Final Project

DLCV Fall 2024

Video

- **Overview** [link]
- **Challenge 1** [link]
- **Challenge 2** [link]

Final Project Overview

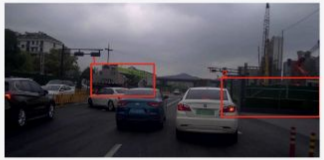
- Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving (ECCV 2024 Challenge)

Task1: General Perception

vehicles:[{
vehicles1_description: A black SUV directly ahead in the left with brake lights on, indicating it is either slowing down or stopped. ,
vehicles1_explanation: The presence and behavior of the vehicle ahead will dictate the ego car's need to ... ,
}]
vulnerable_road_users:[{...}],
traffic lights:[{...}],
traffic cones:[{...}],
barriers:[{...}],other objects:[{...}],
description and explanation:...

Task2: Regional Perception

1: {description and explanation: A construction vehicle is a heavy-duty truck equipped with specialized equipment that caninclude a flatbed, crane arm ... ,
box: [372,284,271,150],
category_name: construction_vehicle
},
2: {description and explanation: A barrier is an obstacle used to block or limit ... ,
box: [933,348,422,181],
category_name: barrier
}

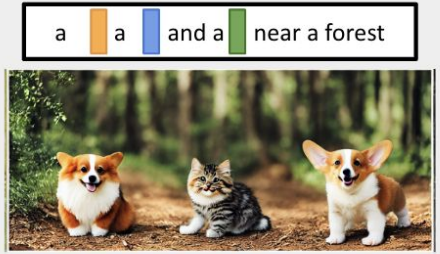


Task3: Driving Suggestions

Maintain a safe following distance from the **black SUV ahead** and be prepared to stop due to the **red traffic light** at the intersection. Stay alert for any potential unexpected maneuvers from the **white sedan on the right**. Do not ...

- Multiple concept personalization

HW2.3: Single Concept Personalization



This task: Multiple Concept Personalization

Outline

- General Rules
  - Teaming up
  - GitHub / CodaLab / Poster / Presentation
  - Grading
- Challenges
  - **Challenge 1** - Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving (ECCV 2024 Challenge)
  - **Challenge 2** - Multiple Concept Personalization

Timeline & Deadlines (GMT+8)

Teaming-up Form Completion	2024/11/30 23:59
Announcement of teaming-up	2024/12/01
Choose when to meet with TAs	2024/12/05 23:59
Project Discussion with TAs	2024/12/09~12/13
Poster Submission	2023/12/23 11:59
CodaLab Submission	2024/12/26 07:59
On-Site **Presentation**	2024/12/26 13:00-17:00
GitHub Code Commit	2024/12/26 23:59

Teaming Up and Challenge Selection

- Please fill in this **form** before **2024/11/30 23:59**
  - Each team should have **4** members
  - Team name
    - English letters (lowercase and uppercase) and numbers only; no spaces
    - **You must use the same team name for GitHub/CodaLab**
  - Team leader
    - Responsible for GitHub team creation and poster/code submission
- We will split the teams equally between the two challenges
  - Your topic choice will be determined by the order of form submissions

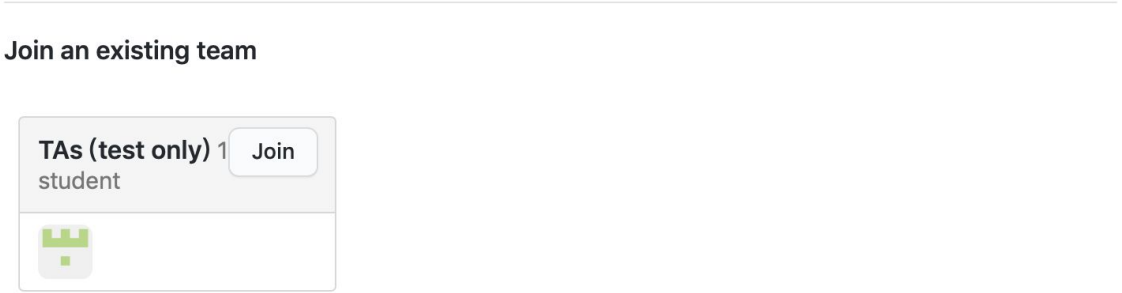
Discussion with TAs

- We will announce more details on 2024/12/01 after teaming-up done.
- Please choose when to meet with TAs before **2024/12/05 23:59**
- Each team will have **20 mins** to discuss.
- **Note that if you do not show on the meeting time, you will lose some points!**

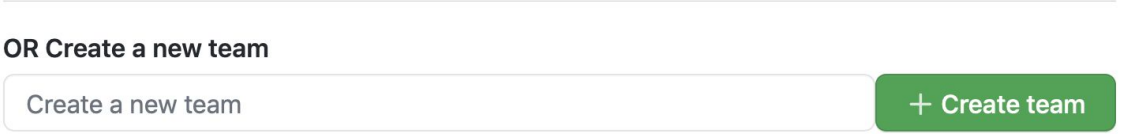
GitHub

- Join the GitHub group assignments (for each challenge) with your **team name**
  - You must use the same team name for GitHub/CodaLab
  - **The team leader creates the team first, and the team members join afterwards**

If you are not the team leader ⟶



If you are the team leader ⟶



CodaLab

- You need to participate in the CodaLab challenge with your **team name**
- Maximum Daily Submissions: 5 times (for each team)
    - CodaLab - will be reset every day
- **Submission Deadline: 2024/12/26 07:59**

Poster for On-Site Presentation

- **PDF format of size A1 (Portrait, 84.1 cm x 59.4 cm)**
- TAs will print it out for your on-site presentation only if you submit it before the deadline.
- **Submission Deadline: 2024/12/23 11:59**
  - Submitted to the root directory of the team's GitHub repository (format: **poster.pdf**)
  - You can leave some blank areas on your poster for further experiment results and fill them up right before the final presentation.
- If you do not submit your poster before the above deadline, you will need to print it out on your own.
- **Remember to put your department, grade, name, and student ID on the poster!**

On-Site Presentation

- **Schedule: 2024/12/26 13:00-17:00**
- Location: 電二142
- 休息室: 電二144

13:00 - 13:20	**Challenge #1 - Poster Readying**
13:20 - 14:40	**Challenge #1 - Presentation**
14:40 - 15:00	**Tea Break / Challenge #2 - Poster Readying**
15:00 - 15:10	**Challenge #1 - Awarding Ceremony**
15:10 - 16:30	**Challenge #2 - Presentation**
16:30 - 16:50	**Tea Break**
16:50 - 17:00	**Challenge #2 - Awarding Ceremony**

On-Site Presentation

- **Poster Readying**
  - 13:00-13:20 for Challenge-1 and 14:40-15:00 for Challenge-2
  - Prepare your posters (i.e., pasting them onto to the boards) in the given time slots
- **Presentation**
  - Proceed team-by-team according to the **Team ID** for each challenge (Excel)
  - **Time Limit - 5 mins per team**
    - Each team will be given a maximum of 4 minutes for presentation
    - An additional 1 minute will be reserved for Q&A from the lecturer and the TAs
    - As we have a tight schedule, we will control your time strictly!
  - For each team, if no members show up for the final presentation, all team members will receive 0 points for this part (0 out of 25 points)

Code Submission

- **Code Submission Deadline: 2024/12/26 23:59**
- Submit all the training/testing code to your team's Github repository
- Provide a detailed **README.md** file with example scripts for TAs to reproduce your results (including model training and inference)
- If TAs cannot reproduce your results, you will receive 0 points in the code part (unless minor errors)

Grading

- Model Performance - CodaLab
  - Baseline
  - Relative ranking
- Peer Review (For Challenge 2)
- Approach & Presentation
  - Discussion with TAs
  - Novelty and Technical Contributions
  - Completeness of Experiments
  - Poster & Oral Presentation
  - Bonus - Intra / Inter-Team Evaluation

Grading - Intra/Inter-Team Evaluation

- Intra-Team Evaluation
  - You must participate and work with your team member
  - We might adjust your final scores based on the evaluation
- Inter-Team Evaluation
  - The top 3 teams selected by (lecturer, guest, & TA) judges will receive cash prizes
  - The most-voted teams for each challenge will receive bonus points (or gifts)

Challenge 1 -
Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving (ECCV 2024 Challenge)

GitHub Classroom Link

CodaLab Competition Link

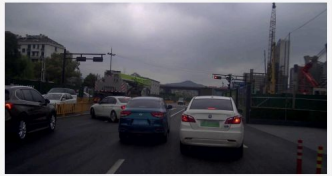
(Do not join them until we announce the final topics for your teams)

CODA-LM

- Real-world autonomous driving dataset
- **Input**: image, question
- **Output**: Description of the scene
- Question types
  - **General Perception**
  - **Region Perception**
  - **Driving Suggestions**

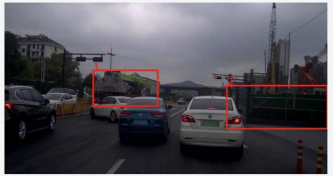
Task1: General Perception

vehicles:[{
vehicles1_description: A black SUV directly ahead in the left with brake lights on, indicating it is either slowing down or stopped. ,
vehicles1_explanation: The presence and behavior of the vehicle ahead will dictate the ego car's need to ... ,
}]
vulnerable_road_users:[{...}],
traffic lights:[{...}],
traffic cones:[{...}],
barriers:[{...}],other objects:[{...}],
description and explanation:...



Task2: Regional Perception

1: {description and explanation: A construction vehicle is a heavy-duty truck equipped with specialized equipment that caninclude a flatbed, crane arm ... ,
box: [372,284,271,150],
category_name: construction_vehicle
},
2: {description and explanation:A barrier is an obstacle used to block or limit ... ,
box: [933,348,422,181],
category_name: barrier
}



Task3: Driving Suggestions

Maintain a safe following distance from the **black SUV ahead** and be prepared to stop due to the **red traffic light** at the intersection. Stay alert for any potential unexpected maneuvers from the **white sedan on the right**. Do not ...

Question type: General Perception

- Q: ……Please **describe** each object's appearance, position, direction, and **explain** why it affects the ego car's behavior.

General Perception

One black car driving ahead in the same lane as the ego car, maintaining a straight trajectory.... **Multiple pedestrians** are visible on both sides of the streets, some closer to the roadway edge on the right side, which may indicate potential road .... **Sanitation workers** on the right side of the road.... **A no-entry sign** is visible on the right side of the road, indicating a restriction for vehicles entering that lane.... **Orange traffic cones** line the right side of the road, marking out a section likely for work ....

Model Answers

GPT-4V:
In this image, several objects .... **Pedestrians** walking along the sidewalk and potentially crossing the road.... Additionally, there is a **car** ahead of the ego car, dictating the speed.... Visible too is **a no-entry sign** for a side road .... On the right side of the road, **traffic cones and road maintenance workers** present a potential hazard ....

Gemini Pro:
In the scene, the objects that greatly influence the driving behavior of an autonomous car include **pedestrians**, **parked cars**, **traffic signs**, the **road surface**, and **weather conditions** ....

InternLM2-vl:
In the depicted scene, **the perspective visible traffic sign**s scattered throughout the scene play .... **Pedestrians**, notably present and in some instances crossing the street .... The image also shows **other vehicles**, some stationary and others in motion .... T**he condition of the road**, though seemingly well-maintained with clear markings .... **The surrounding environment**, characterized by lined trees.

LLaVA:
In the image, there is a **black car** driving down the street, and **a red stop sign** is visible in the scene. The stop sign is an object that has a significant influence on ego car driving behavior ....

Question type: Driving Suggestions

- Q: ……Please provide driving suggestions for the ego car based on the current scene.

Driving Suggestions

The ego vehicle should **maintain a safe following distance** from the SUV ahead, be prepared to react to pedestrian movements, especially from the right side where workers are present, and comply with the no-entry traffic sign by **not turning into the restricted lane**. The ego car should also **monitor the delineated pathway** defined by the traffic bollards and be prepared to take evasive actions if necessary while **maintaining a safe speed to respond** to any unexpected pedestrian or worker movements.

Model Answers

GPT-4V:

In the image, the ego car **finds itself navigating a road teeming** with critical elements requiring the driver's attention. Firstly, the area is bustling with pedestrian activity, featuring individuals walking alongside the road and others crossing directly ahead of the vehicle. This scenario necessitates a cautious approach, urging the driver to decelerate and remain on high alert, ready to **come to a complete stop** should a

InternLM2-vl:

The driver of the ego car should be cautious and **maintain a safe distance** from the car ahead and **be aware of the do not enter sign** ..... Additionally, the driver should be prepared for any sudden movements from the pedestrians or other vehicles, as they may be **crossing the road or changing lanes ....**

LLaVA:

Question type: Regional Perception

- Q: Please **describe** the object inside the red rectangles (bounding boxes) in the image and explain why it affect ego car driving.

Regional Perception

Object1: This is a **traffic cone** often used for traffic management. It's a high-visibility safety device with a cone shape, designed to direct traffic or mark off areas for safety or construction ....

Object2:This is a **traffic sign**, specifically **a 'No Entry' sign**, which is a regulatory sign that indicates that a roadway is closed to vehicular traffic entering in the direction the sign is facing. The sign ....

Model Answers

GPT-4V:

Object1: This is a **traffic cone** often used to indicate areas of roadwork, potential hazards, or to guide traffic away from a particular section of the road. The presence ....

Object2:This is a **"No Entry" traffic sign**, which is a type of regulatory sign used to indicate that vehicles are not allowed to enter the road from the direction faced ....

Gemini Pro:

Object1: This describes the **traffic cones**. These typically bright orange, cone-shaped markers are used

InternLM2-vl:

Object1: The object inside the red rectangle is **a sign with a red circle** and a diagonal line through it ....

Object2:The object inside the red rectangle in the image is a **sign with a red circle** and a diagonal line through it. This sign is commonly known as a "Do Not Enter" sign ....

LLaVA:

Object1: The red rectangle in the image contains a **stop sign**. The stop sign is placed on the side of the road to regulate traffic and ensure ....

CODA-LM

- **Challenges**
  - Diverse Scenes / objects
  - VLM (i.e. LLaVA) cannot handle real-world traffic scene well
  - Small objects perception
- **Application**
  - Provide instant suggestion for drivers/self-driving cars

Goal

- Build a system that can solve the QA problems in CODA-LM
- Rules:
  - You can only use **llava-1.5-7b-hf** in you system to output text
    You should not use other language models (open source / api)
  - Fine-tune LoRA, projectors in LLaVA are allowed
  - Additional modules (pretrained / train by yourself) are allowed
  - Public available external datasets are allowed
- We will use LLM evaluation and BLEU score of prediction on “private test” data for your performance score

Data

- **Source: huggingface** (English version, LLaVA format)
- **Public available data**
  - Training Data: 6884 samples
  - Validation Data: 2084 samples
- **Private test data**
  - **300** samples
    - **Curated by DLCV TAs**, not the official test set
    - You are not allowed to utilize private test data in any form, e.g. data augmentation, fine-tuning.

Huggingface Datasets

- **Usage:**

```
datasets.load_dataset("ntudlcv/dlcv_2024_final1", split = split)
```

- **This will return an IterableDataset**
- **split can be: “train”, “val”, “test”**
- **You can use the “streaming” argument to avoid downloading whole data**

Data format

- Data format (for each data):
  - dictionary with keys: id, image, conversations

subset sample index

{ “id”: Test_general_0, “image”: PIL image, “conversations”: input text and output description}

question type

  - “conversations” share the same format of LLaVA’s instruction tuning dataset format, you can see LLaVA-instruct-150K for further details

Evaluation Metric

- **We will use the following metrics to evaluate prediction on “prevate_test”**
- **LLM judges**
  - Use LLM to evaluate the correctness of model’s response compared with the ground truth annotation, scoring from **1-10**.
  - LLM: Gemini API
    - The free quota is enough to run inference on the test set **once per day**.
    - For reference, we provide evaluation code that LLM base is LLaMA, which you can use to test before evaluating with Gemini. (~16G during inference)
  - It takes time to compute LLM judges
- **NLP metric**
  - **BLEU-4 score**
- The Toal score is calculated by: **0.8 * LLM judges + 0.2 * Bleu-3**
- **Baseline: Total score = 4**

Submission

- **Submission deadline: 2024/12/26 07:59**
- **.zip** file include following two files
  - API_key.txt
    - Save the API key needed for Gemini Evaluation
    - Please ensure your API key has **sufficient quota** for the evaluation
    - **Invalid API keys** will result in **no score** for the LLM judgment section
  - submission.json (your prediction on private test data)
    - key: <id>
    - value: <output>

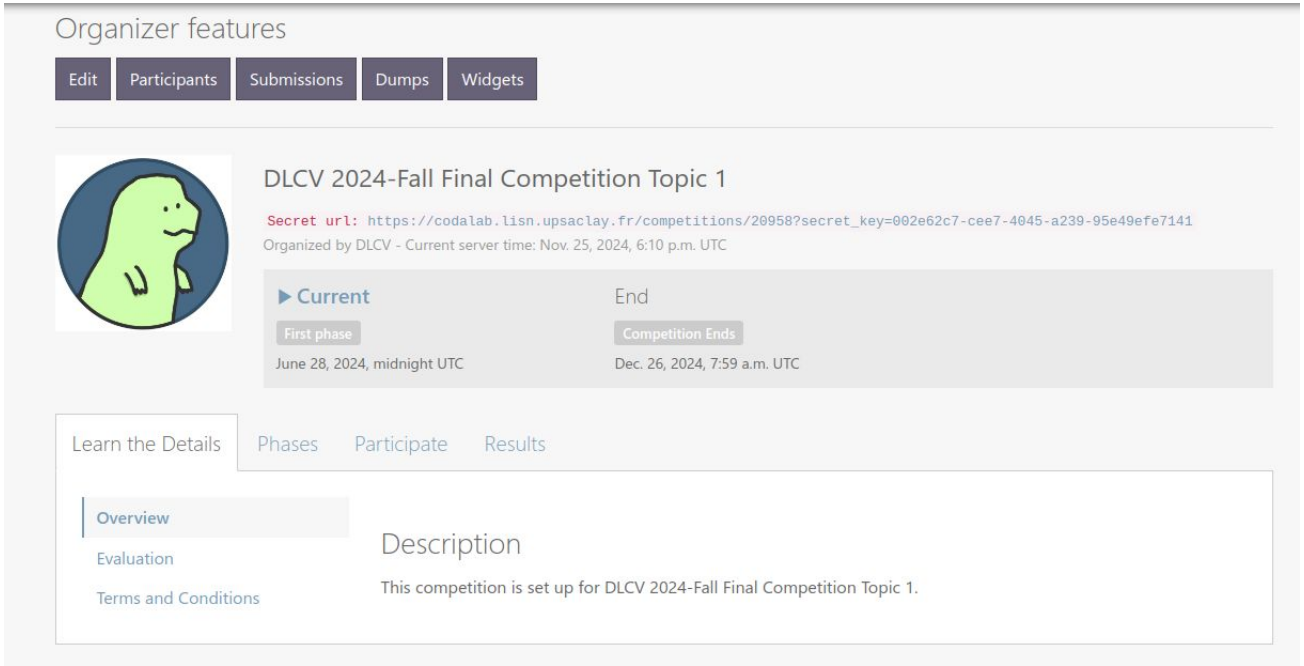
```
{
    "Test_general_0": "In the image, there is a green bus driving down the street.
    "Test_suggestion_0": "Based on the image, the ego car is driving on a street wi
}
```

Evaluation Platform - Local Eval

- We provide a script in GitHub Repo to utilize **Llama 3.1** for local evaluation.
- please refer to **README.md**
- The results from LLaMA-3 may differ from Gemini's evaluation. **Please use LLaMA-3's results only as a reference.**

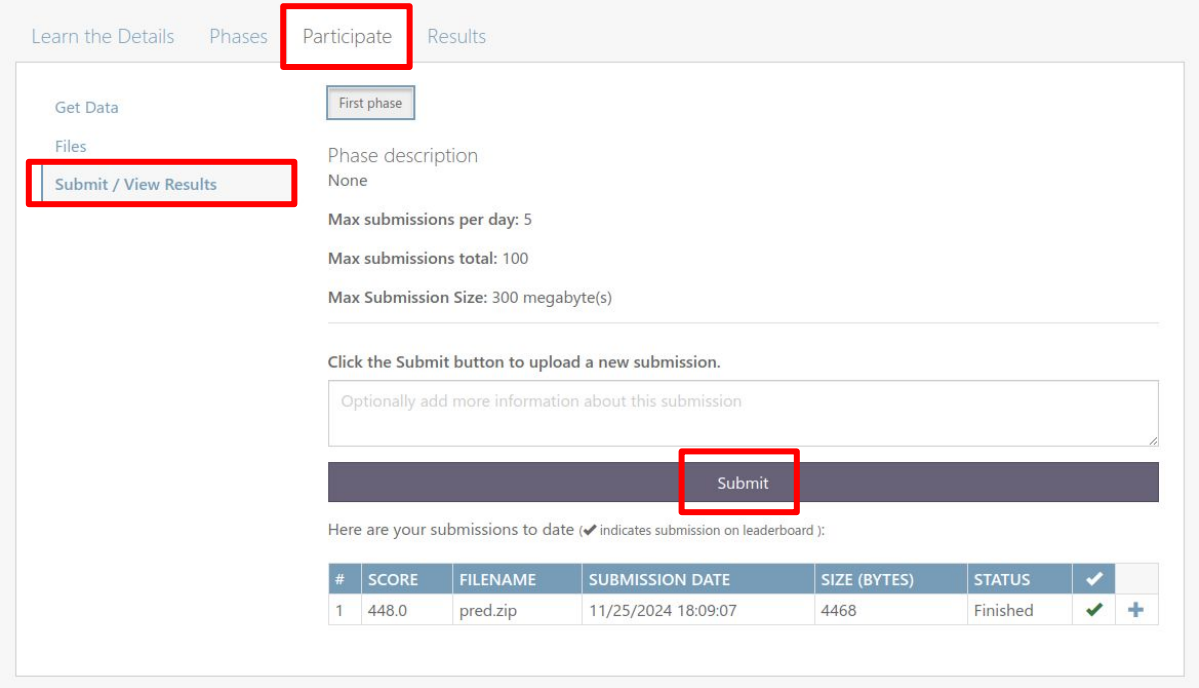
Evaluation Platform - Private Eval

- **Codalab: Link**



Evaluation Platform - Private Eval

- The submission process might take several minutes.



Evaluation Platform - Private Eval

- The submitted file must be named in pred.zip
- The submitted file must contain the following:
  - api_key.txt
  - submission.json
- Note that you should compress the two files directly
  - Do not put them in a directory
- We provide an example submission file in GitHub repo

Summary - Links

- Evaluation platform: Codalab (Will be released right after team up)
- GitHub repo: classroom (Will be released right after team up)
- Huggingface dataset: huggingface
- CODA LM original website: project page

CODA-LM - Grading

- **Final 34%** (Bonus up to **3%**)
  - **Model Performance - CodaLab 9%**
    - Baseline(**Total score = 4**): **4%**
    - Relative ranking in class **5%**
  - **Approach & Presentation 25% + 3%**
    - Novelty and technical contributions **10%**
    - Completeness of experiments **10%**

      (e.g., ablation study, visualization, etc.)
    - Poster & Oral Presentation **5%**
    - Bonus (intra / inter-team evaluation) up to **3%**

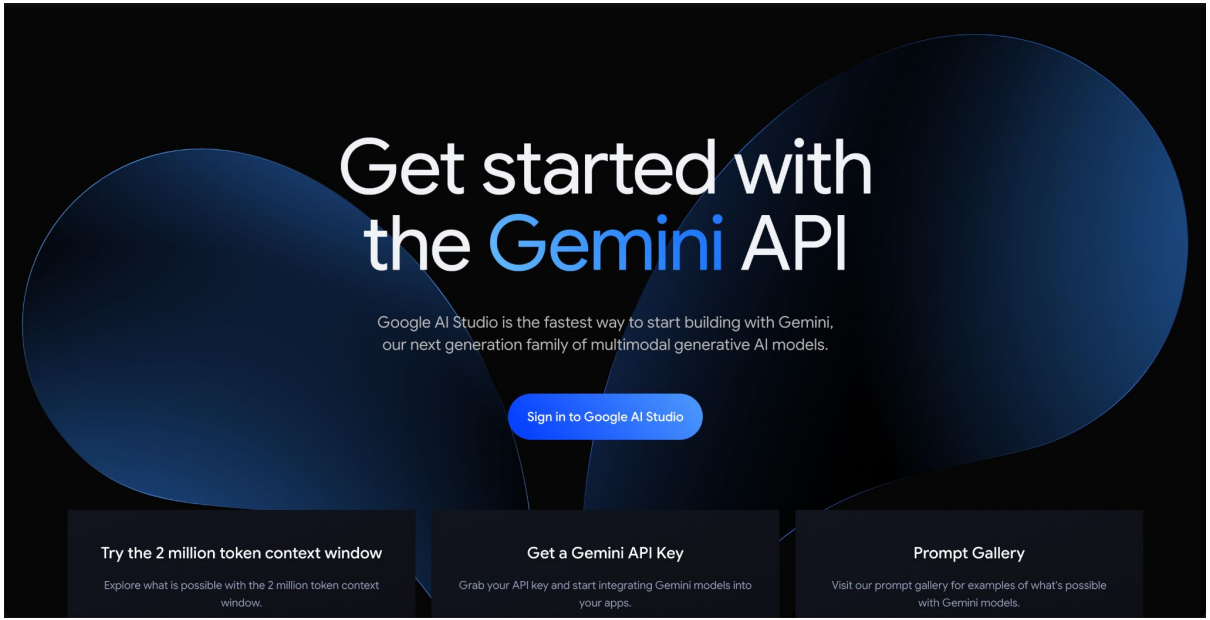
Points	Team Ranking
5	top 0% - 20%
4	top 20% - 40%
3	top 40% - 60%
2	top 60% - 80%
1	top 80% - 100%

Warning and Suggestion

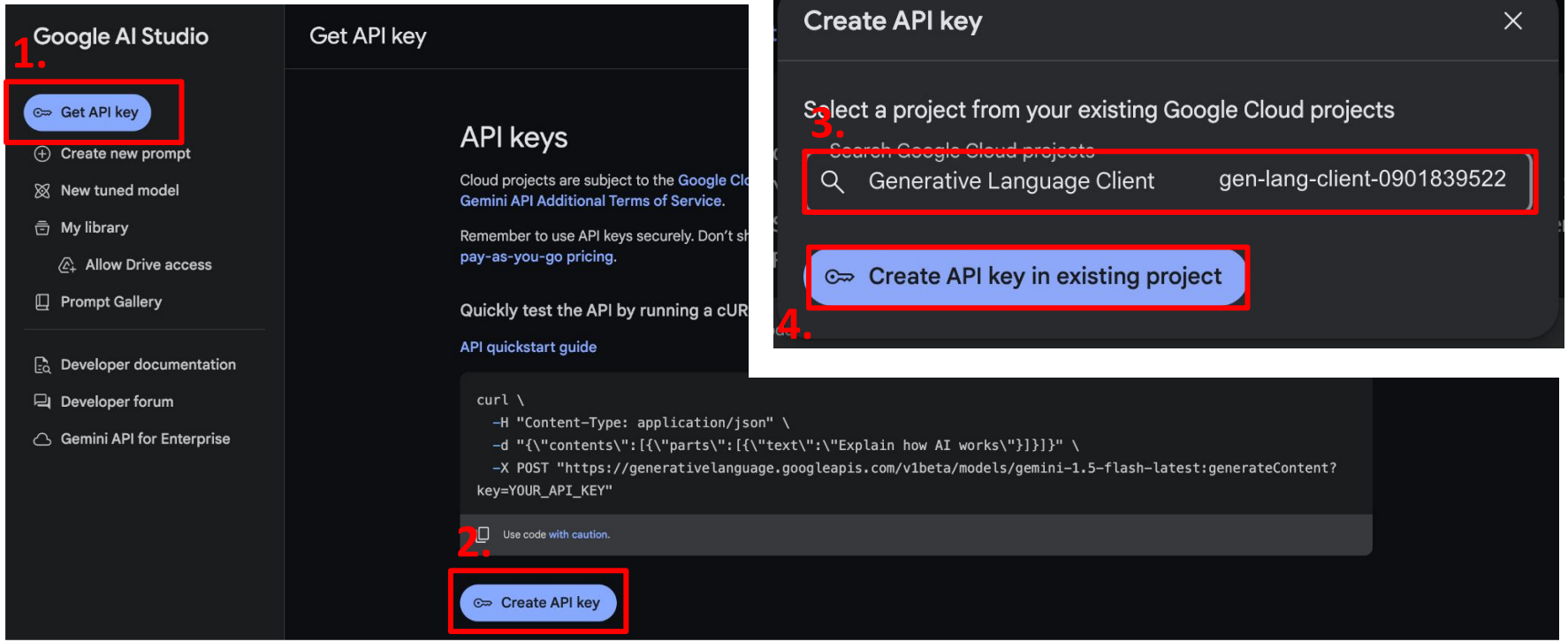
- **Warning**
  - You will not get points for novelty if you only naively finetune.
  - You cannot use APIs like ChatGPT, Gemini, or Claude in this final project, including for synthesizing data or generating predictions, etc.
  - You are not allowed fintune, conduct data augmentation… on test data
- Suggestion:
  - Retrieval augmented generation
  - Other visual clues (e.g., segmentation, depth map)
  - Different encoders
- Your results must be reproducible with your submitted code and models
- Any violation mentioned in this slide will result in **0 score** for your final project

Supplementary – Gemini API (1/3)

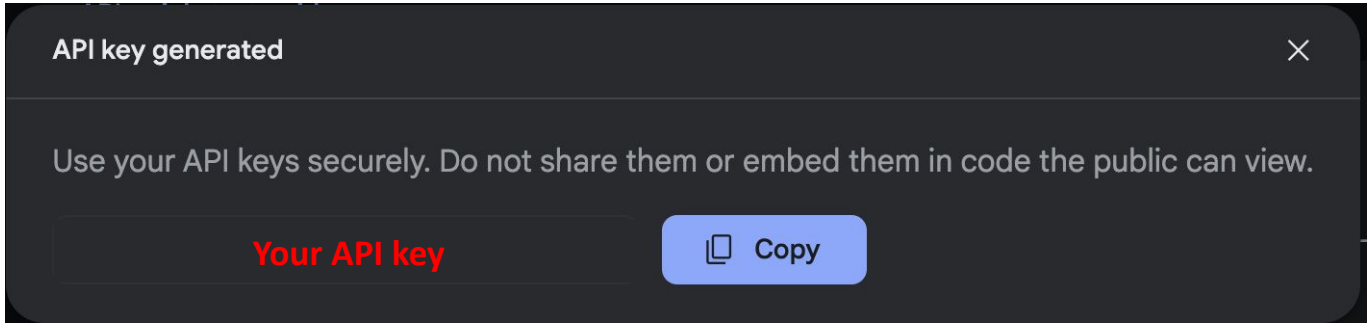
https://ai.google.dev/aistudio



Supplementary – Gemini API (2/3)



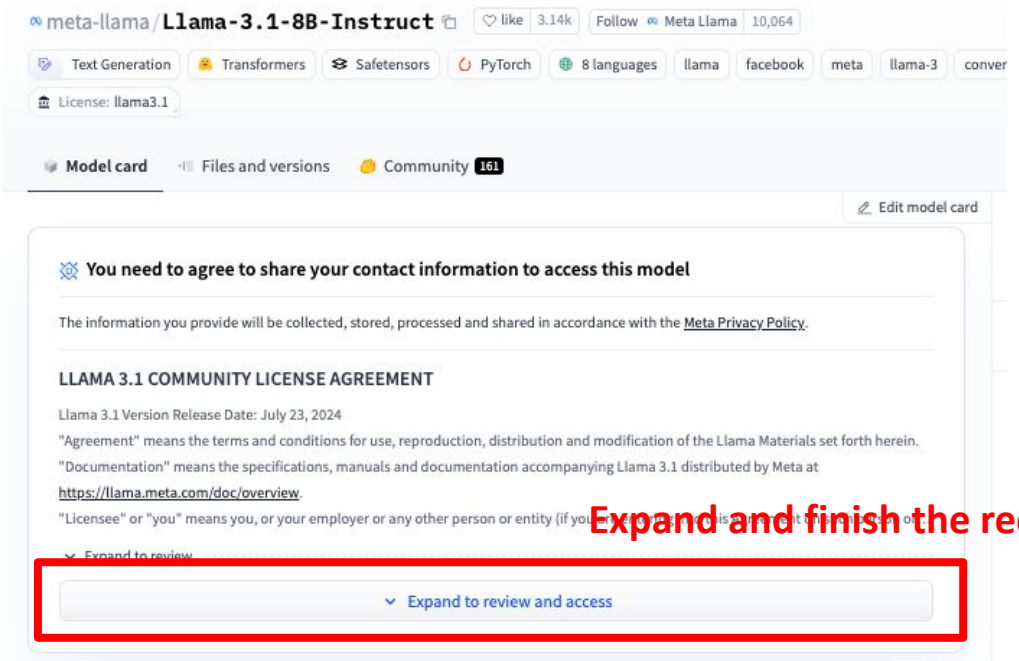
Supplementary – Gemini API (3/3)



- Please be mindful of your quota usage (free version).
  - request limit per day: 1500
- You can create more google accounts or use llama_eval.py to evaluate your prediction first

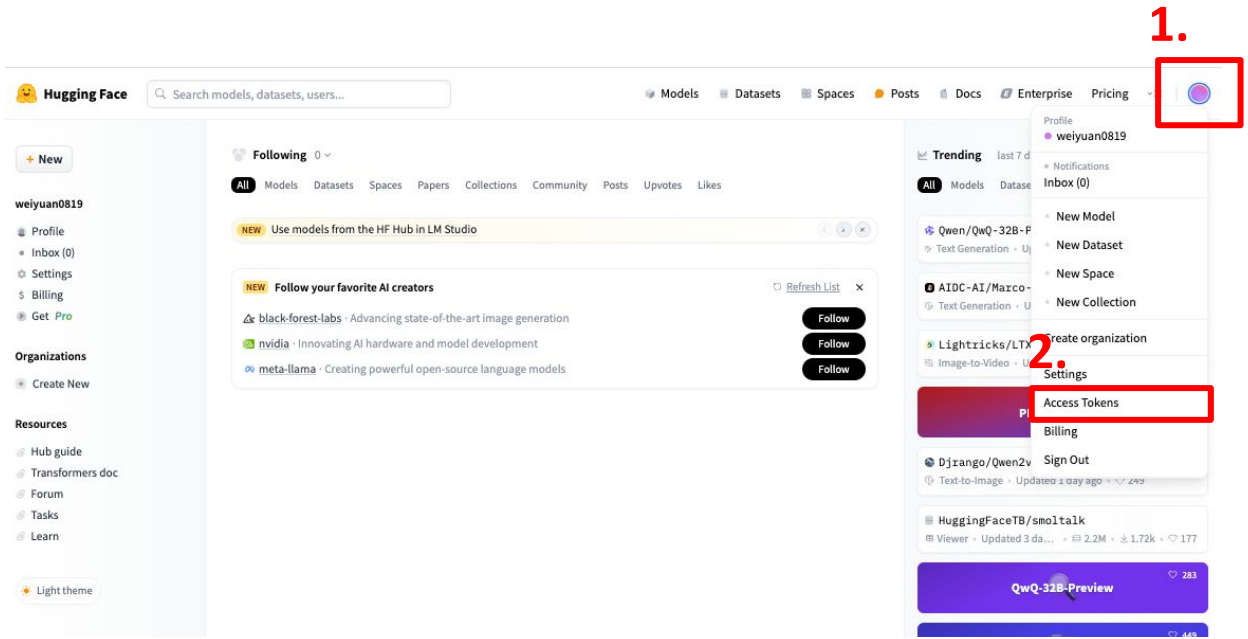
Supplementary – LLaMA-3.1 (1/4)

You need to finish the license to access LLaMA3

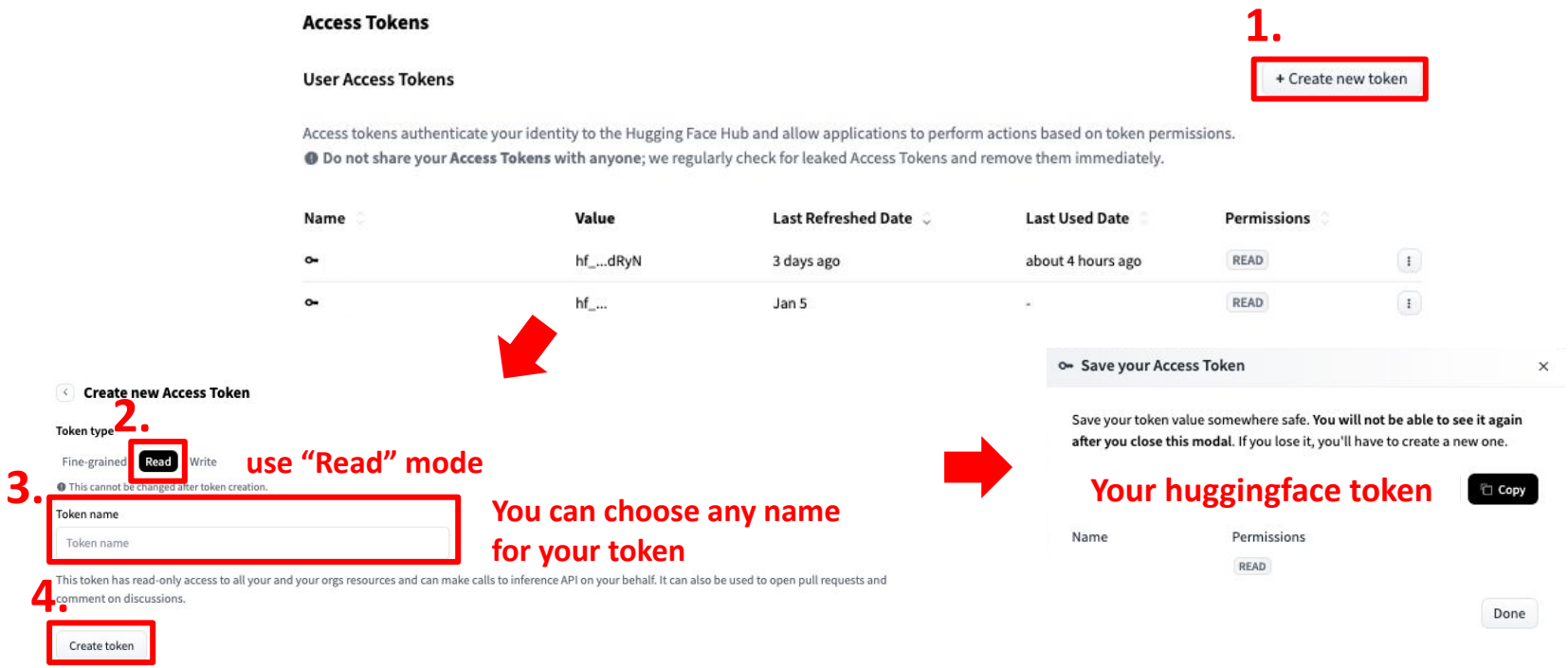


Expand and finish the requirement

Supplementary – LLaMA-3.1 (2/4)



Supplementary – LLaMA-3.1 (3/4)



Supplementary – LLaMA-3.1 (4/4)

1. run the following command in your terminal

```
>>> huggingface-cli login
```

2. paste your generated token here

3. you can log in huggingface and get access to LLaMA3.1. For more details, please refer to [huggingface]

Challenge 2 -
Multiple Concept Personalization

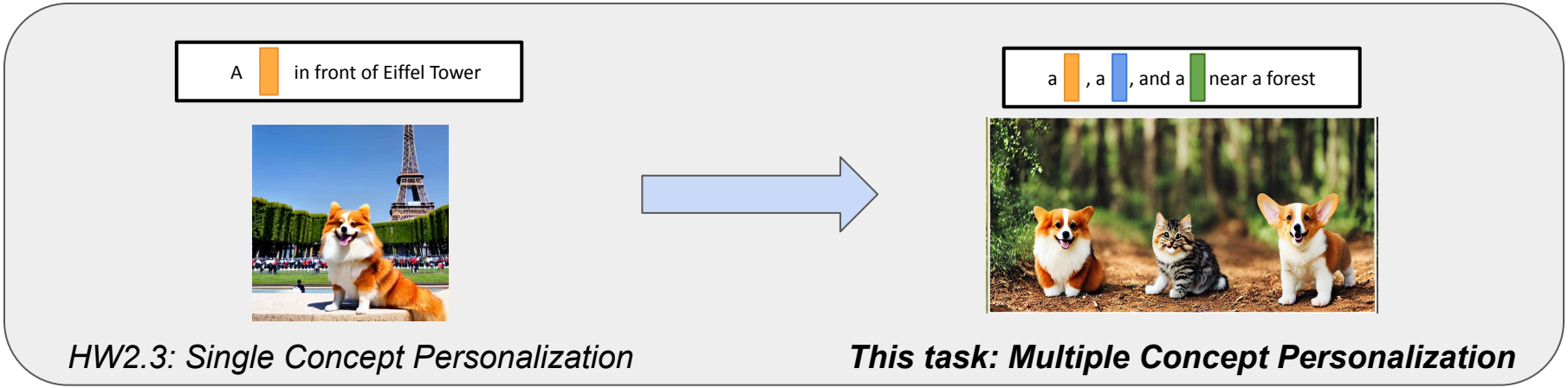
GitHub Classroom Link

CodaLab Competition Link

(Do not join them until we announce the final topics for your teams)

Multiple Concept Personalization - Task (1/2)

- Extending from *homework 2.3 on Single Concept Personalization*, this task challenges you to personalize text-to-image generation for **multiple concepts**.
  - **Input**: **Prompts** and **concept images** for multiple distinct concepts
  - **Output**: **Generated images** depicting the input prompts with the specified multiple concepts



Multiple Concept Personalization - Task (2/2)

- The **prompts** and **concepts (including object and style)** to address:
  1. Dual-concept prompts
     a. a cat **in the right** and a dog **in the left**
     b. a flower **in** a vase
  2. Tri-concept prompts
     a. a dog1, a dog2, and a cat near a forest
     b. a cat **wearing** a glasses in watercolor style

(2.a) Example concepts (objects)

(2.a) Example output

(2.b) Example concepts (objects and style)

(2.b) Example output

Multiple Concept Personalization - Dataset

- **Data Format:**
  - Data/
    - prompts.json
    - concept_images/
      - cat2
        - 0.jpg
        - 1.jpg
        - …
      - dog
        - …
      - …


```
{
    "0":{
        "src_image": ["cat2", "dog6"],
        "token_name": ["<cat2>", "<dog6>"],
        "prompt": "A <cat2> on the right and a <dog6> on the left.",
        "prompt_4_clip_eval": "A cat on the right and a dog on the left."
    },
    "1":{
        "src_image": ["flower_1", "vase"],
        "token_name": ["<flower_1>", "<vase>"],
        "prompt": "A <flower_1> in a <vase>.",
        "prompt_4_clip_eval": "A flower in a vase."
    },
    "2":{
        "src_image": ["dog", "pet_cat1", "dog6"],
        "token_name": ["<dog>", "<pet_cat1>", "<dog6>"],
        "prompt": "A <dog>, a <pet_cat1> and a <dog6> near a forest.",
        "prompt_4_clip_eval": "A dog, a pet cat and a dog near a forest."
    },
    "3":{
        "src_image": ["cat2", "wearable_glasses", "watercolor"],
        "token_name": ["<cat2>", "<wearable_glasses>", "<watercolor>"],
        "prompt": "A <cat2> wearing <wearable_glasses> in a <watercolor> style.",
        "prompt_4_clip_eval": "A cat wearing wearable glasses in a watercolor style."
    }
}
```

Multiple Concept Personalization - Evaluation (1/6)

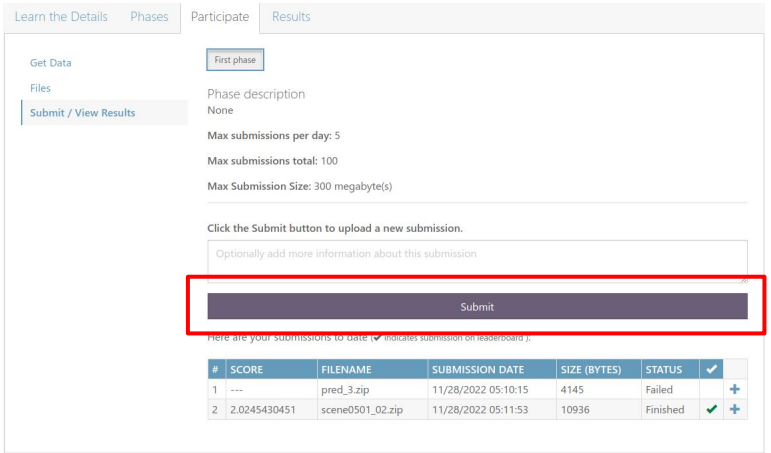
- There are two evaluation methods for this task:
  a. CodaLab quantitative evaluation (see ***Evaluation p.2 ~ Evaluation p.5***)
  b. Peer review (see ***Evaluation p.6***)

Multiple Concept Personalization - Evaluation (2/6)

- CodaLab Quantitative Evaluation
  - We will use the **CLIP-I / CLIP-T** to quantify your model performance
    - image alignment is measured as the average visual similarity between the generated image and each concept image
  - You should submit your **zip** file to the CodaLab competition
  - Please note that the **zip** file should contain **4 folders**, each contains **10 images**
    - folder names **MUST** be **0, 1, 2, 3**
    - image name & extension are not restricted (either **.jpg** or **.png** is ok)
- **Submission format**
  - sample_submission.zip
    - 0/…(10 images)
    - 1/…(10 images)
    - 2/…(10 images)
    - 3/…(10 images)

Multiple Concept Personalization - Evaluation (4/6)

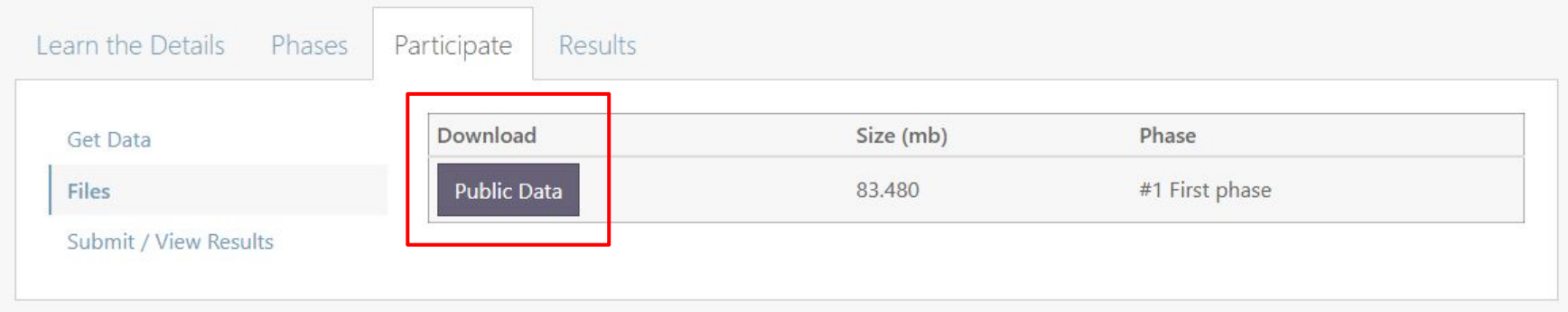
- Submit you **zip** file in the Submit/View Results page.
- You can check your score and ranking in the Results page.
- After you upload the **zip** file to the "Submit/View Results" page. You can wait for the evaluation.
  - The submission status should go as "Submitting" -> "Submitted" -> "Running" -> "Finished".
  - Because the submission file is quite large, please wait until it show the status “Submitting”.
  - Click "Refresh status" too many times may cause the submission stuck at "Submitted"!
  - If it has been stuck at the "Submitted" status for a long time (more than 30 minutes), you can e-mail TAs to re-run your submission.



Multiple Concept Personalization - Evaluation (3/6)

- Create an account and participate the competition (with your team name)
  - CodaLab Competition
  - All members in a team should use the same account.

    (i.e. one team, one account)
- Download the dataset from the link or the **Files** page.
  - Download dataset



Multiple Concept Personalization - Evaluation (5/6)

- **CodaLab submission deadline: 2024/12/26 07:59 (UTC+8)**
- Max submissions: 5 times per day / 100 in total
- Baseline
  - **CLIP-I : 61**
  - **CLIP-T : 28**

Multiple Concept Personalization - Evaluation (6/6)

- Peer review
  - Submission:
    - Upload 10 images per prompt to **Peer Review**
      - each image should not larger than 512*512
    - **Peer review images submission deadline: 2024/12/22 23:59 (UTC+8)**
  - Execution:
    - TA will release a google form for peer review **2024/12/24 (UTC+8)**
    - **Please make sure to finish the peer review before final presentation 2024/12/26 11:59 (UTC+8)**
      - Or you will get 0 point for your final project

Multiple Concept Personalization - Grading

- **Final 34%** (Bonus up to **3%**)
  - **Model Performance - CodaLab 4%**
    - Baseline **1%**
    - Relative ranking in class **3%**
  - **Model Performance - Peer Review 5%**
    - Relative ranking in class **5%**
  - **Approach & Presentation 25% + 3%**
    - Novelty and technical contributions **10%**
    - Completeness of experiments **10%**
      (e.g., ablation study, visualization, etc.)
    - Poster & Oral Presentation **5%**
    - Bonus (intra / inter-team evaluation) up to **3%**

Points	Team Ranking
3	top 0% - 20%
2.5	top 20% - 40%
2	top 40% - 60%
1.5	top 60% - 80%
1	top 80% - 100%

Points	Team Ranking
5	top 0% - 20%
4	top 20% - 40%
3	top 40% - 60%
2	top 60% - 80%
1	top 80% - 100%

Multiple Concept Personalization - Tips

- To address relative position or attribute binding issue
  - **attention calibration loss** while training stage
  - **region control** while inference stage
  - …

Multiple Concept Personalization - Rules

- You are **NOT** allow to apply any image editing work on the challenge
  - please check with TAs if you would like to apply certain techniques and you are not sure if you are allowed to do so
- Your results need to be reproducible with your submitted code and models
- Please use **python3** instead of python for your scripts
- Any violation would result in 0 score for your final project