Mô tả cuộc thi

To participate in the EVENTA 2025 Grand Challenge, please first register by submitting **the form**.

Overview

This track aims to generate captions that provide richer, more comprehensive information about an image. These captions go beyond simple visual descriptions by offering deeper insights, including the names and attributes of objects, the timing, context, outcomes of events, and other crucial details—information that cannot be gleaned from merely observing the image. Given an image, participants are required to search relevant articles in a provided external database and extract necessary information to generate an enriched image caption. This retrieval augmentation generation track facilitates the creation of more coherent and detailed narratives, capturing not only the visible elements but also the underlying context and significance of the scene, ultimately offering a more complete understanding of what the image represents.
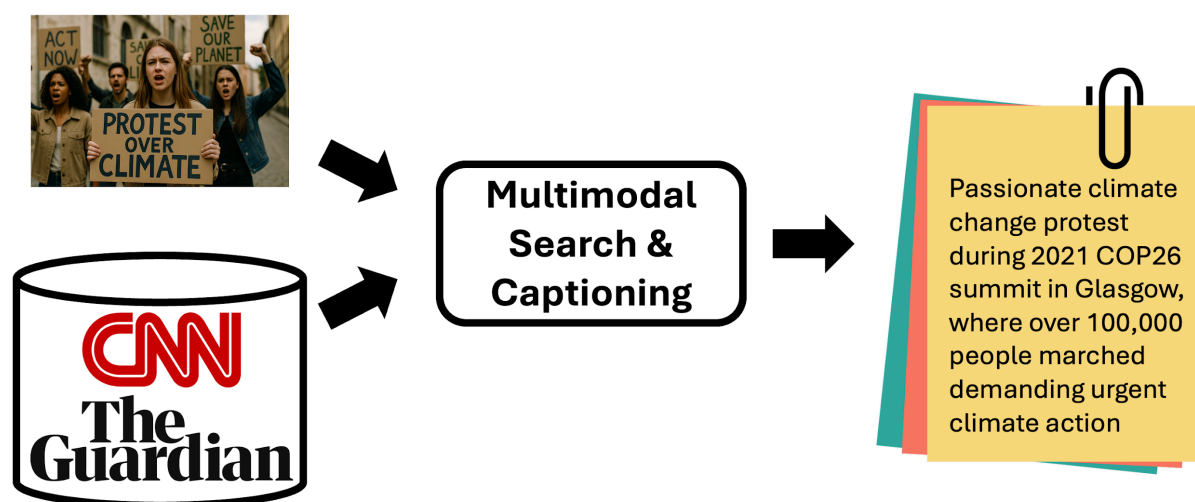
Submission Instructions

Participants must submit a CSV file named using the following format: `TeamName_EVENTA2025_Track1.csv`. This file must be compressed into a ZIP archive named `submission.zip` before uploading to CodaBench platform.

The CSV file should include predictions for all images in the query set. It must contain 12 columns, separated by commas (,), with the following structure:

- Column 1: Query image ID
- Columns 2–11: Top-10 retrieved article IDs, listed in descending order of relevance (from top-1 to top-10). If an article cannot be retrieved, use `#` as a placeholder.

- Column 12: The generated caption corresponding to the query image, enclosed in double quotation marks ("").

CSV Row Format Template:

```
query_id,article_id_1,article_id_2,...,article_id_10,generated_cap
tion
```

```
<query_id>,<article_id_1>,<article_id_2>,...,<article_id_10>,"<gen
erated_caption>"
```

```
<query_id>,<article_id_1>,#,...,#,"<generated_caption>"
```

We provide a submission example:

```
query_id,article_id_1,article_id_2,...,article_id_10,generated_cap
tion
```

```
12312,56712,56723,56734,56745,56756,56767,56778,56789,56790,56701,
"A group of children playing soccer on a sunny afternoon."
```

```
12334,56712,#,#,#,#,#,#,#,#,#,"A man riding a bicycle through a
busy city street."
```

```
12345,56712,56723,56734,56745,56756,#,#,#,#,#,"A cat sitting on a
windowsill looking outside at the rain."
```

Submission Requirements:

- Each query_id in your submission must be unique and must come from the official provided list.
- All image_id values must correspond to valid entries in our image database.
- There is no requirement to sort the rows by query_id—this will be automatically handled during evaluation.
- Each generated_caption must not exceed 8,000 characters in length.
- Your raw submission.csv file must not contain any blank lines.
- Submit the submission.csv file as a single ZIP file named `submission.zip` — do not include any containing folder.
- The submission.csv must include all headers and columns, in the exact order as shown in the example file.

Submission Limits:

- For public test submissions, each team is allowed a maximum of 10 submissions per day and a total of 150 submissions.
- For private test submissions, each team is allowed a maximum of 10 submissions per day and a total of 150 submissions.

Participants also require to submit a detailed paper through **CMT platform** to validate their solutions.

Evaluation

To fairly and comprehensively evaluate submissions in our challenge, which includes both **retrieval** and **captioning** tasks, we define an **Overall Score** that balances key evaluation criteria across both domains.

The **Overall Score** integrates the following six metrics:

Metric	Description	Domain
AP	Average Precision – measures retrieval precision across thresholds	Retrieval
Recall@1	Whether the correct item is ranked at the top	Retrieval
Recall@10	Whether the correct item is within the top 10 results	Retrieval
CLIPScore	Semantic alignment between image and caption	Captioning
CIDEr	Agreement between generated and reference captions	Captioning

These metrics are combined using a **weighted harmonic mean**, which emphasizes balanced performance and penalizes poor performance more heavily than a simple average. The formula is:

$$\text{Overall Score} = r \cdot \left(\sum w_i \sum (w_i m_i + \varepsilon) \right)$$

Where:

m_i is the value of the i-th metric

w_i is the weight for the i-th metric

ε is a small constant (e.g., 10^{-8}) to prevent division by zero

r = (#valid queries) / (#total ground-truth queries)

Default Weights: **[1, 2, 2, 3, 2]** for [AP, Recall@1, Recall@10, CLIPScore, CIDEr]

This **Overall Score** will be used as the **official ranking metric** for evaluating all participant submissions on the leaderboard.

Leaderboard

TBA

Tui có dataset như sau:

Cấu trúc Dataset EVENTA

Tổng quan

EVENTA-DATASET/

├── Database/

│   ├── database.json

│   └── image/   chứa id.jpg

├── Public_Set/

│   ├── image/   chứa id.jpg

│   └── query.csv

└── Train_Set/

    ├── image/   chứa id.jpg

    └── train.csv

Chi tiết

Database

- **database.json**: Chứa thông tin về các bài viết (articles) với cấu trúc:
  - id
  - title
  - content
  - date
  - image_id
- **image/**: Thư mục chứa các hình ảnh được sử dụng trong các bài viết

Public_Set

- **image/**: Thư mục chứa các hình ảnh cần dự đoán caption
- **query.csv**: File chứa thông tin về image_index

Train_Set

- **image/**: Thư mục chứa các hình ảnh huấn luyện
- **train.csv**: File chứa 4 cột:
  - image_index
  - caption
  - retrieved_image_id
  - retrieved_article_id