Performance and Cost Analysis Report

Table of Contents

Overview

This report analyses the performance of AI Marketer V2 in generating social media marketing content using various input configurations, focusing on caption quality, cost-efficiency, accuracy, and practical applicability.

1. Test Objective

The goal of this test is to evaluate the effectiveness of different input methods for caption generation and to inform the future development direction. Specifically, the following aspects were analysed:

- Cost vs. Content Quality
- Suitability for Marketing Objectives
- Information Accuracy and Relevance
- Balance Between User Input Effort and Output

2. Test Methodology

The test consisted of three scenarios to generate marketing content for the tapas bar "Vinos":

1. Cheese Platter Promotion: Product highlight purpose
2. Sangria Limited-Time Promotion: Time-limited discount offer
3. Scallop & Fish and Chips Combo Deal: Product bundle discount promotion

Each scenario was tested using different input methods, considering the following variables:

- AI Caption Generator Model: GPT-4o vs. GPT-4o-mini
- Input Information for AI Caption Generator:
  - Basic Information: Configurable in Business Settings (Name, Type, Target Customers, Vibe)
  - Additional Prompt: Entered by users

    - Object Detection: Image analysis results from AWS Rekognition (user-modifiable)
    - Image Input: Uploaded by users
    - Item Descriptions: Name entered by users and the corresponding description retrieved from Square.

The cost is based on 1,000 caption sets, each containing 5 captions.

3. Key Findings

3.1 Model Selection: GPT-4o vs. GPT-4o-mini

See results 1-a and 1-b in the link

GPT-4o resulted in a cost of $2.31 for 1,000 runs, which is more than 10 times higher than GPT-4o-mini ($0.18). However, the actual performance difference was minimal. This indicates that high-cost models are not necessarily needed for marketing content generation.

Conclusion: GPT-4o-mini is more cost-efficient for marketing content generation, reducing costs by approximately 90% without compromising quality.

3.2 Product Promotion (Cheese Platter Scenario)

See results 1-b to 1-f in the link

Testing with various input methods yielded the following results:

- **Basic Information** only results in a general marketing message like "Savor the night with our artisanal tapas," but it lacks specific mention of the cheese platter (cost: $0.18)
- **Additional Prompt** method with "Cheese sharing platter" generates a caption like "Indulge in the perfect pairing of flavors with our artisanal cheese sharing platter," incorporating the target product info effectively at a low cost ($0.18)
- **Object Detection** method identified elements such as "Weapon, Cheese, Bread, Candle," requiring user modification (e.g., changing "Weapon" to "Cheese, Cracker") (cost: $0.19, $1.19 after 12 months)
- **Image Input** method generated basic descriptions like "artisanal cheese board," but failed to reflect product characteristics adequately despite higher cost ($0.63)
- **Item Description** method provided a detailed caption like "exquisite cheese platter, featuring a curated selection of four artisanal cheeses paired with delightful condiments" ($0.19)

Conclusion: The additional prompt and item description methods provide the most effective cost-to-benefit ratio for product promotion. The object detection method requires user intervention due to inaccuracies in recognising elements, and image input method shows limited performance improvement despite higher costs.

3.3 Time-Limited Promotion (Sangria Discount Scenario)

See results 2-a to 2-e in the link

Testing with various input methods yielded the following results:

- **Basic Information** using a prompt like "3-5pm 20%" led to a vague result like "Join us at Vinos from 3-5 PM for a cozy tapas experience and enjoy 20% off your favorites" (cost: $0.17)

- **Additional Prompt** with "3-5pm 20% Sangria" provided the precise message "20% off all Sangria from 3-5 PM" ($0.21)
- **Object Detection** method identified elements like "Soda, Alcohol, Cocktail," resulting in a wrong message like "20% off on all cocktails and drinks" (cost: $0.20, $1.20 after 12 months)
- **Image Input** method recognised the Sangria image but failed to significantly improve the promotion message delivery despite the higher cost ($0.60)
- **Item Description** method accurately provided "Vinos Sangria (white, rose, or red), 500ml $22.0, 1L $38.0" and generated "20% off on 500ml and 1L options" ($0.20)

Conclusion: In time-limited promotions, additional prompt or item description methods are the most effective. The object detection method leads to incorrect information, such as mistakenly identifying "Sangria" as "Cocktail, Alcohol," which carries the risk of generating an incorrect caption. The image input method produces results similar to the additional prompt method, but instead of requiring extra effort from the user, it costs three times more.

3.4 Product Bundle Promotion (Scallop & Fish and Chips Scenario)

See results 3-a to 3-e in the link

Testing with various input methods yielded the following results:

- **Basic Information** using a prompt like "3-5pm 20%" led to a vague result like "Join us at Vinos from 3-5 PM for a cozy tapas experience and enjoy 20% off your favorites" (cost: $0.17)
- **Additional Prompt** with "scallops fish n chips combo deal 10% off" generated an effective message like "our exclusive scallops fish n chips combo—now 10% off" ($0.19)
- **Object Detection** only recognised "Food presentation," resulting in a vague "combo deal" expression, leading to an unsuitable outcome ($0.16, $1.16 after 12 months)
- **Image Input** method failed to recognise complex food combinations despite higher cost ($0.58)
- **Item Description** method provided detailed item information like "fish & chips [Regular $23.0], pan-seared Hokkaido scallops [4pcs $20.0]" and calculated the 10% discount, providing accurate pricing ($0.24)

Conclusion: For complex product bundles, the additional prompt and item description methods are most effective. Object detection only produced generic expressions, and image input method struggled to recognise the combination of items correctly. Item descriptions were especially useful in accurately calculating prices and offering detailed information.

4. Performance and Cost Comparison by Input Method

Input Method	Cost (Per 1,000 Sets)	Advantages	Disadvantages	Suitable Use Cases
Basic Information	$0.16-0.18	- Low cost - Minimal user effort	- Lack of product specificity	General marketing not specific to image or product

Input Method	Cost (Per 1,000 Sets)	Advantages	Disadvantages	Suitable Use Cases
Additional Prompt	$0.17-0.21	- Low cost - Effective for specific product promotions - Simple to input	- Requires user input	Simple product promotions
Object Detection	$1.16-1.20 (after 12 months)	- Basic object recognition - User can modify recognised items	- Inaccurate or unnecessary recognition - Requires user efforts - Increased cost after 12 months	Unsuitable (costly, unreliable, requires effort)
Image Input	$0.58-0.63	- No user input required - Recognises visual elements	- High cost - Failure in complex combinations - Limited performance improvement	Simple product promotions for high-cost receptive users
Item Description	$0.18-0.24	- Reflects detailed information like price and components - Accurate discount calculation	- Square API integration needed - Item info must be pre-registered - Requires user input	Complex product promotions require accurate pricing

5. Test Conclusions

- **Model for AI Caption Generator:** GPT-4o-mini is the most cost-effective model, reducing costs by approximately 90% without compromising quality.
- **Input Method:**
    - **Additional Prompt:** Cost-effective and generates effective marketing content with simple keywords but requires user input.
    - **Object Detection:** Inaccurate recognition and high cost make it unsuitable.
    - **Image Input:** Struggles with complex images and has limited performance improvement for the high cost.
    - **Item Description:** Ideal for complex promotions requiring precise information like price, composition, and discount calculations, though it requires prior registration and user input.

6. User Segmentation

Based on the test conclusion, I suggest dividing users into two groups:

- Basic Users
  - Users who prioritize lower costs and are willing to put in some effort.
  - Provide services using basic information, additional prompts and item description methods.
- Premium Users
  - Users who prioritize a no-effort experience, even at a higher cost.
  - Provide services using additional prompts and image input method.

However, an improvement to the image input method is necessary, as the current method does not fully meet expectations.

7. Improvement Plan for Image Input

Business-specific model training is required. This involves training separate models for each business (e.g., Vinos) and fine-tuning them through a Fine-tuning API.

- Train the model with the business's social media posts to learn the tone.
- Train the model with the business's menu names and descriptions.
- Train the model with the business's menu photos to enable learning with images.

The benefit is that users would no longer need to input data manually but can generate captions directly from images. However, effective training requires a sufficient amount of labeled data, so a plan for this must be developed. The cost will be approximately twice the current testing costs, with additional training costs.

Cost	**Basic GPT 4o Mini**	**Fine tuned GPT 4o Mini**
1M input tokens	$0.15	$0.30
1M output tokens	$0.60	$1.20
1M training tokens	N/A	$3.00
1000 caption sets	**$0.58-0.63**	**$1.16-1.26 (training cost extra)**