

CAP : Evaluation of Persuasive and Creative Image Generation

Supplementary Material

In this work, we introduced three new evaluation metrics to assess: 1. alignment of images with implicit messages, 2. creativity, and 3. persuasiveness. In the supplementary material, we provide:

1. Implementation details.
2. A detailed table for text-image alignment, including metrics from prior works, such as ImageReward, VQAScore and CLIPScore (Table 1).
3. Results from baseline metrics for evaluating creativity and persuasiveness (Table 2).
4. Two examples showing the benefit of combining the persuasiveness components for better alignment with human evaluations.
5. A comparison of PA components with human evaluations for persuasiveness (Table 3) on real advertisements (I_{real}).
6. Example showing impact of different α values.
7. Examples of failures and discussion of limitations.
8. The prompts used for MLLMs and LLMs.
9. Human annotation interface.

1. Implementation Details

We fine-tuned LLAMA3-Instruct-8B (LLAMA3) and Qwen-2.5-7B-Instruct (QwenLM) LoRA [8] parameters, for generating accurate action-reason statements for each image (AR_{gen}), using the CPO trainer [13], with a batch size of 4, learning rate of $5e-5$, and 3,000 training steps. We sampled 250 images, utilizing each of the three to five AR_m messages from the PittAd dataset [9] as accepted statements in the training data. Each accepted statement was paired with all 4 hard negatives for the corresponding image from [11], to create negative statements. This process created a training set of 3,500 data points.

For LLAMA3-Instruct-8B [4], Qwen-2.5-7B-Instruct [1], QwenVL-2-7B-Instruct [2], and InternVL-2-26B [3], AuraFlow-2 [5] and FLUX.1-dev [6] we used the default temperature settings with 8-bit quantization. For SDXL [12] and FLUX.1-dev [6], we applied the default guidance scale, no negative prompt, and 28 inference steps. For AuraFlow-2 [5], we set a guidance scale of 5 and used 28 inference steps.

All experiments were conducted on A100 and L40s GPUs.

2. Alignment Results

Table 1 represents the text-image alignment scores for different images and evaluated by AIM and baseline metrics.

Image		COM Ads				PSA Ads			
$I_{input-text}$	T2I	AIM	IR	VQA	CS	AIM	IR	VQA	CS
I_{AR}	SDXL	0.50	0.44	0.72	0.23	0.32	-0.05	0.75	0.24
	AuraFlow	0.50	0.00	0.75	0.22	0.31	-0.21	0.72	0.24
I_{LLM}	AuraFlow	0.53	0.29	0.76	0.23	0.43	-0.13	0.69	0.23
I_{real}		0.55	-0.69	0.75	0.26	0.49	-1.03	0.72	0.26

Table 1. Comparison of alignment of the images with action-reason message (AR_m) using our proposed AIM and baseline metrics. Best result per column bolded (including real images). IR = ImageReward. CS = CLIPScore. This result uses InternVL and LLAMA3.

Image		COM Ads				PSA Ads			
$I_{input-text}$	T2I	C_{obj}	C_{LLM}	P_{c+A}	P_{LLM}	C_{obj}	C_{LLM}	P_{c+A}	P_{LLM}
I_{AR}	SDXL	2.03	0.63	0.62	0.60	1.33	0.59	0.48	0.52
	AuraFlow	2.12	0.73	0.64	0.60	1.36	0.72	0.42	0.57
I_{LLM}	AuraFlow	2.25	0.77	0.70	0.76	1.87	0.71	0.60	0.67
I_{real}		2.28	0.67	0.98	0.66	2.04	0.65	0.60	0.61

Table 2. Comparison of persuasiveness and creativity proposed (C_{obj} , P_{c+A}) and baseline metrics. Best gen. image result bolded.

Since I_{real} are the ground-truth images in the dataset and AR_m are the interpretation of real images, we expect I_{real} to have higher alignment score than I_{gen} (i.e. I_{AR} , I_{LLM}). However, the table highlights that ImageReward (IR) [14] and VQAScore (VQA) [10] assign lower scores to I_{real} (ground-truth) compared to generated images. This contradicts our expectation and highlights the limitations of these metrics when evaluating the alignment of images and implicit prompts. CLIPScore (CS) [7] performs reasonably as it rates I_{real} the highest, but it rates the alignment of I_{LLM} and I_{AR} equally. This is problematic because human annotators preferred I_{LLM} in 92% of comparisons, demonstrating its superior alignment. AIM assigns the highest score to I_{real} , followed by I_{LLM} which highlights the higher accuracy of our proposed evaluation method.

3. Creativity and Persuasiveness Baseline Metrics

I_{real} are images from PittAd dataset [9] designed to be creative and persuasive and are expected to be more creative than I_{gen} . As observed in Table 2 the C_{LLM} score for I_{real} , which represents the ground-truth images, is lower than the scores for both I_{LLM} and I_{AR} generated by AuraFlow-2, by 10% and 6% respectively, across both Commercial and PSA advertisements. This raises questions about the reliability of LLMs in accurately scoring creativity, as real images receive lower scores than generated ones. In contrast, our proposed metrics assign the highest scores to I_{real} ,

As illustrated in Fig. 3, top-right example, when the image fails to convey the intended message accurately, the model may hallucinate while generating AR_{gen} . For instance,

<p>I should buy Elizabeth Arden because it will make me feel natural</p> <p>I should buy Elizabeth Arden's Blue Grass because it will make me feel young again</p>			
✗	Human Annotators	✓	✓
0.54	✓	AIM	✗ 0.48
<p>I should buy this soap because it's good for your skin</p>		<p>I should buy this cologne because it's sophisticated</p>	
AR _{gen}			

<p>I should buy this cow because she is nice</p> <p>I should use this service because I can help animals</p>			
✗	Human Annotators	✓	✓
0.51	✓	AIM	✓ 0.51
<p>I should buy Nivea because it's nice to us</p>		<p>I should buy b/bubba cow milk because it is delicious</p>	
AR _{gen}			

<p>I should buy this brand because it will feel very free</p> <p>I should shop at Lacoste because they will make me sporty</p>			
✓	Human Annotators	✓	✓
0.56	✓	AIM	✗ 0.55
<p>I should buy some Lacoste clothes because it's sporty</p>		<p>I should buy this product because it'll protect me from the cold</p>	
AR _{gen}			

<p>I should smoke camels because doctors like to smoke them, and they know what's best.</p> <p>I should BUY CAMEL CIGARETTES because I ENJOY SMOKING CAMELS</p>			
✗	Human Annotators	✓	✓
0.44	✓	AIM	✗ 0.39
<p>I should buy a camel cigarette because it's so smooth it feels like you're getting a massage</p>		<p>I should buy camels because 3 of them are only 7 cents</p>	
AR _{gen}			

Figure 3. Example of failure of AIM in both generating accurate AR_{gen} and choosing the more aligned image with AR_m . **Red text** indicates inaccurate action/reason considering the image. **Green text** shows accurate action/reason. **Red circle** shows inaccurate choice of image for which image is more aligned with the AR_m according to the human annotation.

in this example, the image depicts wind, which can imply coldness, but it does not effectively represent the clothing’s ability to keep someone warm. Due to the unclear intent of the image, the LLM struggles to interpret the message correctly and hallucinates while generating the action-reason statement. Additionally, when the MLLM incorporates textual elements into the image description (as seen in the left image of each example in Fig. 3), the fine-tuned LLM often generates an AR_{gen} that closely resembles the intended AR_m , even if the *visual* content is misaligned. However, due to the unclear or incorrect spelling in the text, human annotators do not consider these images aligned with the statement. To address this issue, we force the MLLM to ignore the texts that starts with “I should” or “I shouldn’t” (which are common starting words for action-reason statements). Additionally, we prompt the MLLM to determine whether the image is text-only. If the image is identified as text-only, a score of 0 is returned to account for the lack of meaningful visual content.

6.2. Limitation

Given the focus of this work, we train and evaluate our methods exclusively on the PittAd dataset [9]. For the evaluation of persuasiveness, we employ an LLM to address seven questions in a single evaluation. While this approach is less efficient, it ensures comprehensive assessment, as the subjective nature of the components makes it essential to in-

clude all components for an accurate evaluation of persuasiveness.

7. Prompts

In this section, we present the prompts used in different evaluation methods. We begin with the prompts for generating descriptions, as shown in Listings 1 and 2. Next, we provide the prompts for generating action-reason statements based on image descriptions in Listings 3 and 4. Then, we present the prompts used for evaluating the creativity and persuasiveness of images with LLMs, detailed in Listings 5 and 6. Finally, we outline the prompts for evaluating individual P_{comp} components, including elaboration, synthesis, originality, imagination, audience targeting, benefit conversion, and appeal category, in Listings 7, 8, 9, 10, 11, 12, and 13.

Listing 1. Description generation for images using InternVL-v2-26B[3] prompt templates in generating train data.

```
Describe the image in detail in one paragraph.
Only return the description. Do not include any further explanation.
```

Listing 2. Description generation for images with MLLMs prompt templates for inference. If the answer to Q1 is “No”, then the AIM score for that image is 0.

```
Carefully analyze the image and respond only in the specified format, without any interpretations or inferences. Focus on only the visible elements in the image. Ensure that any object seen in the image is included in Q1, even if it is described in more detail in Q2.
```

Response Format:

```
Q1: ${answer to Q1}
Q2: ${answer to Q2}
```

Questions:

```
Q1: Are there any objects in the image, excluding text-only logos, and text? List at most 5 such objects if present.
Q2: Describe the image in detail, focusing only on visible objects and elements without adding any interpretation, opinion, or analysis.
```

Listing 3. Prompt for Action-Reason Prediction with zero-shot LLAMA3 [4] and fine-tuned QwenLM [1]. {Blue} denotes elements added dynamically, based on the description of the image.

```
What is the correct interpretation for the described image:
```

Description: {description}.
The interpretation format is: I should
\${action} because \${reason}. ONLY RETURN A
SINGLE SENTENCE IN THIS FORMAT

Listing 4. Prompt for Action-Reason Prediction with fine-tuned LLM[4] in AIM evaluation. {Blue} denotes elements added dynamically, based on the description of the image.

What is the correct interpretation for the
described image:
Description: {description}.

Listing 5. Prompt for evaluation of creativity with LLMs. {Blue} denotes elements added dynamically, based on the description of the image.

USER:
Description: {description}
Question: Score the creativity of the
described image in range of (0, 5).
Your output format is only Answer: score\n
form, no other form. Empty is not allowed.
ASSISTANT:

Listing 6. Prompt for evaluation of persuasiveness with LLAMA3-Instruct[4]. {Blue} denotes elements added dynamically, based on the description of the image.

USER:
Context: If the image convinces the
audience to take an action, it is
considered persuasive.
Description: {description}
Question: Based on the context score the
persuasiveness of the described image in
the range of (0, 5).
Your output format is only Answer: score\n
form, no other form. Empty is not allowed.
ASSISTANT:

Listing 7. Prompt for scoring elaboration with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the
image based on the descriptions.
You are given the description of an image
and you are asked to score the image based
on the following question.
Description: {description}
How visually detailed is the image? Do not
consider the text in the image. Return a
score in the range (0,5).
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}
ASSISTANT:

Listing 8. Prompt for scoring synthesis with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the
image based on the descriptions.
You are given the description of an image
and you are asked to score the image based
on the following question.
Description: {description}
How well does the image connect the
objects that are usually unrelated? Return
the score in range (0,5).
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

Listing 9. Prompt for scoring originality with LLM. {Blue} denotes elements added dynamically.

Context: You are supposed to score the
image based on the descriptions.
You are given the description of an image
and you are asked to score the image based
on the following question.
Description: {description}
How out of the ordinary, and unique the
image is, and how well it breaks away from
habit-bound and stereotypical thinking?
Return a score in range (0, 5).
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

Listing 10. Prompt for scoring imagination with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the
image based on the descriptions.
You are given the description of an image
and you are asked to score the image based
on the following question.
Description: {description}
Assume you are a human evaluating the
given image. How well does the image allow
you to form images you have not directly
experienced before more easily? Do not
consider the text in the image. Return a
score in the range (0,5)
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

Listing 11. Prompt for scoring how well the image targets audience with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the
image based on the descriptions.

You are given the description of an image and you are asked to score the image based on the following question.
Description: {description}
How well the image targets {audience}?
Return a score in range (0, 5)
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

Listing 12. Prompt for scoring how well the image converts features to benefits for customer with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the image based on the descriptions.
You are given the description of an image and you are asked to score the image based on the following question.
Description: {description}
How well does the image connect the feature of the products to the benefits for customers? Return the score in range (0,5).
Please follow the format of:
Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

Listing 13. Prompt for scoring how well the image targets correct appeal category with LLMs. {Blue} denotes elements added dynamically.

Context: You are supposed to score the image based on the descriptions.
You are given the description of an image and you are asked to score the image based on the following question.
Description: {description}
To answer the question consider the following explanation:
There are three types of rhetorical appeals or ways to convince the audience:

- Ethos is a persuasive technique that appeals to an audience by highlighting credibility. Ethos advertisement techniques invoke the superior character of a speaker, presenter, writer, or brand.
- Pathos is a persuasive technique that tries to convince an audience through emotions. Pathos advertisement techniques appeal to the senses, memory, nostalgia, or shared experience.

- Logos is the persuasive technique that aims to convince an audience by using logic and reason. Also called the logical appeal, logos examples in advertisements include the citation of statistics, facts, charts, and graphs.

Question: Assume you are a human evaluating the image. Based on the context score, how well the image appeals to {appealcategory} in the range of (0,5).
Please follow the format of:

Explanation: \${explanation}
Answer: \${answer}

ASSISTANT:

**Previewing Answers Submitted by Workers**

This message is only visible to you and will not be shown to Workers.

You can test completing the task below and click "Submit" in order to preview the data and format of the submitted results.



Evaluation of Advertisement Images

Please do not complete more than 5 forms.

Instruction of part 1

This experiment compares advertisement images based on their persuasiveness and image-text alignment. Each image is based on an advertisement message with two components:

- Action: What action the audience should take
- Reason: Why the audience should take that action

Part A: Persuasiveness Comparison

In this task, you are given 2 images and asked to compare the images on how persuasive they are.

When comparing the persuasiveness, consider the following:

- Which image better convinces you to take the action in the action-reason statement?
- How well the reason in the image aligns with the reason in the given action-reason statement?

Part B: Creativity Comparison

In this task, you are given 2 images and asked to compare the images on how creative they are.

Part C: Action Representation Comparison

Next, you are asked to compare two images based on how well they represent the action in the given action-reason statement. Ignore the persuasiveness, and focus only on the action representation.

Part D: Reason Representation Comparison

Finally, compare each two images and choose which image better represents the reason in the given action-reason statement. For example, if the reason is "Because it is fresh", consider how well each image represents the freshness.

Note: Do not make assumptions about the texts in the images. Ignore typos, but if the text does not make sense, do not infer its meaning.

For the following questions, one option is "Equal". However, do not select "Equal" for 30% of the questions.

Example:

Next, we will show two images as an example of how to compare their persuasiveness and alignment with the action-reason statement.

Example Image 1



Example Image 2



Action-reason statement

I should driving a Subaru because they have great brakes

Comparison of the alignment of images with the action-reason statement

Both Image 1 and Image 2 are showing a car. Image 1 is showing a Subaru car in a road, convincing the audience to drive a Subaru car. Image 2 is simply showing a Porsche with no further element.

While Image 1 is conveying Subaru cars are reliable and have great brakes, by showing brake system in the bottom of the iamge and using the "Stop with confidence" slogan, Image 2 does not depict any specific reason for why the audeicne should buy a Porsche car.

So, Image 1 better aligns with both action and reason in the sentence.

Comparison of the persuasiveness of the images

Image 1 is picturing both a brake system and a Subaru car in a road, and is using a slogan to convince the audience to buy Subaru car because the cars are reliable.

On the other hand, Image 2 is simply showing the car without any futher visual element to make it intresting.

So, Image 1 is better convincing the audience to buy Subaru car using the same reason in the action-reason statement.

Comparison of the persuasiveness of the images

Image 1 is picturing both a brake system and a Subaru car in a road, and a slogan -Stop with confidence.

On the other hand, Image 2 is simply showing the car without any futher visual element to make it creative.

Image 1 has more details, elements, and because of the slogan it is more original. So, Image 1 is more creative.

Questions - Part 1**Part A: Persuasiveness**

Instructions

Shortcuts

Which image is more persuasive considering the given action-reason statement?



Action-Reason Statement:

I should buy 7up Because its popular

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image is more persuasive considering the given action-reason statement?



Action-Reason Statement:

I should drink 7up because is is sugar free and the light of the party

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image is more persuasive considering the given action-reason statement?

Action-Reason Statement:

I should buy this soda Because they sell fast

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Part B: Creativity Comparison

Instructions

Shortcuts

Which image is more creative?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Part C: Alignment with Action

Instructions

Shortcuts

Which image better represents the action in the given action-reason statement?

Action-Reason Statement:

I should buy 7up Because its popular

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image better represents the action in the given action-reason statement?

Action-Reason Statement:

I should drink 7up because is is sugar free and the light of the party

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image better represents the action in the given action-reason statement?

Action-Reason Statement:

I should buy this soda Because they sell fast

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Part C: Alignment with Reason

Instructions

Shortcuts

Which image better represents the reason in the given action-reason statement?

Action-Reason Statement:

I should buy 7up Because its popular

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image better represents the reason in the given action-reason statement?

Action-Reason Statement:

I should drink 7up because is is sugar free and the light of the party

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image better represents the reason in the given action-reason statement?

Action-Reason Statement:

I should buy this soda Because they sell fast

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instruction of part 2

In this experiment, you are given two images and three messages, and you are asked to answer the following questions:

Part A - Maslow's Pyramid of Needs

In the next section, you are given three messages and two images, and you are asked to choose which need in the Maslow's Pyramid each image or message targets:

The definition of each need is as follows:

- **Physiological needs:** Biological requirements for human survival (e.g., air, food, shelter).
- **Safety needs:** people want to experience order, predictability, and control in their lives (e.g., security, safety)
- **Love and belonging needs:** emotional need for interpersonal relationships, affiliating, connectedness, and being part of a group (e.g., intimate relationships, friends)
- **Esteem needs:** Self-worth, respect, prestige, and a feeling of accomplishment.
- **Self-actualization needs:** Realization of a person's potential, self-fulfillment, seeking personal growth, and peak experiences.

Please note that you should choose the need for each image or message independently from others. You can choose multiple needs.

Part B - Rhetorical Appeals

In the next section, you are first asked to choose the types of rhetorical appeals or ways to convince the audience: ethos, pathos, or logos.

- **Ethos** is a persuasive technique that appeals to an audience by highlighting credibility. Ethos advertisement techniques invoke the superior “character” of a speaker, presenter, writer, or brand.
- **Pathos** is a persuasive technique that tries to convince an audience through emotions. Pathos advertisement techniques appeal to the senses, memory, nostalgia, or shared experience.
- **Logos** is the persuasive technique that aims to convince an audience by using logic and reason. Also called “the logical appeal,” examples of logos in advertisements include the citation of statistics, facts, charts, and graphs.

Next, you are asked to choose the image that better appeals to each category. To do so, you should consider the design and components of each image. And based on that, choose the image.

Part C - Benefit Scores and Audiences

In this section, you are asked to compare each of the two images in terms of the following factors.

- **Benefit:** Which image better connects the features of the advertised product to benefits for consumers? Your comparison should be based on how well each image shows the impact of the product in improving consumers' lives.
- **Audiences:** Which image better targets a specific group of audiences? Each image can be designed to target a specific group of audiences by representing their needs and what they like or dislike. In this question, you are asked to find the target audience of each image based on the components represented in each image and then choose which image better targets the given group of audiences.

Part D - Creativity Questions

In the next section, you are asked to compare each two image in terms of the following factors:

- **Originality:** How rare and surprising are the elements in the image?
- **Elaboration:** How complicated is the image? How distant is the image from the basic idea? How detailed is the image?
- **Synthesis:** How well normally unrelated objects have been connected in the image?
- **Artistic Value:** How attractive are the colors and shapes in the image? Does it include a verbal impression?
- **Imagination:** How well does the image help the audience to form vivid mental images or make something unreal come to life?

Please do not select 'equal' or 'none of the above' for more than 30% of the questions.

Now we show examples of each part.

Part A: Maslow's Pyramid of Needs - examples

Example of an advertisement message that targets Physiological Needs

I should drink DASANI water because I cannot live without water.

This advertisement message demonstrates the Physiological need in Maslow's hierarchy of needs. The ad expresses that water is very important and it is impossible to survive without it. This ad fits the Physiological need because water is something that people will always need, and thirst is something that people will always have.

Example of an advertisement message that targets Safety Needs

I should get this insurance because not having insurance is as dangerous as using the air dryer in the bathtub.

This advertisement message exemplifies the Safety needs in Maslow's hierarchy of needs. The strategy of this ad is to convey that if you rent a home or an apartment and do not have insurance, it is equivalent to doing the most dangerous activities in your home. This ad fits the Safety need because people feel secure and safe when they have insurance.

Example of an advertisement message that targets Love and Belonging Needs

I should use Bumble because I need human connection.

This advertisement message focuses on the desire to be connected with people and have confidence in life. Brand marketing communication revolves around the need to find a partner and a sense of belonging.

Example of an advertisement message that targets Esteem Needs

I should buy a new BMW because I want to release the power.

To maintain a sense of prestige, exclusivity, and status, the advertised message does not wish to come across as an everybody-brand. The ad is symbolized by BMW, which wants to stand out and offer a sense of accomplishment to its buyers.

Example of an advertisement message that targets Self-actualisation Needs

I should buy Nike because I want to move more and better.

The advertisement message conveys that we should improve our performance on the track and to be more competitive. It is targeting the Self-actualization need in Maslow's pyramid.

Example on advertisement image that targets Physiological Needs



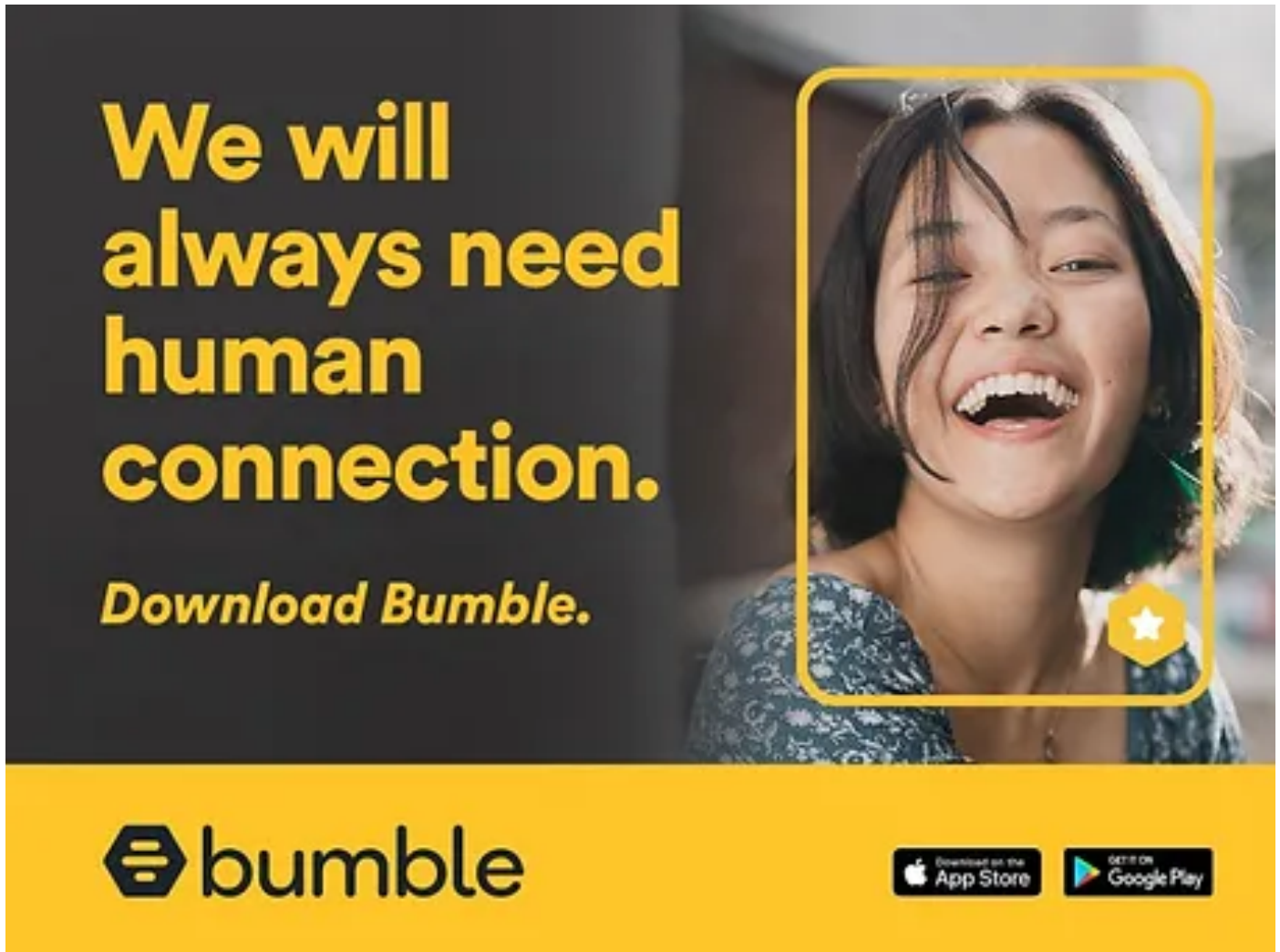
This ad demonstrates the Physiological need in Maslow's hierarchy of needs. The ad expresses that water is very important and it is impossible to survive without it. This ad fits the Physiological need because water is something that people will always need, and thirst is something that people will always have.

Example on advertisement image that targets Safety Needs



This ad exemplifies the Safety need in Maslow's hierarchy of needs. The strategy of this ad is to convey that if you rent a home or an apartment and do not have insurance, it is equivalent to doing the most dangerous activities in your home. This ad fits the Safety need because people feel secure and safe when they have insurance.

Example on advertisement image that targets Love and Belonging Needs



**We will
always need
human
connection.**

Download Bumble.

bumble

Download on the
App Store

GET IT ON
Google Play

This dating app ad focuses on the desire to be connected with people and have confidence in life. Brand marketing communication revolves around the need to find a partner and a sense of belonging.

Example on advertisement image that targets Esteem Needs



A silver BMW 5 Series sedan is shown from a front-three-quarter view, driving on a road at night. The car's headlights are on, and the background is dark with some light trails. In the top right corner, there is a BMW logo and the text "Sheer Driving Pleasure". Below the car, the text "The all-new BMW 5 Series" is visible on the front license plate area.

**RELEASE THE
POWER WITHIN.**

THE ALL-NEW BMW 5 SERIES.

Exclusive drive-away offer.
Monthly installment Rs.185,584*
Available ex-stock.

Contact our sales team for more information.
Teran - 0766 914 441, Dilanka - 0773 487 434
Viraj - 0773 711 444, Suneth - 0773 733 733

Prestige Automobile (Pvt) Ltd., 234-238, Pannipitiya Road, Battaramulla, Sri Lanka.
Tel: +94 112 075 666 | Email: sales@prestigeauto.lk

*Conditions apply

To maintain a sense of prestige, exclusivity, and status, the advertised brand does not wish to come across as an everybody brand. The ad is symbolized by BMW, which wants to stand out and offer a sense of accomplishment to its buyers.

Example on advertisement image that targets Self-actualization Needs



The ad is encouraging us to improve our performance on the track and to be more competitive. It targets the Self-actualization need in Maslow's pyramid.

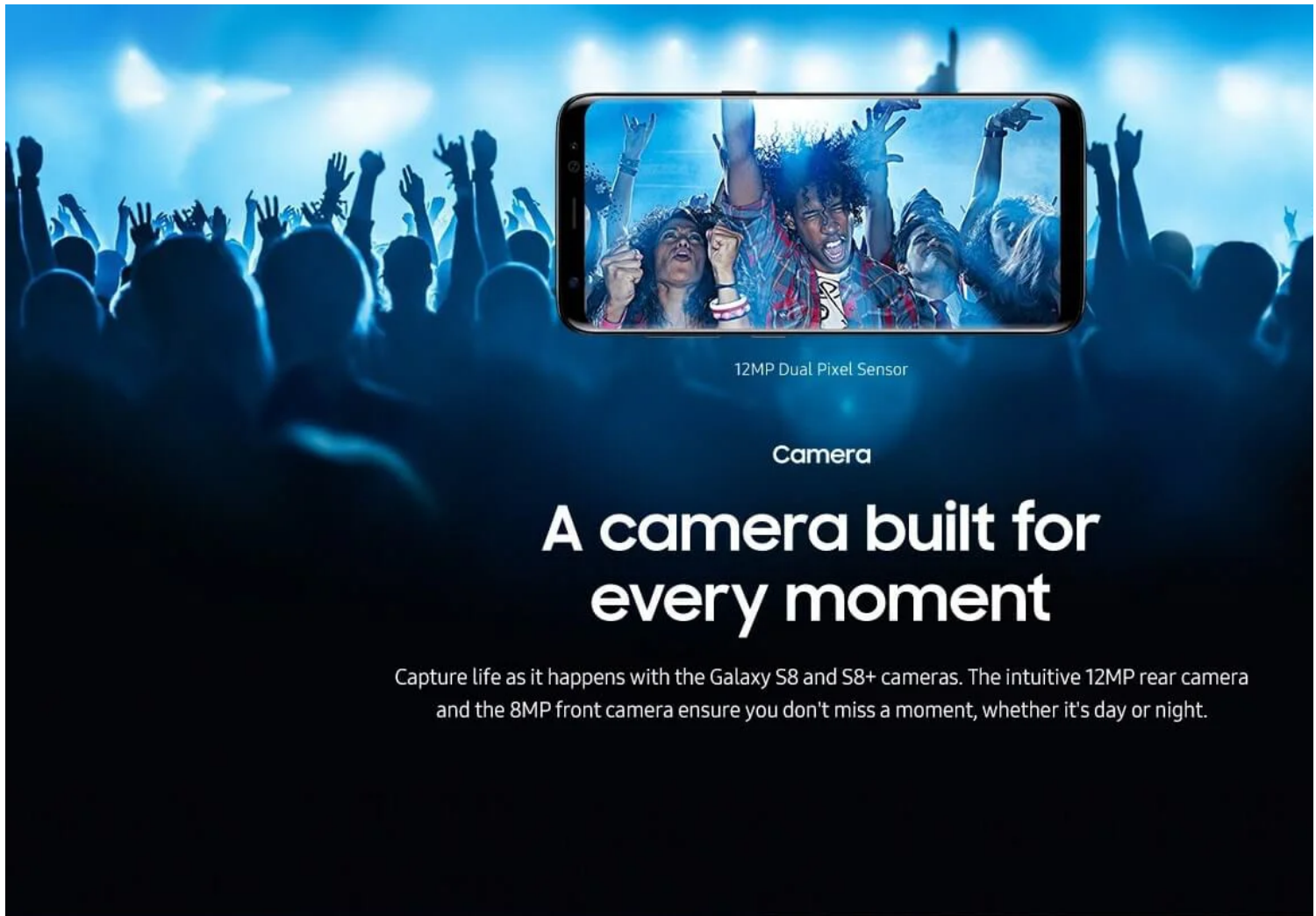
Part B - Rhetorical Appeal - examples

Example of advertisement image with ethos appeal category



This example uses an image of a celebrity holding the Smart water to establish credibility and authority.

Example of advertisement image with logos appeal category



12MP Dual Pixel Sensor

Camera

A camera built for every moment

Capture life as it happens with the Galaxy S8 and S8+ cameras. The intuitive 12MP rear camera and the 8MP front camera ensure you don't miss a moment, whether it's day or night.

This image uses the causal relation between the image on the left side and the right side to invoke logic in the audience. The image suggests if the audiences do not stop buying cigarettes, they will get the oxygen mask for free as a result of lung disease. So, this image uses a causal relationship between the events to convince the audience.

Example of advertisement image with pathos appeal category



The image includes the slogan "Open a Coke, open happiness," inferring that drinking Coke causes happiness, which targets emotions in the audience.

Example of comparison of the images on which image appeals to ethics, logic, or pathos more.



The image on the right side is simply showing the product without adding more detail. On the other hand, the image on the left side shows a famous actress promoting the product, which highlights its credibility. So, the image on the left side appeals to ethos more.

Part C - Benefit Score and Audiences - examples

Example of comparison of the images on which image better connects the features in the product to the benefits for the audiences.

**STORAGE FOR
1GB OF MP3S**



**1,000 SONGS
IN YOUR POCKE**



The image on the left side directly shows the features of the item; however, the image on the right side shows the benefits of those features. So, the image on the right side is better, connecting the features of the product to the benefits for customers.

Which image better targets families?

-I SHOULD DRIVE DRYE DORD
BERUE6 GOOD **FOOD**
IN MY FAIMILE BEIKS AFFFOCABLE
FOX TIAR MY BUY 'E
FORD **FRS AMBLE**
11 **1 1 FORE**
d **CARD**
HO
HEGANCE
tr. + + + + of MANY
BEOLLING D Y S E 595 6 FORD
BE # 701 6 REY WE ONFF AN?INTIY D!
BEQU S BE AR BE BIG FAMILY



ONE BI
ONE Bi

Image on the left is simply showing a car surrounded by some text with unknown audience. On the other hand, image on right is showing a family with a Ford car, and is using the slogan of "ONE BIG FAMILY, ONE BIG FORD" directly targeting families. So, image on the right side, better targets families.

Part D - Creativity Questions - examples

We now show two example images and explain below how we might compare them in terms of:

- Originality
- Elaboration
- Synthesis
- Artistic
- Imagination

Example Image 1



Example Image 2



Originality: Which image is more out of the ordinary, unique, and better breaks away from habit-bound and stereotypical thinking?

Image 1: The image is depicting the coca-cola in the grass field, using a unique slogan of 'open Coke, open happiness' which is more unique, out of the ordinary and better breaks away from habit-bound and stereotypical thinking than the image 2 which simply shows number of coca-cola cans in the image.

Synthesis: Which image better connects the objects that are usually unrelated?

None of the images represents the connection between unrelated objects; however, Image 1 includes multiple different objects in the image so it better connects different objects than Image 2 which is only showing a single object.

Elaboration: Which ad has more details?

Image 1 has many more details (grass field, sky, Coke, 'open Coke, open happiness', etc.) than Image 2 which is a number of Coke cans.

Imagination: Which ad allowed you to form images you have not directly experienced before more easily? Which image better made something unreal come to life?

Image 1, with all the details and scenes it has depicted, better helps with forming the mental images. This is because it has much more detail in the image. The more detailed the image is, the easier it is to form the mental image seeing the image.

Artistic Value: Which image is more visually/verbally distinctive and better makes ideas come to life graphically/verbally?

Image 1 shows the Coke in a sense with different objects with many colors rather than just showing the Coke cans in the image. So, image 1 is more artistically produced compared to Image 2 with a single object in the image and fewer colors used.

Questions - Part 2**Part A - Maslow's Pyramid of Needs****Advertisement Messages**

- 1. I should wear Athletica clothing because they empower women.
- 2. I should use this activewear because I will feel more motivated.
- 3. I should buy Athleta clothing because they are empowering.

☐ Physiological Need ☐ Safety Needs ☐ Love and Belongingness Need ☐ Esteem Needs ☐ Self-actualization Needs ☐ None of the above ☐ Other

If chose other:

Short answer text

Additional Comments (Optional)

Short answer text

Advertisement Images



☐ Physiological Need ☐ Safety Needs ☐ Love and Belongingness Need ☐ Esteem Needs ☐ Self-actualization Needs ☐ None of the above ☐ Other

If chose other:

Short answer text

Additional Comments (Optional)

Short answer text

Generate Abvviiset

7A moult UP- Sooda is
made eeks 77s Sooda are
pudonds le Ada mill
bur is hay llyh JPP!

• I Soolul UP ed a
freesa totl Free Pao
oreaulq his Sela Sodar

• Isocola Drinkes and
Incurat: Freed...

• Aleaul: Buy Drink Buy
begea Buy pscuas
votae hnt's Soda in
and fre.

• Agcevald fe Silen
Bas 770 yo 77es Up at the
Kopoe ch Think the
Fegul Fas fel

A Besoululn you
Pucce Wepder f goe
D orce ra fogt Seel

• I f free feilur
the Laus to in nee asect
neegus of the Poes to
oumra pphor f fre
ref Ab d
Wepoe Te Neipe fresae
hnt ugent or 6 korm hnt
fee de. soul Sooda

• Tibhs Sool UP, f Tye Fep Up
to D'o- Sed Be o f eoon
hure hns wld hto Assy
soda to pzoas sod soda



Advortlenierfts Bu

• It Shnould is seusse up
resulcag / 2n deen 77 o fts
teah of untk Sed de dcln

• IS Sola Buy pccasse
77in Ategr.

• The souulo free the ne
oury up, fei Aul's 7 lber
upryu gmag.

• A Ssdol oita free resave
uugdor the gelftes in dees
sccer the totl Sga sere
te blinc dnk soul it.

• Is. is zeme the 77 fie be
veagr thad oas and fre

• I oeah .7 Sek to ia free
frees m for thn teout
Sed yart ell 77 UP 77s.

• Teosede Eholuase in
oe egru te

• I Beegst Eo paece vntoso
tagg poeul ofrs a se ilnie
Sggind ITH Bap.

• I Teo thre fe SilneTe up
vare paval sool ade to uy
sag Sic!

• 7 / 1 I Teoc Tee Tremon
sccomea toni lgh is fee
piseu hnt to teie prep
sts ia year Sooda.

7UP SODA

☐ Physiological Need ☐ Safety Needs ☐ Love and Belongingness Need ☐ Esteem Needs ☐ Self-actualization Needs ☐ None of the above ☐ Other

If chose other:

Short answer text

Additional Comments (Optional)

Short answer text

Part B - Rhetorical Appeal

InstructionsShortcuts

Which type of rhetoric appeal is best represented in this image? Even if no type of appeal is definitely utilized in the image, choose the one that is more relevant than the others.

Image 1



Select an option

Ethos	1
Logos	2
Pathos	3
None of the above	4

Submit

Optional: Further Comments

InstructionsShortcuts

Which type of rhetoric appeal is best represented in this image? Even if no type of appeal is definitely utilized in the image, choose the one that is more relevant than the others.

[illegible]

Ethos	1
Logos	2
Pathos	3
None of the above	4

Submit

Optional: Further Comments

Now compare the images and choose which image better appeals to ethos/pathos/logos.

Instructions

Shortcuts

Which image appeals to ethos more?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image appeals to logos more?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image appeals to pathos more?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Part C - Benefit Score and Audiences

Instructions

Shortcuts

Which image better connects the feature of products to the benefits for costumers?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Which image better targets health-conscious costumers?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Part D - Creativity Questions

Instructions

Shortcuts

Originality: Which image is more out of the ordinary, unique, and better breaks away from habit-bound and stereotypical thinking?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Synthesis: Which image better connects the objects that are usually unrelated?

⌵

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Elaboration: Which ad has more details?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Imagination: Which ad allowed you to form images you have not directly experienced before more easily? Which image better made something unreal come to life?

{

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Instructions

Shortcuts

Artistic Value: Which image is more visually/verbally distinctive and better makes ideas come to life graphically/verbally?

Image 1



Select an option

Image 1	1
Image 2	2
Equal	3

Submit

Optional: Further Comments

Thank you so much for completing this form. Please provide any additional comments on the questions.

Optional: Further Comments

References

- [1] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023. 1, 3
- [2] Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-vl: A versatile vision-language model for understanding, localization. *Text Reading, and Beyond*, 2, 2023. 1
- [3] Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24185–24198, 2024. 1, 3
- [4] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. 1, 3, 4
- [5] Fal. Introducing auraflow v0.1, an open exploration of large rectified flow models, 2024. Available at: <https://blog.fal.ai/auraflow/> [Accessed: 2024-11-07]. 1
- [6] FLUX. Flux.1 dev: Frontier ai image generator model, 2024. Available at: <https://flux1ai.com/dev> [Accessed: 2025-03-07]. 1
- [7] Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7514–7528, 2021. 1
- [8] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022. 1
- [9] Zaem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka. Automatic understanding of image and video advertisements. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1705–1715, 2017. 1, 3
- [10] Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation. In *European Conference on Computer Vision*, pages 366–384. Springer, 2025. 1
- [11] Sina Malakouti, Aysan Aghazadeh, Ashmit Khandelwal, and Adriana Kovashka. Benchmarking vlms’ reasoning about persuasive atypical images. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4788–4798. IEEE, 2025. 1
- [12] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. In *The Twelfth International Conference on Learning Representations*. 1
- [13] Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation. In *Forty-first International Conference on Machine Learning*. 1
- [14] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36, 2024. 1