

🔑 API Credentials

Enter or load your OpenAI API Key for evaluation:

Enter OpenAI API Key: 

🔍

API key set successfully!

Load API Key from Environment

📊 Baseline Metrics Configuration

Set baseline values to compare your evaluation results against these reference points.

🖼️ → 🇺🇸 CapsBench Accuracy Baseline

CapsBench Accuracy (%)

0.00 - +




📊 Token Metrics

Average Token Length per Caption

0.00 - +

Vocabulary Size

0.00 - +

 → 

# CapsBench Benchmark

A Benchmark for evaluating the accuracy of synthetic captions against the CapsBench dataset.

📘 Understanding the Evaluation

The **CapsBench Evaluation Framework** measures how accurately generated captions answer a set of ground-truth questions derived from the CapsBench dataset.

**Scoring Methodology**

The score is calculated based on the following tasks to be accomplished for `gpt-4o-2024-08-06` :

1.

For each question, respond "yes", "no", or "n/a", based only on the caption's content.

2.


Compare these answers to the reference ground truth answers.

3.

Compute the percentage of correct responses:

$$\text{Score} = \left( \frac{\text{Number of Correct Answers}}{\text{Total Number of Questions}} \right) \times 100$$

[View an Example DataFrame Output in .md format.](#)




## Upload the CapsBench Dataset


You can download `.parquet` Or `.feather` files from [HuggingFace/playgroundai/CapsBench](#).

**Recommendation:** Perform the evaluation at least three times and average the results for stable findings.

Choose a Parquet or Feather file

 Drag and drop file here  
Limit 500MB per file • PARQUET, FEATHER

Browse files

 capsbench\_dataset.parquet 320.8MB


✕

✅ File 'capsbench\_dataset.parquet' uploaded successfully.

📊 Dataset Preview

Below is an overview of the columns in your dataset.

Column	Type	Sample Values
questions	List(Struct({'answer': String, 'category': String, 'question': String}))	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answer': 'yes', 'category': 'text', 'question': "Doe
sonnet	String	This image is a striking black and white movie poster for the film 'Nightmare Alley'. The poster features a dramatic close-up po
pg-captioner	String	A high-contrast, black-and-white movie poster for 'Nightmare Alley' featuring a close-up of Bradley Cooper's face dominating i
gpt-4o	String	This is a black-and-white movie poster featuring a man in formal attire with a prominent double exposure effect. The main sut
image	Struct({'bytes': Binary, 'path': Null})	(Struct({'bytes': Binary, 'path': Null})) Image bytes not displayed



## Processing Parameters

Advanced Configuration

Image Bytes Column

🔍

image

Caption Column

🔍

pg-captioner

Output File Path

🔍

results\_gpt-4o-2024-08-06\_pg-captioner\_iter\_1\_.parquet

Questions Column

🔍

questions

Start Index

🔍

0 - +

End Index (not inclusive)


🔍

200 - +

Step Size




🔍

1 - +

 Process File

Processing: 100%

✅ Processing complete! Results saved to 'results\_gpt-4o-2024-08-06\_pg-captioner\_iter\_1\_.parquet'.


 → 

## CapsBench Accuracy

CapsBench Accuracy (%)

76.59


↑ 76.59%



## Snapshot of Evaluation Results

Below are the first 10 evaluated entries:

candidate_caption	captioned_by	ground_truth	openai_answers	score	token_count
A surreal 3D digital illustration features a woman's face in an avant-garde composition	pg-captioner	[{'answer': 'yes', 'category': 'color', 'question': 'Is the color of the butterflies primarily	1. Is the color of the butterflies primarily blue? - yes 2. Is the face mask on the right	66.67	357
A vintage-style coffee label features an ornate frame with goldenrod (#d4a964) text on	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the text include the word 'COFFEE'?	90.91	390
A surreal digital illustration features a central female figure with long, straight teal hair	pg-captioner	[{'answer': 'yes', 'category': 'color', 'question': "Is the hair's color light green?"}] {'answ	1. Is the hair's color light green? - no 2. Is the character located in the center of the	72.73	457
Three men are seated at a black table covered with a dark blue cloth, engaged in what	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the image contain the text 'AKSHAY'	68.18	370
A close-up photograph captures an intimate moment between two individuals seated	pg-captioner	[{'answer': 'Yes', 'category': 'color', 'question': 'Is the background color off-white?'} {'a	1. Is the background color off-white? - yes 2. Is the person with the white bow tie o	72.73	425
A high-contrast, black-and-white movie poster for 'Nightmare Alley' featuring a close-	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the image contain the text 'NIGHTMA	86.67	334
A high-angle aerial photograph captures an intense nighttime scene of a large bonfire	pg-captioner	[{'answer': 'yes', 'category': 'color', 'question': 'Is the fire orange?'} {'answer': 'yes', 'ca	1. Is the fire orange? - yes 2. Is the fire in the center of the image? - n/a 3. Are the	84.62	386
A promotional graphic featuring a vibrant contrast between a bright red background and	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the image contain the text 'MOTIVATI	72.22	399
A cozy kitchen scene with an emphasis on vintage and rustic elements, featuring a white	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the image contain the text 'LINEN'?	75	280
A movie poster for 'The Secret Life of Pets' features a close-up view of a brown and white	pg-captioner	[{'answer': 'yes', 'category': 'text', 'question': 'Is there any text in the image?'} {'answe	1. Is there any text in the image? - yes 2. Does the image contain the text 'JULY 8'?	82.35	421



## Token Metrics

Average Token Length

357.04

↑ 357.04

Vocabulary Size

6456.00

↑ 6456.00

Total Number of entries successfully processed for CapsBench Accuracy: 200

✅ Evaluation completed successfully.

Developed by [alexferdigi](#) | Enhanced by the community