

# UW A3D3 ML Hackathon



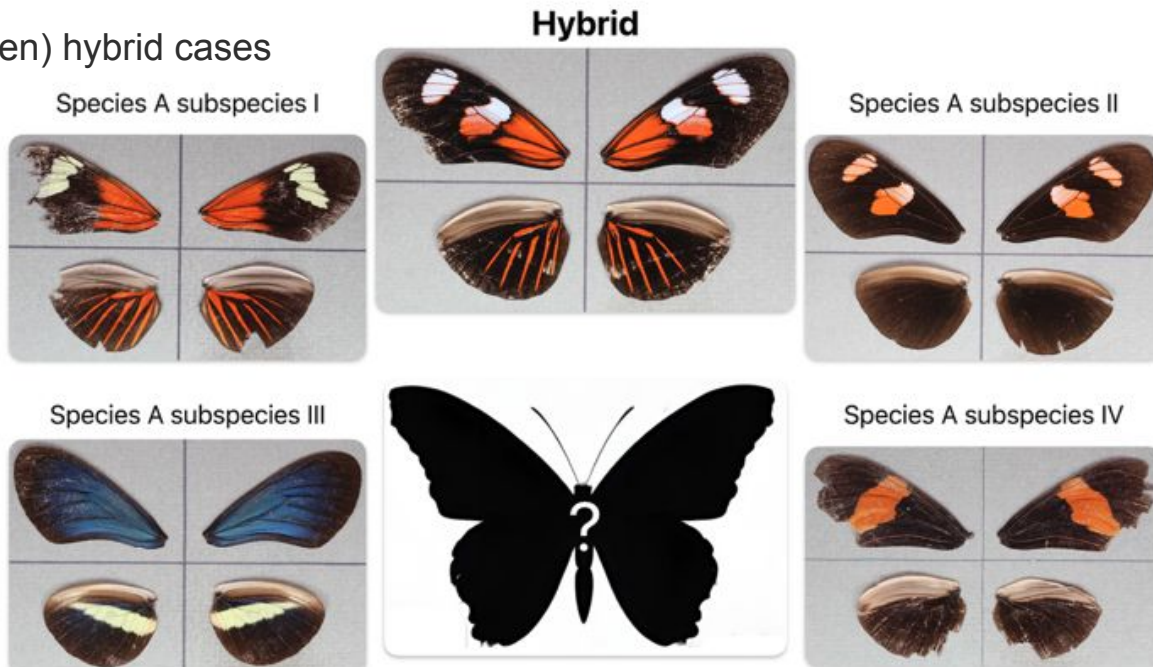
**NSF HDR ML Challenge**

# Agenda

1. Introduction to three challenges
  - a. Butterfly Hybrid Detection
  - b. iHARP HDR Anomaly Challenge (Sea level anomaly detection)
  - c. Detecting Anomalous Gravitational Wave
2. Submission Platform
3. Example submission with Detecting Anomalous Gravitational Wave

# Butterfly Hybrid Detection

Automatically identify (unseen) hybrid cases



Adapted from slides: Elizabeth G. Campolongo and The  
Imageomics ML Challenge Team

# Training Data

~2200 images of Species A:

- Multiple subspecies.
- Selected signal hybrids of two subspecies.

# Dev & Test Data Description

Includes:

- All Species A subspecies.
- Signal hybrids from training data.

Further introduces:

- Other Species A hybrids (non-signal).
- Species B: Mimics of Species A signal hybrid parents (& their hybrids).

The numbers:

- Validation Data (Dev): ~1100 images
- Test Data: ~2200 images

# Butterfly Hybrid Detection

Among Species A & B, can your algorithm find...

- Species A signal hybrids?
- Species A non signal hybrids?
- Species B hybrids (mimics of Species A signal hybrids)?

Competition Page: <https://www.codabench.org/competitions/3764/>

Git Repo: <https://github.com/Imageomics/HDR-anomaly-challenge/tree/main?tab=readme-ov-file>

More dataset description:

<https://github.com/Imageomics/HDR-anomaly-challenge/blob/main/pages/overview.md>

Start kit: [https://github.com/Imageomics/HDR-anomaly-challenge/blob/main/pages/starting\\_kit.md](https://github.com/Imageomics/HDR-anomaly-challenge/blob/main/pages/starting_kit.md)



# Detecting Anomalous Sea Level Rise Events



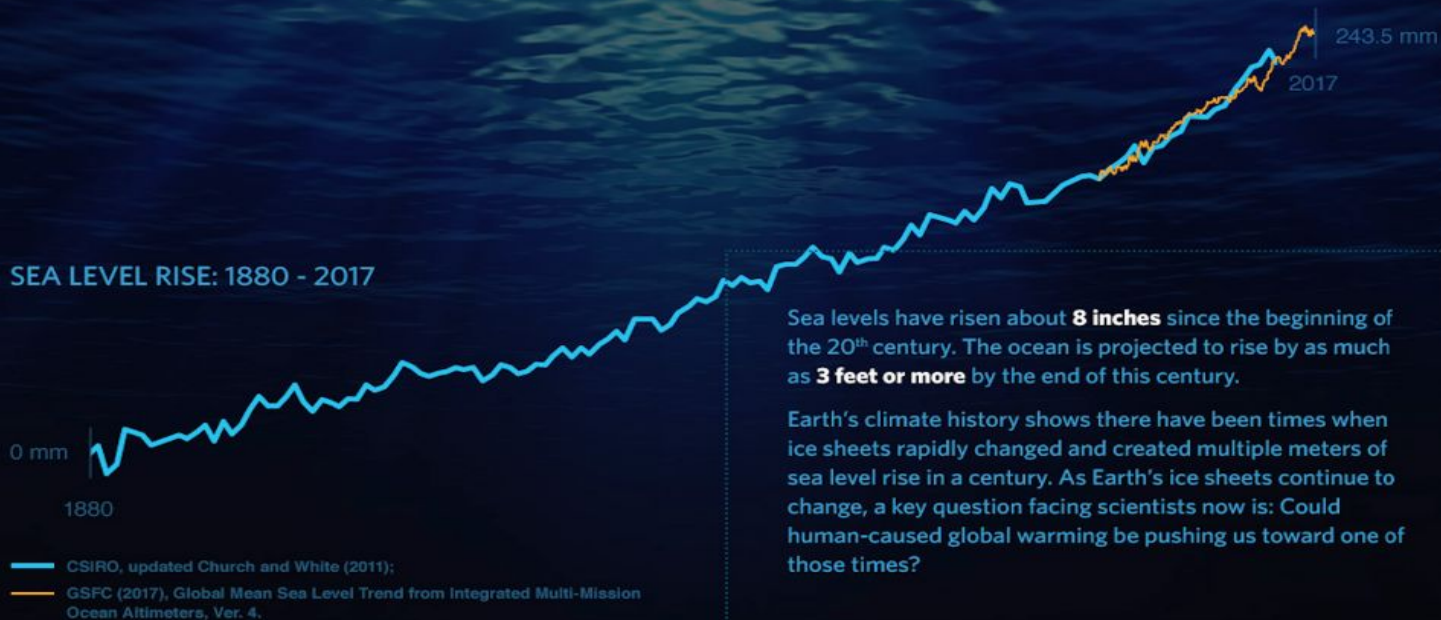
Adapted from slides: **Subhankar Ghosh & Aneesh Subramanian w/ Shashi Shekhar, Vandana Janeja, Josephine Namayanja**





# AS OUR OCEAN WARMS, SEA LEVEL RISES

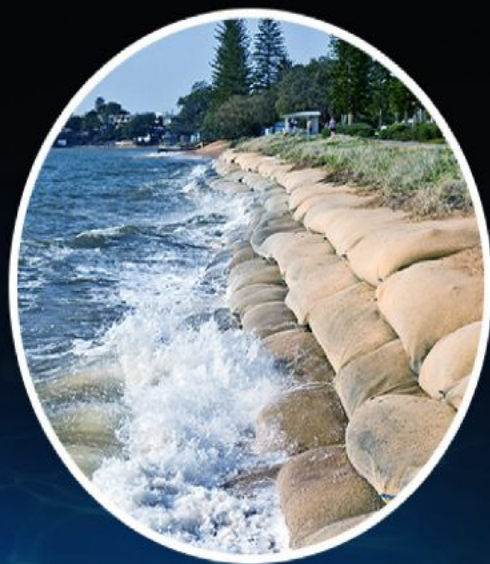
We know seas are rising and we know why. The urgent questions are by how much and how quickly.





# SEA LEVEL RISE AFFECTS US ALL

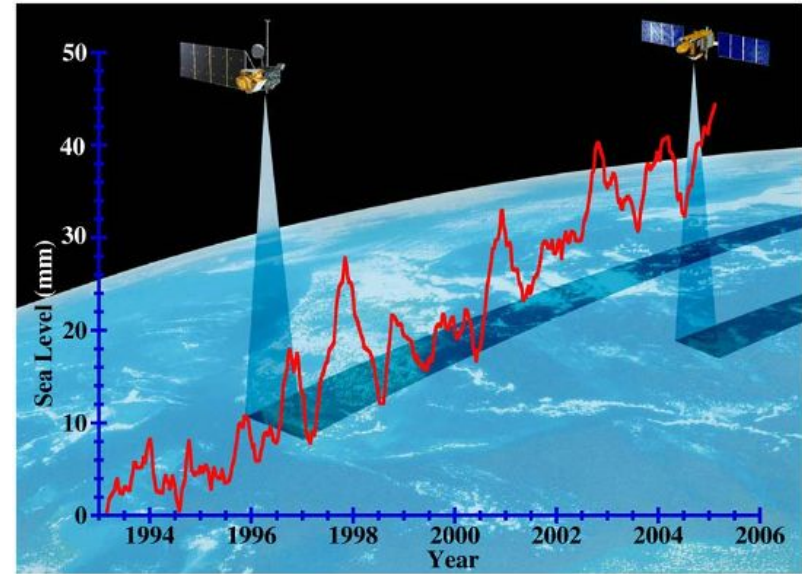
More than **160 million people** live along coasts in the U.S., about half the nation's population. **Eleven of the world's 15 largest cities** lie along shores, including New York City. Sea level rise means the ocean will gradually inundate low-lying areas, and storms like hurricanes, bolstered by even higher seas, will extend their reach inland. All of society bears the burden for storm damage and those costs are expected to rise: Annual losses from flooding in the world's biggest coastal cities could rise from about **\$6 billion a year** today to **\$1 trillion a year** by 2050.



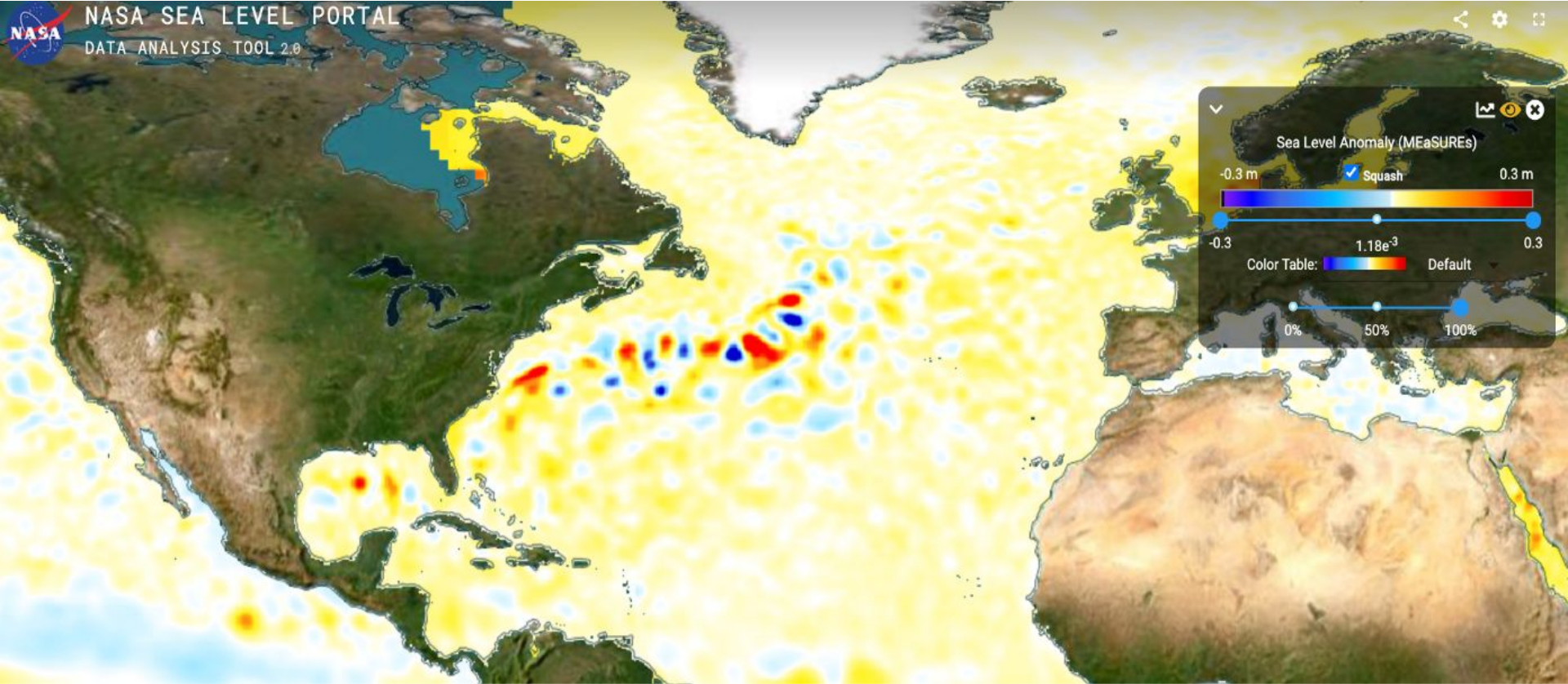
# Sea Level Rise Prediction

## Continual Tracking

- Motivation: understand **how** and **why** sea level is changing – > plans for adaptation
- Tools: Monitor and understand individual factors contributing to sea level rise
  - Satellites to track global ocean level and ice sheet thickness



# Detect anomalous flooding events from satellite sea level maps



# Challenge

**Goal:** Detect anomalous flooding events along the US East Coast with the maps of sea level over the North Atlantic.

## Training data

**Maps of Sea Level:** Satellite sea level anomaly data over the North Atlantic for the past 30 years.

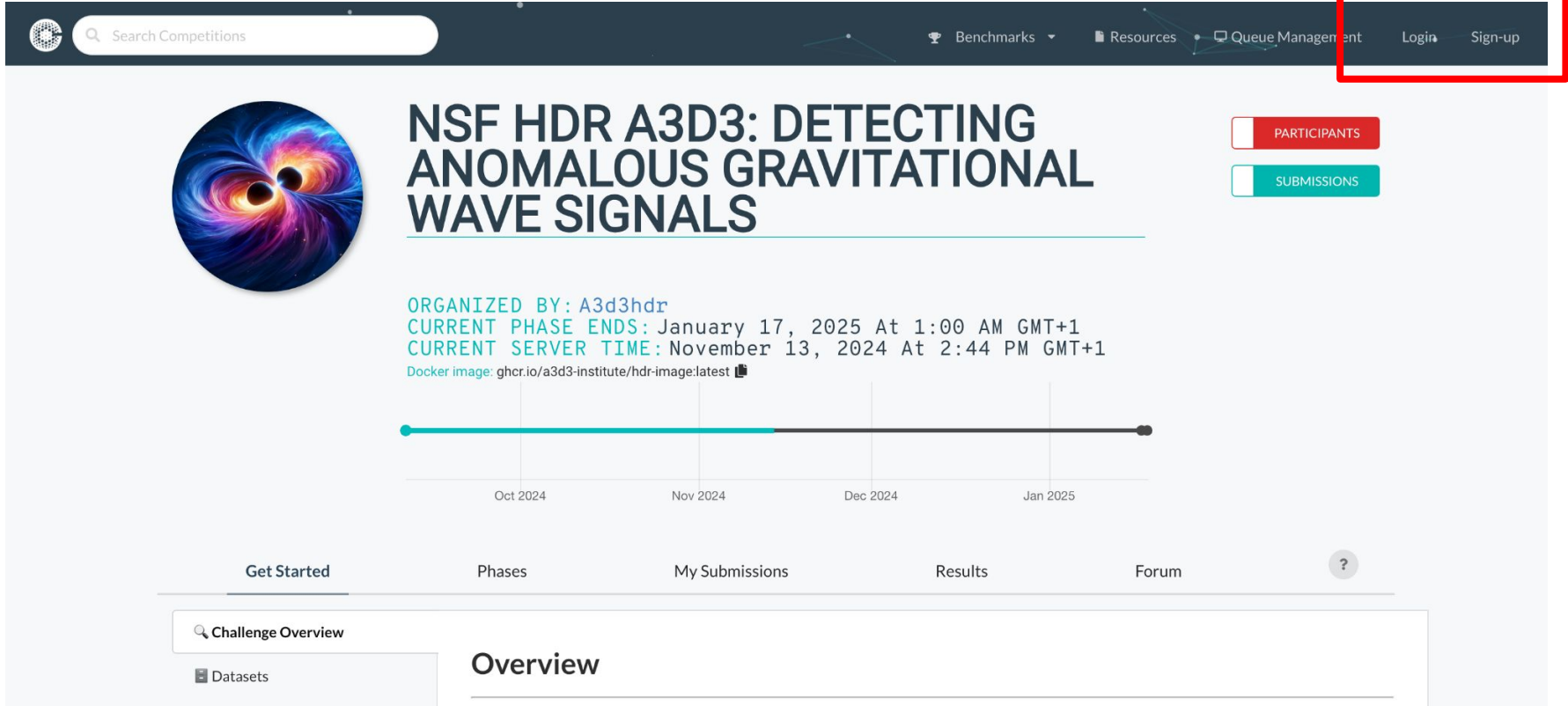
**Labeled Anomalous Flood:** Dates of anomalous flooding along the US East Coast stations for the past 30 years.

# Agenda

1. Introduction to three challenges
  - a. Butterfly Hybrid Detection
  - b. iHARP HDR Anomaly Challenge
  - c. Detecting Anomalous Gravitational Wave
2. Submission Platform
3. Example submission with Detecting Anomalous Gravitational Wave



# 1. Login or Create Account on Codabench



The screenshot displays the Codabench website interface for a specific challenge. At the top, a dark navigation bar contains a search bar on the left and several menu items on the right: 'Benchmarks', 'Resources', 'Queue Management', 'Login', and 'Sign-up'. A red rectangular box highlights the 'Login' and 'Sign-up' buttons, with a red arrow pointing downwards towards it from the top right of the image.

The main content area features a large circular image of a gravitational well on the left. To its right, the challenge title is prominently displayed: **NSF HDR A3D3: DETECTING ANOMALOUS GRAVITATIONAL WAVE SIGNALS**. Below the title, there are two buttons: a red 'PARTICIPANTS' button and a teal 'SUBMISSIONS' button.

Organizational and timing information is provided below the title:

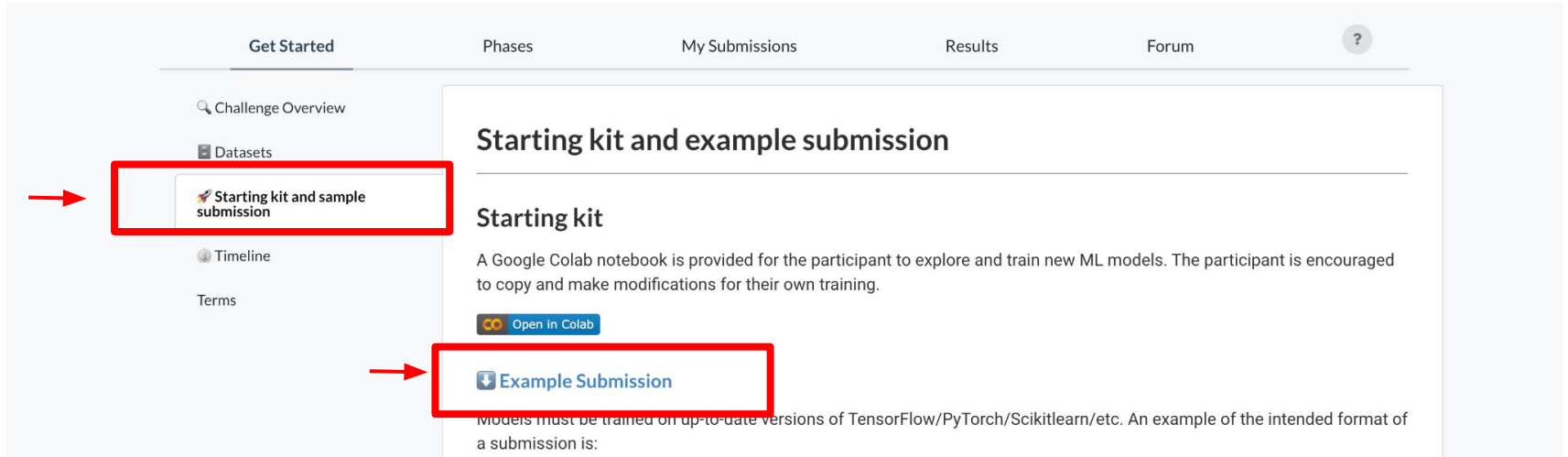
- ORGANIZED BY: A3d3hdr
- CURRENT PHASE ENDS: January 17, 2025 At 1:00 AM GMT+1
- CURRENT SERVER TIME: November 13, 2024 At 2:44 PM GMT+1
- Docker image: ghcr.io/a3d3-institute/hdr-image:latest

A horizontal timeline at the bottom of the challenge details shows the progression from October 2024 to January 2025. The current phase is indicated by a teal bar extending from the start of the timeline to the end of November 2024.

At the bottom of the page, a navigation bar includes links for 'Get Started', 'Phases', 'My Submissions', 'Results', and 'Forum', along with a help icon (question mark).

Below the navigation bar, a search bar contains the text 'Challenge Overview', and a 'Datasets' button is visible on the left side of the page.

## 2. Download Dummy Submission



The screenshot shows a web interface with a top navigation bar containing 'Get Started', 'Phases', 'My Submissions', 'Results', and 'Forum'. A help icon is on the right. A left sidebar contains 'Challenge Overview', 'Datasets', 'Starting kit and sample submission', 'Timeline', and 'Terms'. The main content area has the heading 'Starting kit and example submission' and a sub-heading 'Starting kit'. Below this is a paragraph about a Google Colab notebook and an 'Open in Colab' button. At the bottom, there is an 'Example Submission' link and a partially visible paragraph about training models.

Get Started   Phases   My Submissions   Results   Forum   ?

Challenge Overview

Datasets

Starting kit and sample submission

Timeline

Terms

### Starting kit and example submission

#### Starting kit

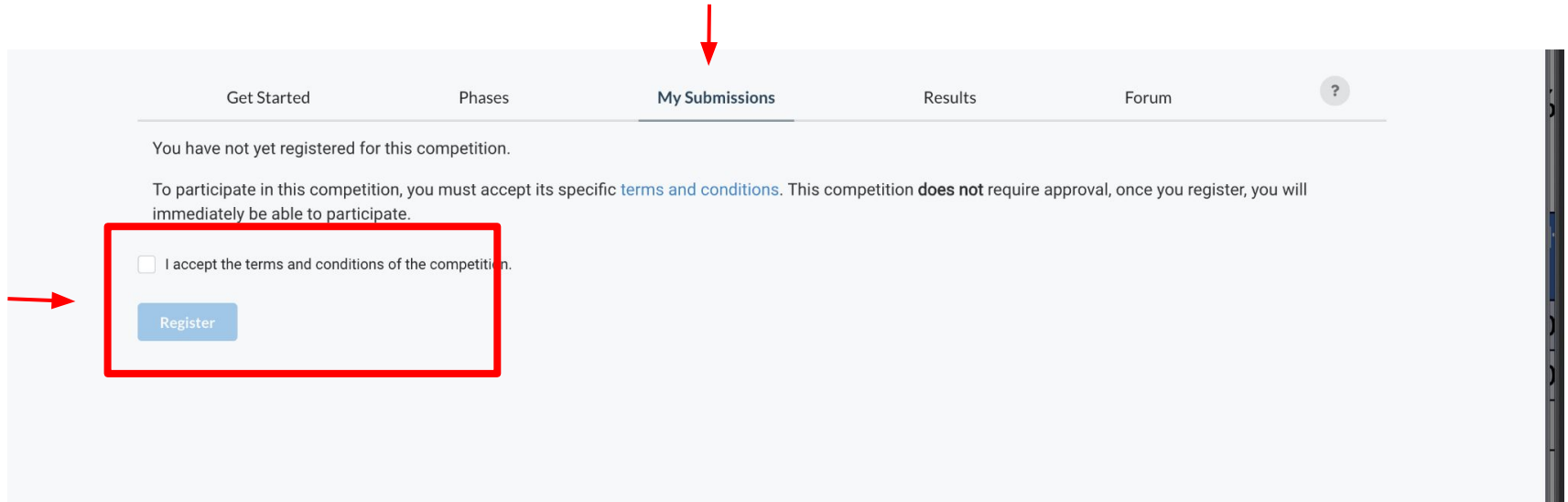
A Google Colab notebook is provided for the participant to explore and train new ML models. The participant is encouraged to copy and make modifications for their own training.

[Open in Colab](#)

[Example Submission](#)

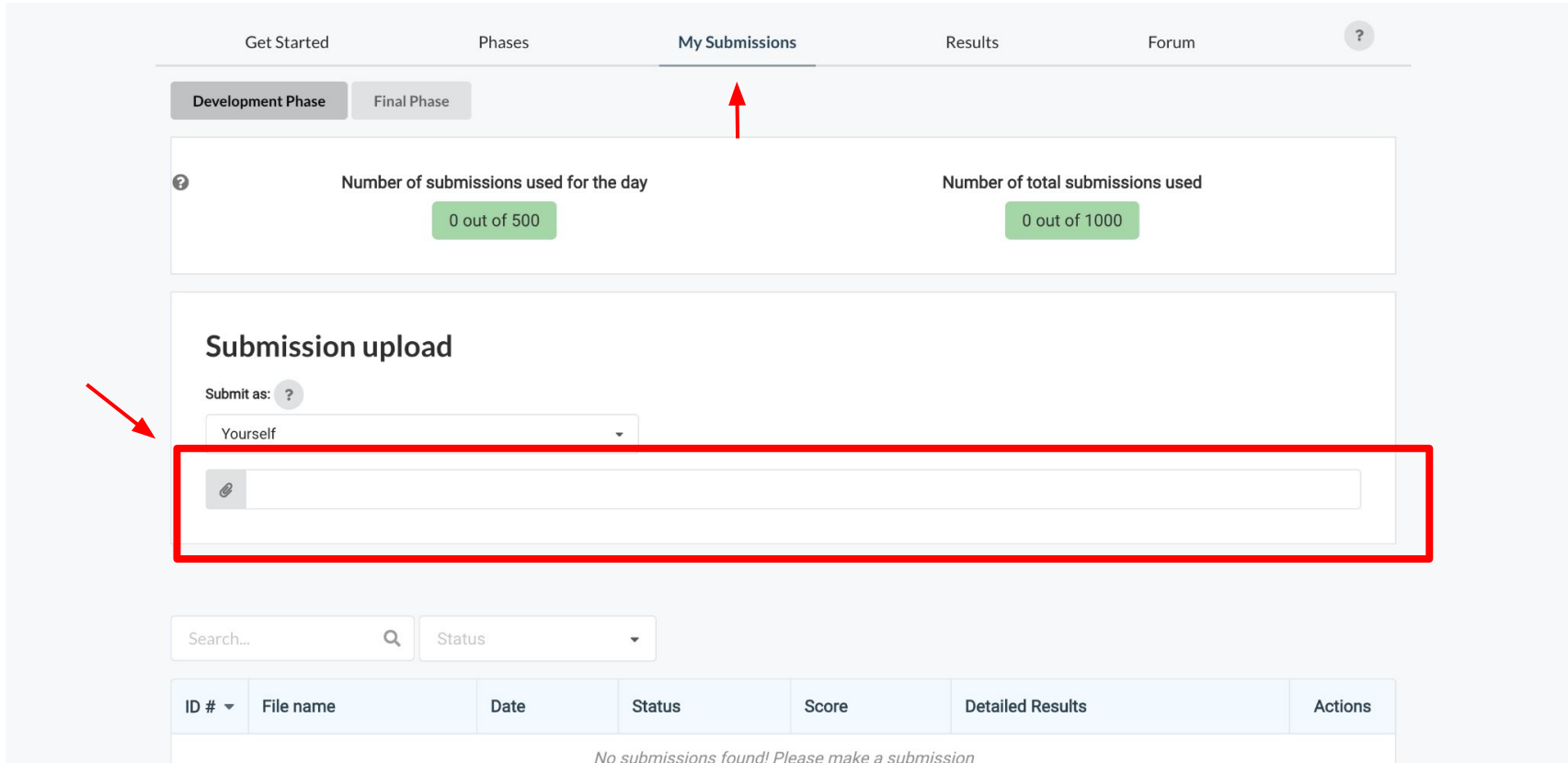
Models must be trained on up-to-date versions of TensorFlow/PyTorch/Scikitlearn/etc. An example of the intended format of a submission is:

### 3. Register in the Competition



The screenshot shows a navigation bar with the following items: "Get Started", "Phases", "My Submissions" (highlighted with a red arrow), "Results", "Forum", and a help icon (a question mark in a circle). Below the navigation bar, the text reads: "You have not yet registered for this competition." followed by "To participate in this competition, you must accept its specific [terms and conditions](#). This competition **does not** require approval, once you register, you will immediately be able to participate." A red box highlights the registration form, which includes a checkbox with the text "I accept the terms and conditions of the competition." and a blue "Register" button. A red arrow points to the checkbox from the left.

## 4. Submit Dummy Submission



The screenshot shows a web interface for managing submissions. At the top, there are navigation tabs: "Get Started", "Phases", "My Submissions", "Results", and "Forum". A red arrow points to the "My Submissions" tab. Below the tabs, there are two buttons: "Development Phase" and "Final Phase".

Below the buttons, there are two summary cards:

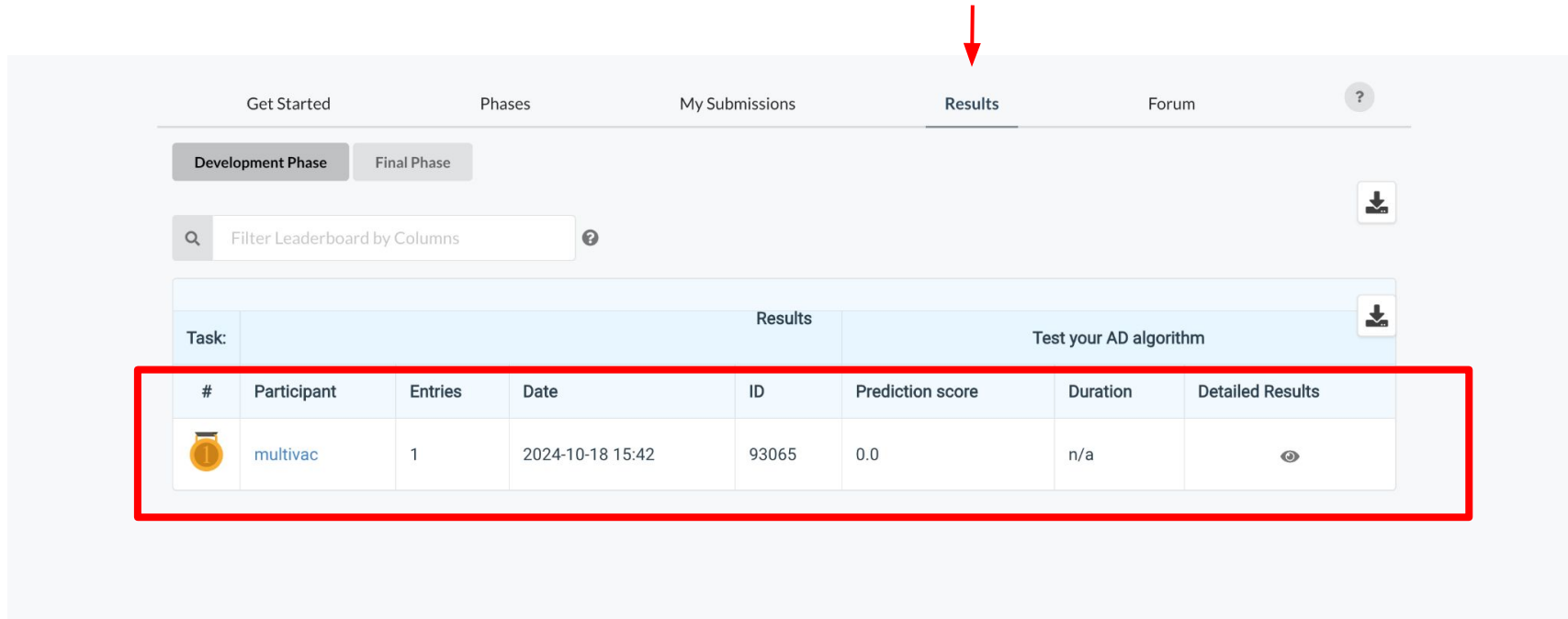
- Number of submissions used for the day:** 0 out of 500
- Number of total submissions used:** 0 out of 1000

The main section is titled "Submission upload". It includes a "Submit as:" dropdown menu with "Yourself" selected. A red arrow points to this dropdown. Below the dropdown is a large file upload field, which is highlighted with a red border. The field contains a paperclip icon and a text input area.



At the bottom, there is a search bar with "Search..." and a magnifying glass icon, and a "Status" dropdown menu. Below these is a table with the following columns:

ID #	File name	Date	Status	Score	Detailed Results	Actions
<i>No submissions found! Please make a submission</i>						

## 5. Check results in the leaderboard



The screenshot shows a web interface with a navigation bar at the top containing 'Get Started', 'Phases', 'My Submissions', 'Results', and 'Forum'. A red arrow points to the 'Results' tab. Below the navigation bar, there are two buttons: 'Development Phase' and 'Final Phase'. A search bar labeled 'Filter Leaderboard by Columns' is present. The main content area features a table with a light blue header. The table has a 'Task' column and a 'Results' column. The 'Task' column contains the text 'Test your AD algorithm'. The 'Results' column contains a table with the following data:

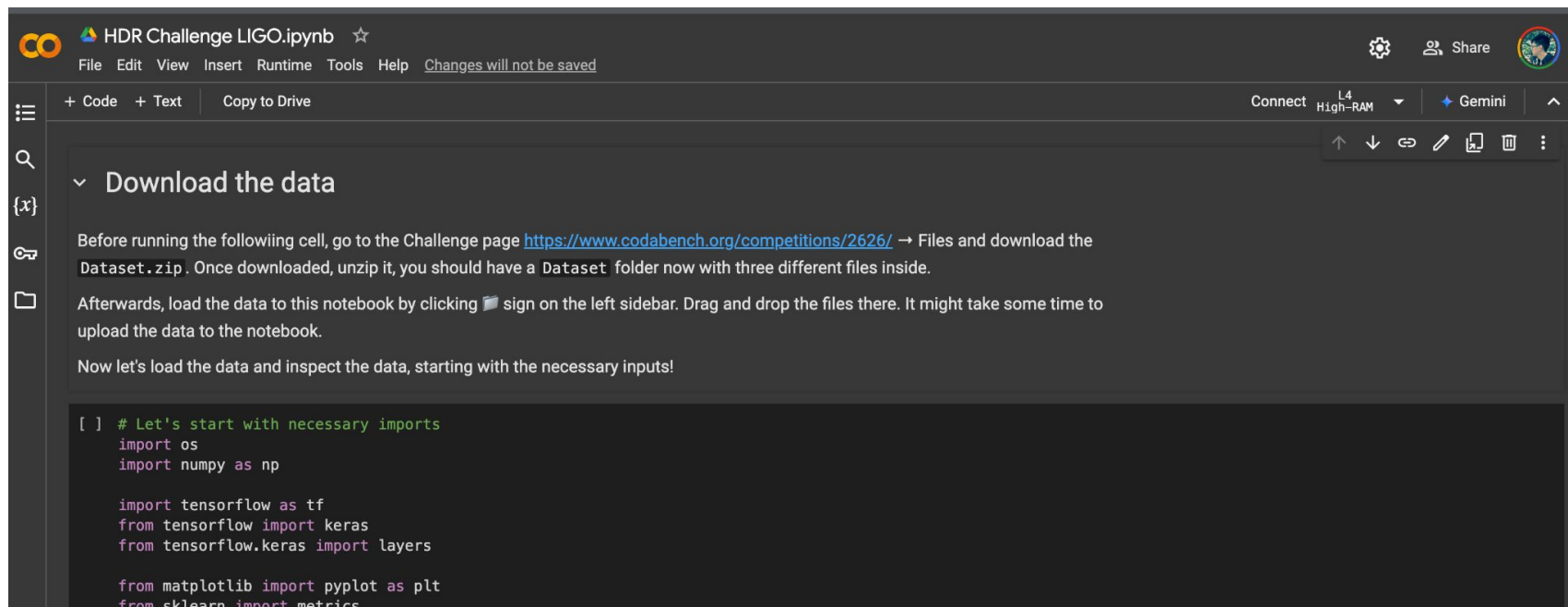
#	Participant	Entries	Date	ID	Prediction score	Duration	Detailed Results
	<a href="#">multivac</a>	1	2024-10-18 15:42	93065	0.0	n/a	



## 6. Check out the starting kit

The screenshot shows a web interface for a challenge. The top navigation bar includes 'Get Started', 'Phases', 'My Submissions', 'Results', and 'Forum'. A sidebar on the left contains a search bar and several menu items: 'Challenge Overview', 'Datasets', 'Starting kit and sample submission' (highlighted with a red arrow), 'Timeline', 'Terms', and 'Files'. The main content area is titled 'Starting kit and example submission'. Under the 'Starting kit' sub-heading, there is a paragraph explaining that a Google Colab notebook is provided for training ML models. Below this text is a blue button with the Colab logo and the text 'Open in Colab', which is highlighted with a red rectangular box. Further down, there is a section for 'Example Submission' with a paragraph of instructions and a code block containing the line `import tensorflow as tf`.


# 7. Starting kit as a Google Colab Notebook



The screenshot shows a Google Colab notebook interface. At the top, the title is "HDR Challenge LIGO.ipynb" with a star icon. Below the title is a menu bar with options: File, Edit, View, Insert, Runtime, Tools, Help. A status bar indicates "Changes will not be saved". On the right side, there are icons for settings, share, and a globe. Below the menu bar, there are tabs for "+ Code", "+ Text", and "Copy to Drive". On the far right, there are options for "Connect" (with "L4 High-RAM" selected), "Gemini", and an upward arrow. The main content area has a search icon, a "{x}" icon, a key icon, and a folder icon. The text in the notebook reads:

Download the data

Before running the following cell, go to the Challenge page <https://www.codabench.org/competitions/2626/> → Files and download the Dataset.zip. Once downloaded, unzip it, you should have a Dataset folder now with three different files inside.

Afterwards, load the data to this notebook by clicking  sign on the left sidebar. Drag and drop the files there. It might take some time to upload the data to the notebook.

Now let's load the data and inspect the data, starting with the necessary inputs!

```
[ ] # Let's start with necessary imports
import os
import numpy as np

import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers

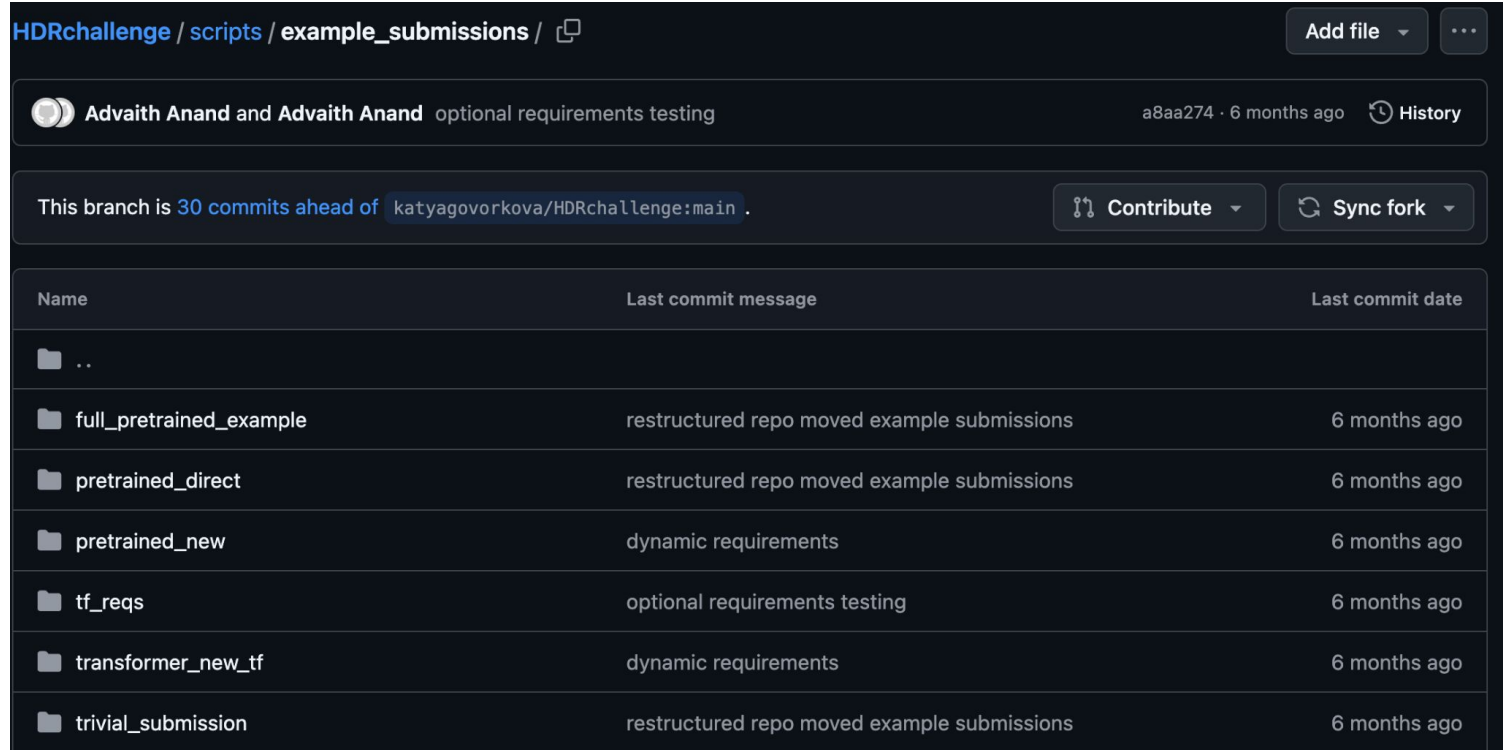
from matplotlib import pyplot as plt
from sklearn import metrics
```


# 8. Get Public Data



The screenshot shows a challenge interface with a timeline at the top and a navigation menu on the left. The timeline spans from October 2024 to January 2025. The navigation menu includes 'Challenge Overview', 'Datasets', 'Starting kit and sample submission', 'Timeline', 'Terms', and 'Files'. A red arrow points to the 'Files' menu item. The main content area displays a table with the following data:



Download	Phase	Task	Type	Size
<a href="#">solution @ 04-09-2024 19:28</a>	Development Phase	Test your AD algorithm	Solution	522 B
<a href="#">Dataset</a>	Development Phase	-	Public Data	473.26 MB








# 9. Checkout example submissions



HDRchallenge / scripts / example\_submissions / 

 Advaith Anand and Advaith Anand optional requirements testing a8aa274 · 6 months ago  History

This branch is [30 commits ahead of](#) [katyagovorkova/HDRchallenge:main](#) .  **Contribute**  **Sync fork**

Name	Last commit message	Last commit date
 ..		
 full_pretrained_example	restructured repo moved example submissions	6 months ago
 pretrained_direct	restructured repo moved example submissions	6 months ago
 pretrained_new	dynamic requirements	6 months ago
 tf_reqs	optional requirements testing	6 months ago
 transformer_new_tf	dynamic requirements	6 months ago
 trivial_submission	restructured repo moved example submissions	6 months ago

# 11. Code submission structure [\[Example\]](#)

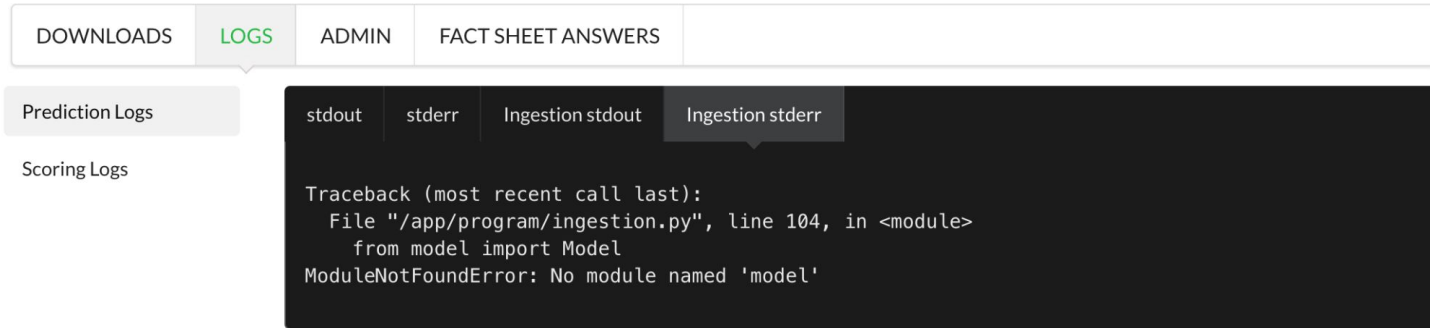
```
1 import tensorflow as tf
2 import json
3 import os
4
5 class Model:
6     def __init__(self):
7         # You could include a constructor to initialize your model here, but all calls will be made to the load meth
8         self.clf = None
9
10    def predict(self, X):
11        # This method should accept an input of any size (of the given input format) and return predictions appropri
12        preds = self.clf.predict(X)
13        print(preds)
14        return preds
15
16    def load(self):
17        # This method should load your pretrained model from wherever you have it saved
18
19        with open(os.path.join(os.path.dirname(__file__), 'config.json'), 'r') as file:
20            for line in file:
21                self.clf = tf.keras.models.model_from_json(line)
22            self.clf.load_weights(os.path.join(os.path.dirname(__file__), 'model.weights.h5'))
```

[\*] Follow the example to load your model. Avoid hard-coded path to model weight



# Common issue

[!!] Do not zip the whole folder. ONLY select the model.py and relevant weight files to make the tarball



The screenshot shows a web application interface with a navigation menu at the top containing 'DOWNLOADS', 'LOGS', 'ADMIN', and 'FACT SHEET ANSWERS'. Below the menu, there are two log categories: 'Prediction Logs' and 'Scoring Logs'. The 'Prediction Logs' category is selected, and the 'Ingestion stderr' tab is active. The log content displays a Python traceback error:

```
Traceback (most recent call last):
  File "/app/program/ingestion.py", line 104, in <module>
    from model import Model
ModuleNotFoundError: No module named 'model'
```

If you see the above error, mostly likely you zip the whole folder when making the tarball

# Submission Links

Butterfly: <https://www.codabench.org/competitions/3764/>

Gravitational Wave: <https://www.codabench.org/competitions/2626/>

Sea Level: <https://www.codabench.org/competitions/3223/>

---

Get connected in slack  
channel



Confidential

Copyright ©



# NSF HDR ML Challenge

