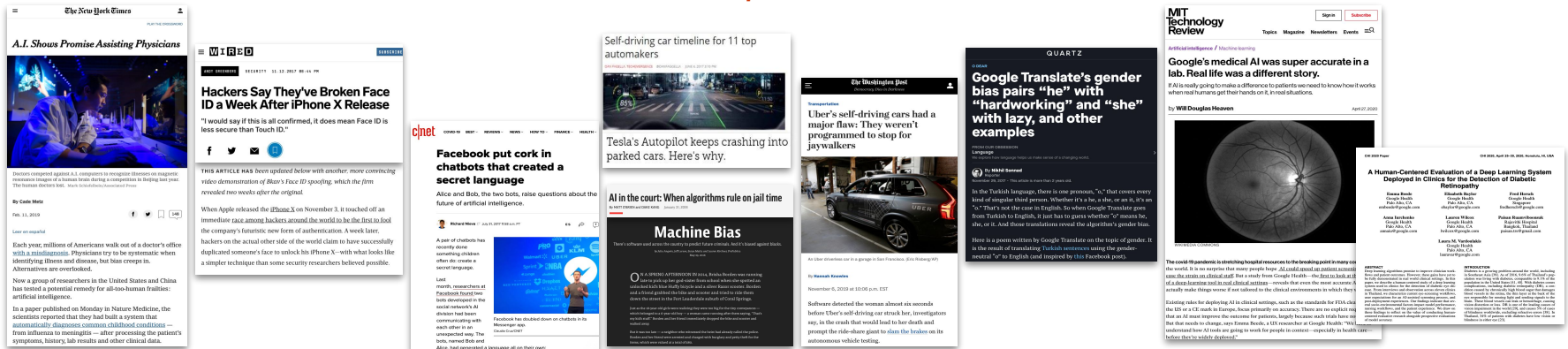
The Life of AI Data



but before it got better ...

“It exists!” → “It is bigger!” → “It is better!”

reactive
data improvement



The Life of AI Data



to reach here

“It exists!” → “It is bigger!” → “It is better!”

*we need proactive
data improvement*

Your AI model is as good as
your **evaluation data**

... but is your evaluation data
missing relevant examples?



cats4ml.humancomputation.com

Your AI model is as good as
your **evaluation data**

... but is your evaluation data
missing relevant examples?

**How can we find such
examples, especially if they are
AI blindspots
(i.e. unknown unknowns)?**



cats4ml.humancomputation.com

CATS4ML Challenge

Crowdsourcing Adverse Test Sets for ML

=

crowdsourced team
for finding **blindspots** of AI models

cats4ml.humancomputation.com

In this first version of the CATS4ML challenge participants will discover **AI blindspots** in the **Open Images Dataset**

These **AI blindspots** are real images with **visual patterns** that confuse AI models
in ways humans might find meaningful

Lipstick?



Thanksgiving?



Construction worker?



Car?



Airplane?



cats4ml.humancomputation.com

Challenge Data from Open Images Dataset

We have selected 1,3M images from the Open Images Dataset, which can be downloaded from the challenge website

<http://bit.ly/cats4ml-data>

```
target_images_large (2).csv
image_id
9625bbe3a59f1688
8fea40e425e6266a
96862c761ebd734b
036e139964de52c7
b1db4ef0e9518a12
1912e83ff784d438
be587ffb4ae2a115
6461546a6fb51da3
c5f70587003b2d4e
25dd27b71d70db51
ee391f2446f37801
c4cfa4d475bb50d1
dab44270a9c79356
bb9328e857b6188a
a415428833bae9c2
403d09165840453d
e6681232185a7fa2
0b8d98b8d4f1eae2
9f5804a36cdd123a
42ee4fe3439ced2b
fa75e3ccd1e5b372
004c0d4a277cc9a7
4a9b34dcadd57bba
98397892729dc07a
860697323ff27286
0bbd1d51fb00274b
143f348fb699ebf1
ebe216ff1f87bce8
6f1e7e79650905c4
01a0508ae2fc1d69
```

We have also selected 24 target labels, for which participants will discover adverse examples

<http://bit.ly/cats4ml-labels>

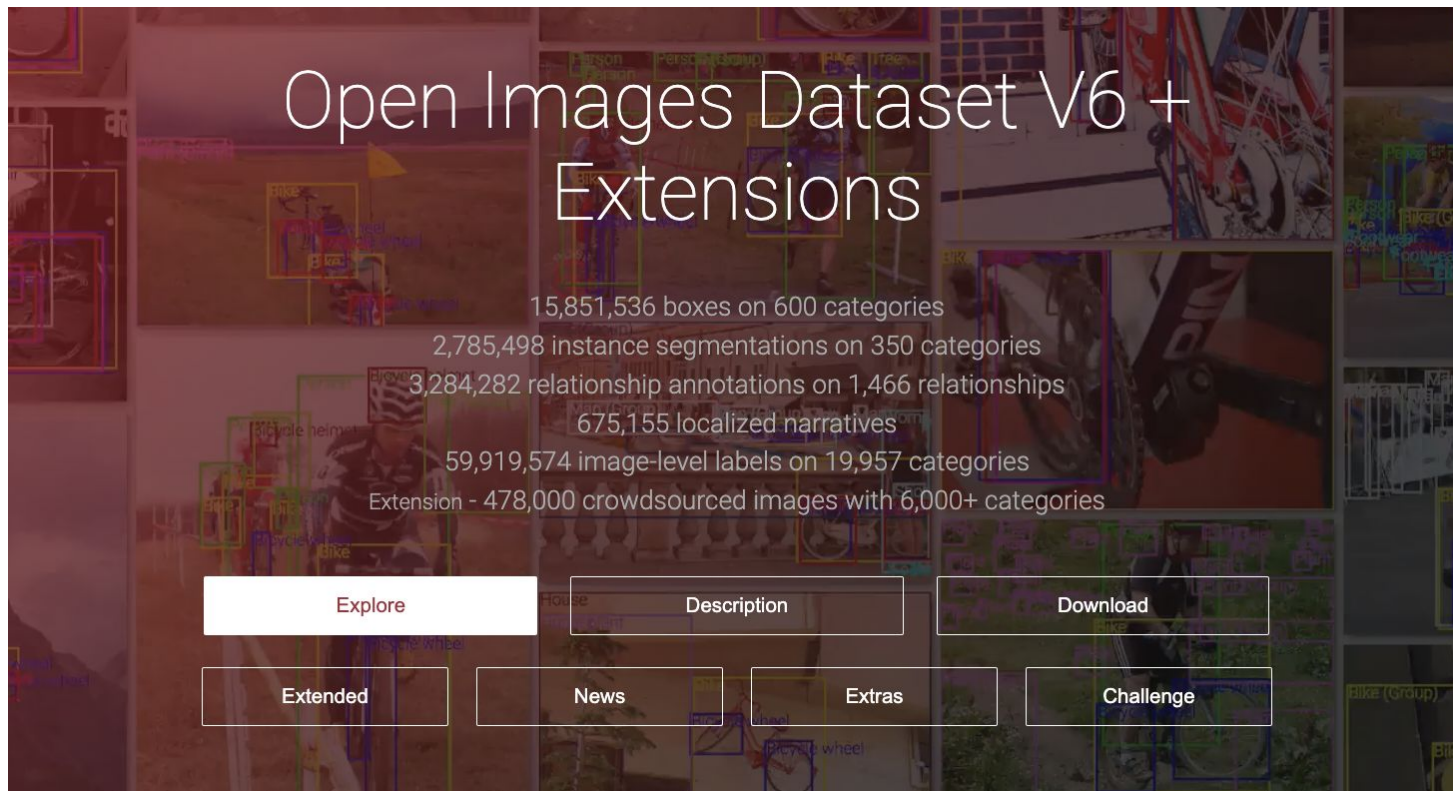
```
target_labels (3).csv
label,display_name
/m/06c7f7,Lipstick
/m/0sqh53y,Selfie
/m/019nj4,Smile
/m/027qtzz,Thanksgiving
/m/02_g0,Funeral
/m/016c3c,Graduation
/m/0ytgt,Child
/m/05t4q,Physician
/m/0fczf,Nurse
/m/01d30f,Teacher
/m/0bk5m9,Bus Driver
/m/012n4x,Firefighter
/m/01pn0r,Chef
/m/047x57,Construction Worker
/m/02y5kn,Coach
/m/01445t,Athlete
/m/0jm,American football
/m/0itc10,Muffin
/m/015wgc,Croissant
/m/0663v,Pizza
/m/0ph39,Canoe
/m/015p6,Bird
/m/01_5g,Chopsticks
```

Some examples of Adverse Images from Open Image Dataset

<https://storage.googleapis.com/openimages/web/index.html>

Examples of adverse examples from Open Images Dataset

You can use the **OID web UI** to explore some of the labels and find manually candidate images



The screenshot shows the Open Images Dataset V6 + Extensions website. The background is a collage of images with various bounding boxes and labels. The main title is "Open Images Dataset V6 + Extensions". Below the title, the following statistics are listed:

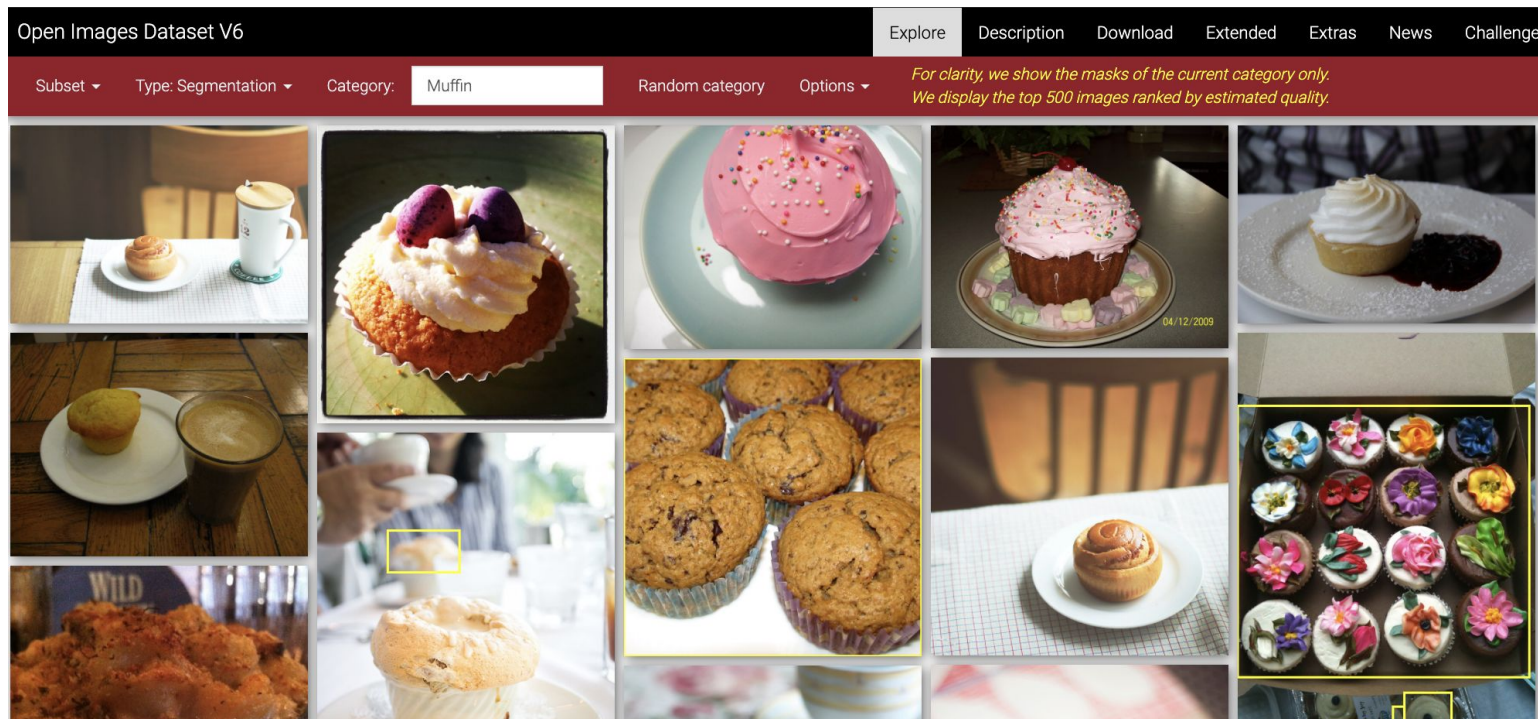
- 15,851,536 boxes on 600 categories
- 2,785,498 instance segmentations on 350 categories
- 3,284,282 relationship annotations on 1,466 relationships
- 675,155 localized narratives
- 59,919,574 image-level labels on 19,957 categories
- Extension - 478,000 crowdsourced images with 6,000+ categories

At the bottom, there are six buttons arranged in two rows:

- Explore (highlighted with a white background)
- Description
- Download
- Extended
- News
- Extras
- Challenge

Using the OID Web UI you can **search for the target labels**

Below you can see image search results for the target label **MUFFIN**

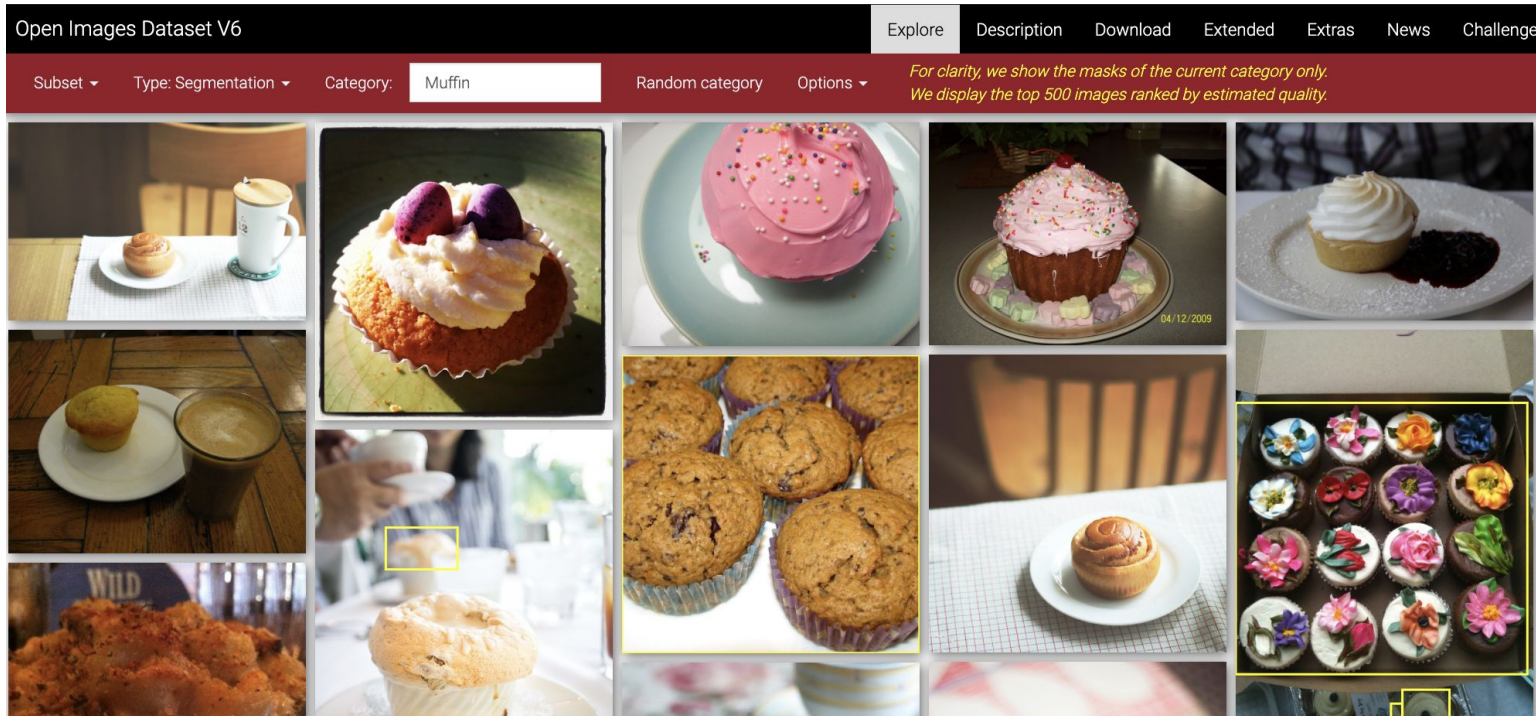


Using the OID Web UI you can **search** for the target labels

Below you can see image search results for the target label **MUFFIN**

Try it yourself:

<https://storage.googleapis.com/openimages/web/visualizer/index.html?set=valtest&type=detection&c=%2Fm%2F01tcjp&id=c6bf6b63014ad396>

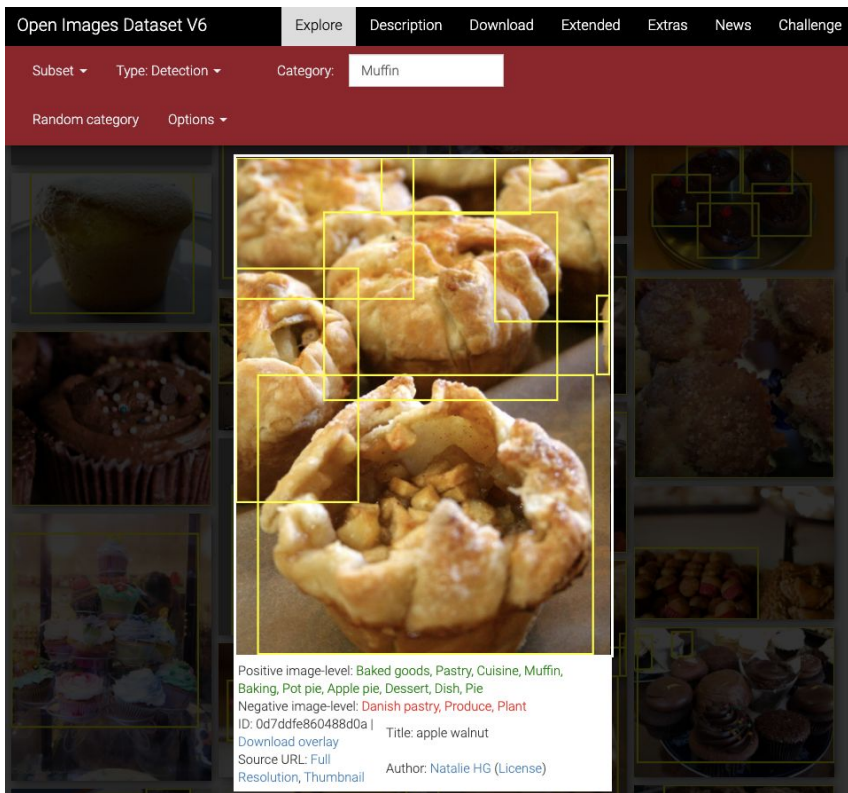


Let's explore some of the search results to find Adverse Images for the target label MUFFIN

<https://storage.googleapis.com/openimages/web/visualizer/index.html?set=valtest&type=detection&c=%2Fm%2F01tcjp&id=c6bf6b63014ad396>

#1: Possible candidate for an adverse example from OID

You can use the OID web UI to explore some of the labels and find manually candidate images, e.g. see the example below for the label “MUFFIN”, which might not contain an actual MUFFIN



Check this image:

(one of the image search results for the label MUFFIN)

https://c3.staticflickr.com/3/2687/4110405062_0273a0d896_o.jpg

#2: Possible candidate for an adverse example from OID


You can use the OID web UI to explore some of the labels and find manually candidate images, e.g. see the example below for the label “MUFFIN”, which might not contain an actual MUFFIN

Open Images Dataset V6

ExploreDescriptionDownloadExtendedExtrasNewsChallenge

Subset ▾Type: Detection ▾Category: Muffin

Random categoryOptions ▾



Positive image-level: Sweetness, Petit four, Chocolate truffle, Rum ball, Snack, Muffin, Cuisine
Negative image-level: Produce, Pastry, Plant, Icing, Baking, Person, Tart
ID: 297938a3233f92e6 | Download overlay Title: rafahinda
Source URL: Full Resolution, Thumbnail Author: mauroguanandi (License)

Check this image:

(one of the image search results for the label MUFFIN)

https://farm7.staticflickr.com/2744/4242818968_5dd1b52f88_o.jpg

#3: Possible candidate for an adverse example from OID


You can use the OID web UI to explore some of the labels and find manually candidate images, e.g. see the example below for the label “MUFFIN”, which might not contain an actual MUFFIN

Open Images Dataset V6

Explore Description Download Extended Extras News Challenge

Subset ▾ Type: Detection ▾ Category: Muffin

Random category Options ▾



Positive image-level: Food, Dessert, Dish, Falafel, Cutlet, Meat chop, Muffin, Curry, Cuisine, Meal, Baked goods, Fried food, Fried chicken

Negative image-level: Person, Meat, Plant

ID: c6bf6b63014ad396 | Download overlay

Source URL: Full Resolution, Thumbnail

Title: fried chicken_greens_pinto beans_mashed potatoes_cornbread

Author: Southern Foodways Alliance (License)

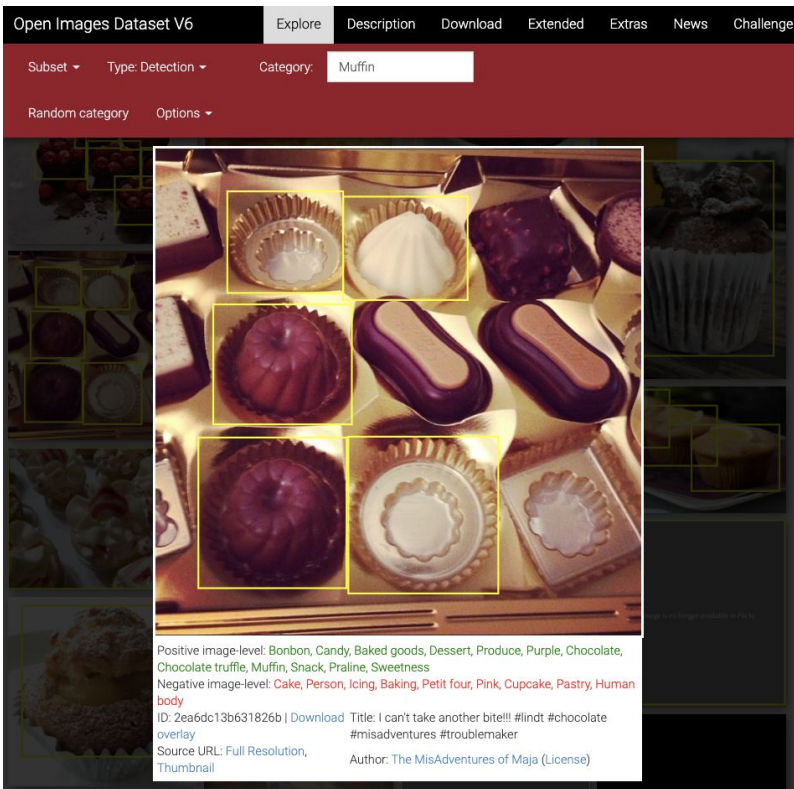
Check this image:

(one of the image search results for the label MUFFIN)

https://farm7.staticflickr.com/1101/5158382341_9384325975_o.jpg

#4: Possible candidate for an adverse example from OID

You can use the OID web UI to explore some of the labels and find manually candidate images, e.g. see the example below for the label “MUFFIN”, which might not contain an actual MUFFIN



Check this image:

(one of the image search results for the label MUFFIN)

https://c4.staticflickr.com/4/3824/11595557444_e6fb94a491_o.jpg

How to Participate in the
CATS4ML Challenge
Crowdsourcing Adverse Test Sets for ML


cats4ml.humancomputation.com

Step 1: Visit <https://cats4ml.humancomputation.com/>

Step 2: Create an account

Email address* Choose password* ☒ I have read and accept the [terms and conditions](#). [SIGN UP](#) [CANCEL](#)

Step 3: Download the data (target labels and images) for the challenge

l.m.arojo@gmail.com [SIGN OUT](#)

Crowdsourcing Adverse Test Sets to Help Surface AI Blindspots

[Home](#)
[Overview](#)
[Participate](#)
Data
[Rules](#)
[Scoring](#)
[Submission](#)
[Queue](#)
[Leaderboard](#)
[Organizers](#)

Data

To keep the competition focused, the participants of the CATS4ML challenge will search for adverse example images only in the target images and for the target labels provided below:

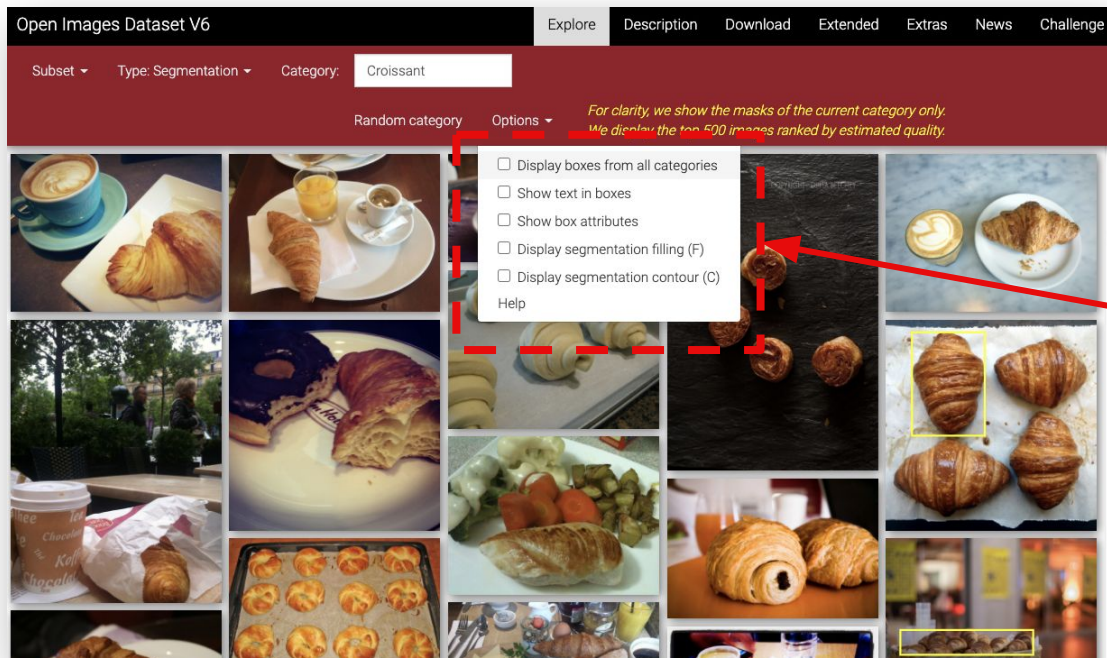
- **Target images:** These are the subset of images from Open Images Dataset (OID). Only image IDs from these target images will be considered for this challenge.
- **Target labels:** a list of the labels for which participants should find adverse examples (AI blindspots).

Check [Participate](#) to learn how to start contributing.

Step 4: Use the OID Web UI to inspect some target labels & their images
bit.ly/OID-WebUI

Step 4a: Uncheck all the display options to see images better on the OID Web UI

Step 4b: Start searching for images for the target labels



Step 5: Study the submission and participation rules on the challenge website

bit.ly/cats4ml-participate

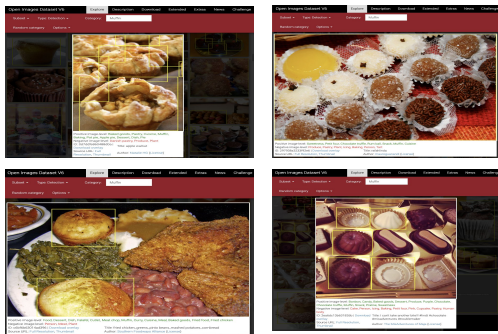
The screenshot shows the CATS4ML website with a navigation menu on the left and a 'Participate' section on the right. Red dashed boxes and arrows highlight specific rules:

- Participate** (Section Header)
- To start participating, sign up and accept the [terms and conditions](#).**
- To participate you need to submit a sample set of images from the [target images](#) for the [target labels](#).**
- Each submission should be a CSV file;**
- Each submission should contain in every row of the CSV file only imageIDs, corresponding target label, and a brief text explaining the discovery rationale;**
- Each submission should contain only image IDs from the [OID dataset](#) and labels from the [target label set](#); Check the [example submission file](#).**
- Synthetically-manipulated images are not accepted;**
- Each submission should not exceed participants **submission quota** (see [Submission](#) for quota explanation);**
- Participants **earn points** by submitting **adverse examples**, which are image-label pairs for which the human verification is in disagreement with a machine prediction, e.g.**
 - human verification = Y, machine prediction = N (false negative);
 - human verification = N, machine prediction = Y (false positive).
- If multiple participants submit the same image-label pair, a point is awarded to the first participant who submits it (based on the timestamp in the submitted images queue)**
- The image-label pairs submitted by all participants will be published on the challenge website and will be visible in the [Queue](#) page;**
- Participants earn **bonus submission quota** with every image-label pair which earned them a point;**
- The [Leaderboard](#) is **updated continuously** as the image-label pairs are verified.**
- Participants are free to use any creative form of sampling methods for discovering images from the OID dataset (including generating synthetic manipulations of the original OID images to identify both false positives and false negatives).**

At the bottom, it states: "This challenge is a co-located event at the [HCOMP2020 Conference](#). Contact us at cats4ml@googlegroups.com. Read [terms and conditions](#) for this challenge."

Step 6: When you found images that you think are adverse examples create a submission file
bit.ly/cats4ml-submit

four images for target label “MUFFIN”



CATS4ML

l.m.arojo@gmail.com SIGN OUT

Crowdsourcing Adverse Test Sets to Help Surface AI Blindspots

Home Overview Participate Data Rules Scoring Submission Queue Leaderboard Organizers

Submission

submission quota

Your remaining quota is 1000. Participants will submit sets of image-label pairs in this format - [example submission file](#)

Check [Participate](#) to learn how to start contributing.

- A submitted set can contain 1 or more image-label pairs not exceeding participants' submission quota;
- Each submission set should contain only image-ID and target-label pairs from the [target images](#) (subset of [OID](#)) and the [target labels](#) set;
- Additionally participants will include a brief free text explanation of their discovery rationale and motivation for every image-label pair;
- Participants can submit multiple times within their submission quota;


Submit a new submission:

This challenge is a co-located event at the [HCOMP2020 Conference](#).
Contact us at cats4ml@googlegroups.com
[Read terms and conditions for this challenge](#)

example submission file for the four images

```
CATS4ML labels - test-submission.csv — Edited
image_id,label,rationale
2ea6dc13b631826b,/m/01tcip,"i used OID WebUI to search for images of muffins; this picture does not depict a muffin, these are other deserts"
c6bf6b63014ad396,/m/01tcip,"i used OID WebUI to search for images of muffins; this picture does not depict a muffin, these are similar shaped other food items"
41f541535d5e80da,/m/01tcip,"i used OID WebUI to search for images of muffins; this picture does not depict a muffin, these are apple crumbles"
297938a3233f92e6,/m/01tcip,"i used OID WebUI to search for images of muffins; this picture does not depict a muffin, these are chocolate balls"
```

Step 7: Once you submit you will wait until the leaderboard updates with your score
bit.ly/cats4ml-leaderboard

l.m.aroyo@gmail.com [SIGN OUT](#)

Crowdsourcing Adverse Test Sets to Help Surface AI Blindspots

[Home](#)
[Overview](#)
[Participate](#)
[Data](#)
[Rules](#)
[Scoring](#)
[Submission](#)
[Queue](#)
[Leaderboard](#)
[Organizers](#)

Leaderboard

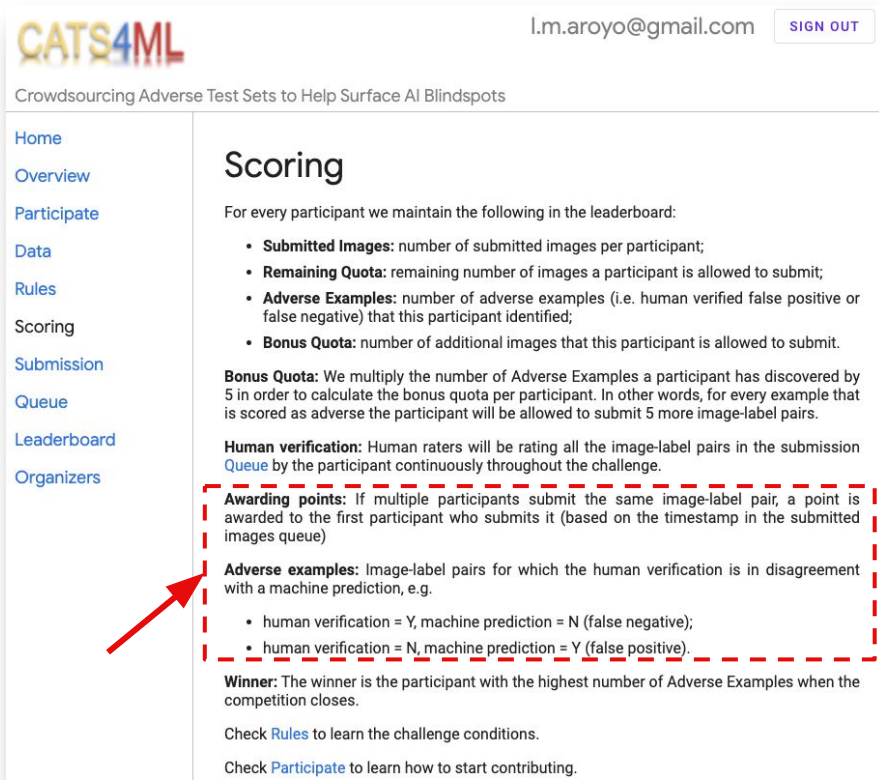
This page maintains the current state of the ranks for all participants based on their scores. With every new submission participants can improve their position in the leaderboard and earn a new bonus quota.

Last updated at 01/12/2021

#	User ID	Score	Bonus Quota	Remaining Quota
---	---------	-------	-------------	-----------------

Step 8: You score points for each image you submitted where the human raters **disagree** with the machine label

bit.ly/cats4ml-scoring



CATS4ML l.m.aroyo@gmail.com [SIGN OUT](#)

Crowdsourcing Adverse Test Sets to Help Surface AI Blindspots

- [Home](#)
- [Overview](#)
- [Participate](#)
- [Data](#)
- [Rules](#)
- [Scoring](#)
- [Submission](#)
- [Queue](#)
- [Leaderboard](#)
- [Organizers](#)

Scoring

For every participant we maintain the following in the leaderboard:

- **Submitted Images:** number of submitted images per participant;
- **Remaining Quota:** remaining number of images a participant is allowed to submit;
- **Adverse Examples:** number of adverse examples (i.e. human verified false positive or false negative) that this participant identified;
- **Bonus Quota:** number of additional images that this participant is allowed to submit.

Bonus Quota: We multiply the number of Adverse Examples a participant has discovered by 5 in order to calculate the bonus quota per participant. In other words, for every example that is scored as adverse the participant will be allowed to submit 5 more image-label pairs.

Human verification: Human raters will be rating all the image-label pairs in the submission [Queue](#) by the participant continuously throughout the challenge.

Awarding points: If multiple participants submit the same image-label pair, a point is awarded to the first participant who submits it (based on the timestamp in the submitted images queue)

Adverse examples: Image-label pairs for which the human verification is in disagreement with a machine prediction, e.g.

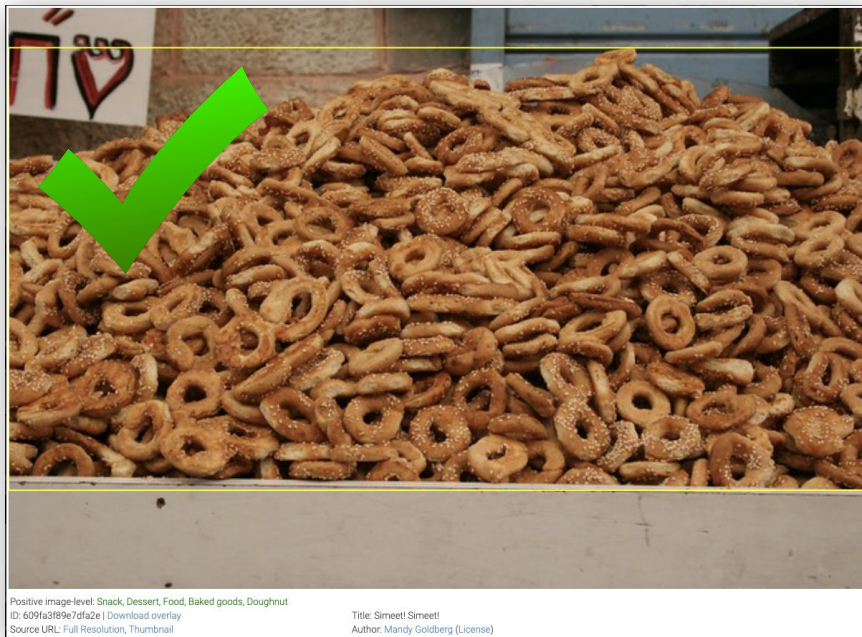
- human verification = Y, machine prediction = N (false negative);
- human verification = N, machine prediction = Y (false positive).

Winner: The winner is the participant with the highest number of Adverse Examples when the competition closes.

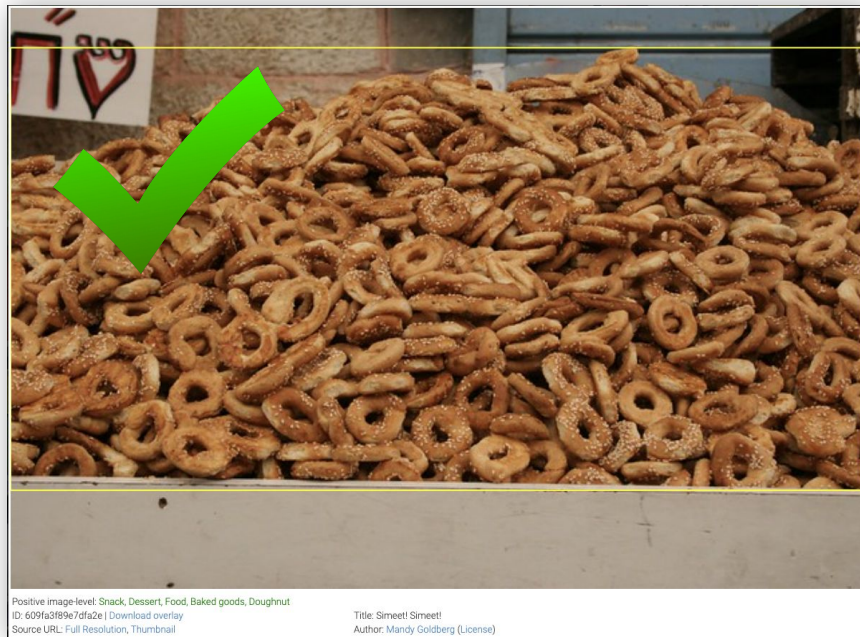
Check [Rules](#) to learn the challenge conditions.

Check [Participate](#) to learn how to start contributing.

Step 8: You score points for each image you submitted where the human raters **disagree** with the machine label
bit.ly/cats4ml-scoring



Human raters = **NO** for DOUGHNUT
Machine score = **YES** for DOUGHNUT



Human raters = **YES** for BAGEL
Machine score = **NO** for BAGEL

Step 8: You score points for each image you submitted where the human raters **disagree** with the machine label
bit.ly/cats4ml-scoring



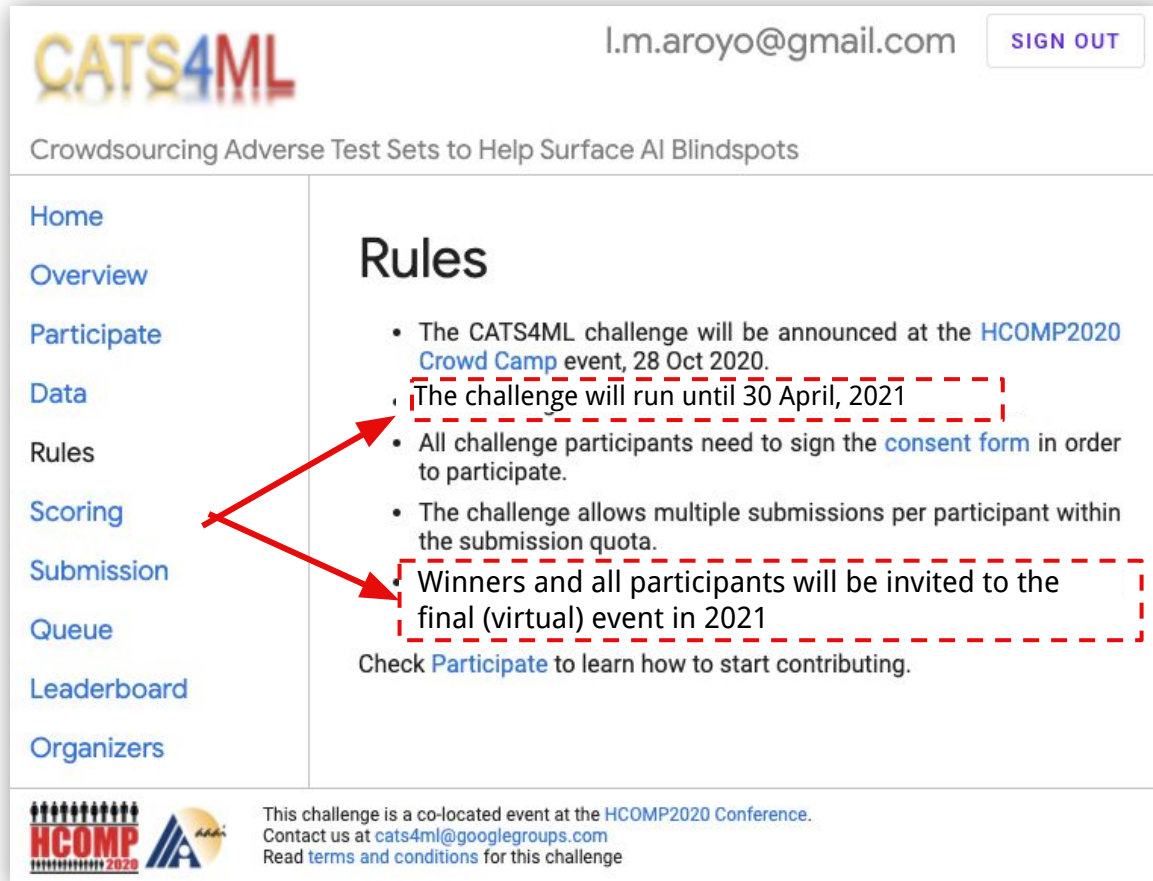
Human raters = **NO** for MUFFIN
Machine score = **YES** for MUFFIN



Human raters = **YES** for MUFFIN
Machine score = **YES** for MUFFIN

Step 9: The challenge allows for continuous multiple submissions. So, keep submitting every time you find new images

bit.ly/cats4ml-submit



CATS4ML l.m.aroyo@gmail.com [SIGN OUT](#)



Crowdsourcing Adverse Test Sets to Help Surface AI Blindspots

- [Home](#)
- [Overview](#)
- [Participate](#)
- [Data](#)
- [Rules](#)
- [Scoring](#)
- [Submission](#)
- [Queue](#)
- [Leaderboard](#)
- [Organizers](#)

Rules

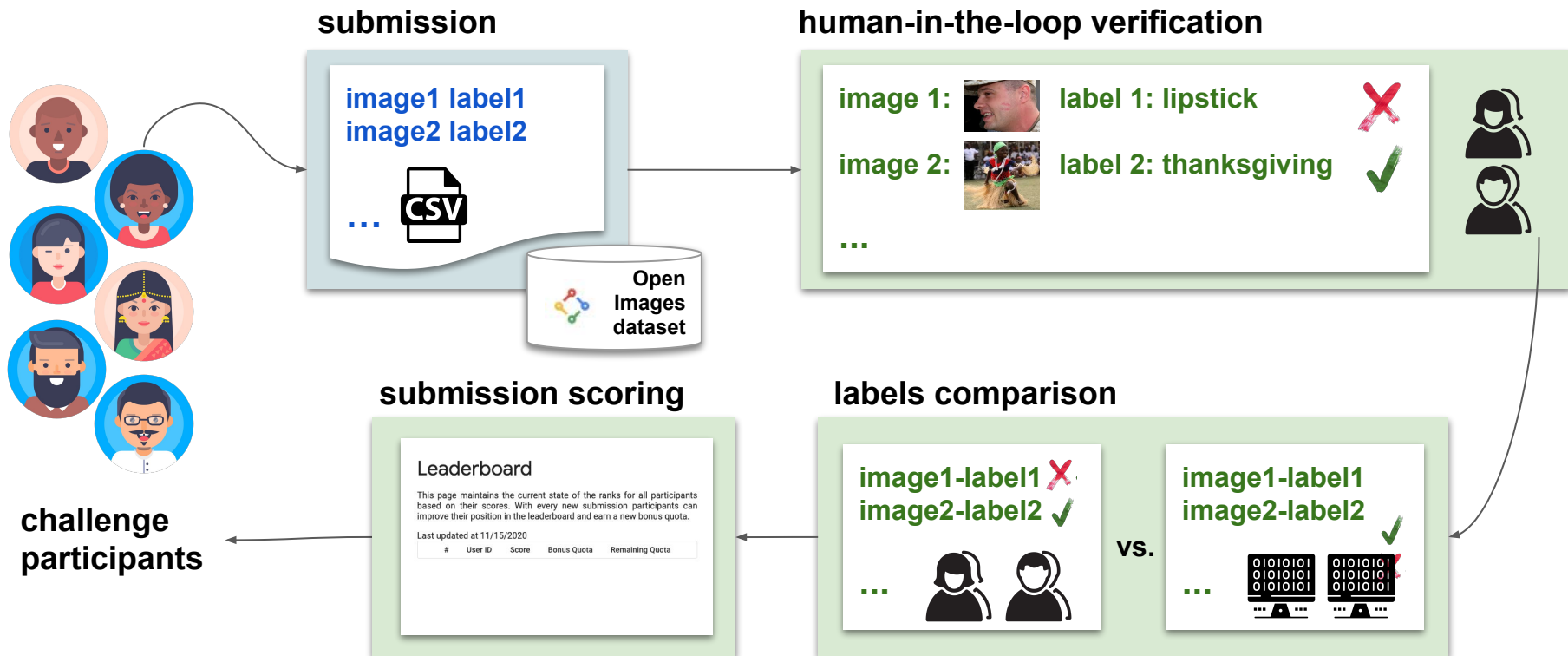
- The CATS4ML challenge will be announced at the [HCOMP2020 Crowd Camp](#) event, 28 Oct 2020.
- The challenge will run until 30 April, 2021
- All challenge participants need to sign the [consent form](#) in order to participate.
- The challenge allows multiple submissions per participant within the submission quota.
- Winners and all participants will be invited to the final (virtual) event in 2021

Check [Participate](#) to learn how to start contributing.

  This challenge is a co-located event at the [HCOMP2020 Conference](#).
Contact us at cats4ml@googlegroups.com
Read [terms and conditions](#) for this challenge

cats4ml.humancomputation.com

Behind the scenes



cats4ml.humancomputation.com

Join us now!

- build a team or join individually:
bit.ly/cats4ml
- multiple submissions from each participant allowed until **30 April, 2021**
bit.ly/cats4ml-faq
- contribute your creative adverse examples from the Open Image Dataset
- spread the word to your research community
- give us feedback
cats4ml@googlegroups.com

cats4ml.humancomputation.com



The Challenge Team



Lora
Aroyo



Praveen
Paritosh



Ka
wong



Tong
Zhou



Mig
Gerard



Kenny
Wibowo

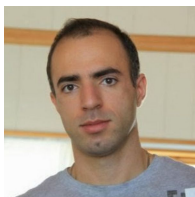


Igor
Karpov

in collaboration with



Ken
Burke



Shahab
Kamali



Google Research

Join CATS4ML challenge

bit.ly/cats4ml

bit.ly/cats4ml-faq

cats4ml@googlegroups.com



Praveen Paritosh



Ka Wong



Lora Aroyo



Devi Krishna