

Case 1 VS Case 2

Llava_next showed better performance (only 6 mismatches) in Case 2 & (18 mismatches) in Case 1

Qwen2.5 and Llava_onevision showed similar error counts in both cases

In **Case 1**, same category/same product errors were more common for Qwen2.5 and Llava_next

In **Case 2**, Llava_onevision had a higher proportion of different category errors compared to Qwen2.5

Model	Case	Total Mismatches
Qwen2.5	1	64
	2	55
Llava_onevision	1	68
	2	57
Llava_next	1	18
	2	6

Case 1

Input : Original Advertisement, Product, Category, Selling Point Options

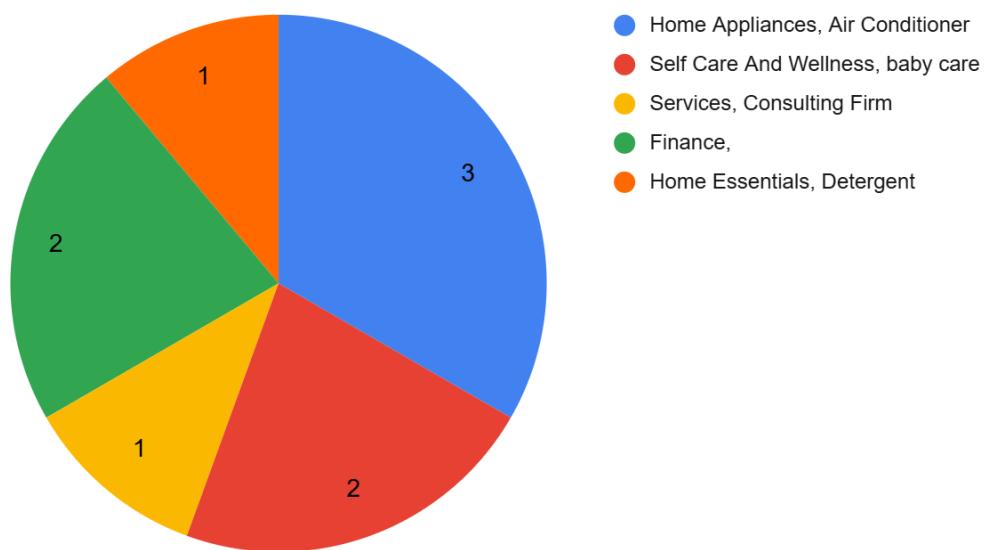
- **Option 1:** Correct selling point - The actual selling points for the specific product in the advertisement
- **Option 2:** Same category but different product - Selling points from a different product in the same category
- **Option 3:** Same category and same/similar product - Same or similar product (sharing similar keywords) in the same category
- **Option 4 & 5:** Different categories - Selling points from products in completely different categories

Common Files

All three models picked the **SAME WRONG SELLING POINTS** for these files - **9 images**

Dove_baby_products_11, Hindustan_Unilever_413,
Huggies_diapers_and_wipes_7_it_001, Kotak_Life_8, LG_refrigerators_432,
lic_insurance_008, mitsubishi_air_conditioners_703, Soft_And_Cool_2

common wrong answers across all three models



Qwen2.5 & llava_onevision picked the **SAME WRONG SELLING POINTS** for these files - **32 images**

at&t_001, beauty_001, Blue_Star_air_conditioners_411, car_brands_423_BMW,
clothing_002, daikin_702, Dell_laptops_and_computers_15, dove_365,
Dove_baby_products_11, food_007, ford_400, Hindustan_Unilever_413,
home_essential_003, honda_civic_001, hotel_001,
HP_laptops_and_computers_21, HRX_43, Huggies_diapers_and_wipes_7,
ICICI_Bank_412, irvasa_204, it_001, Kotak_Life_8, LG_air_conditioner_45,
LG_refrigerators_432, lic_insurance_008, mitsubishi_air_conditioners_703,
Soft_And_Cool_2, State_Bank_Of_India_16, State_Bank_Of_India_406,
Tide_pods_008, volvo_001, Yatra_429.

same wrong option selected (qwen2.5 & llava_next)

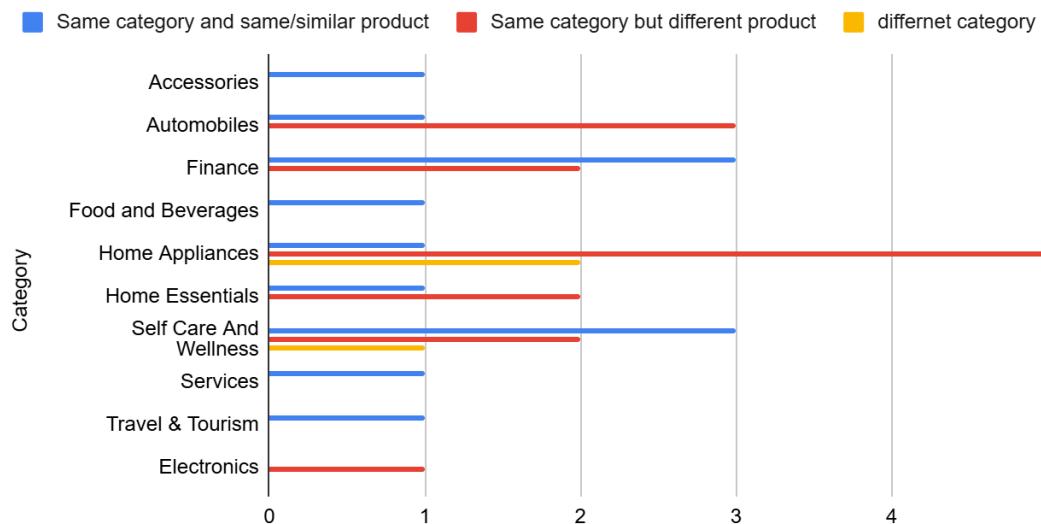
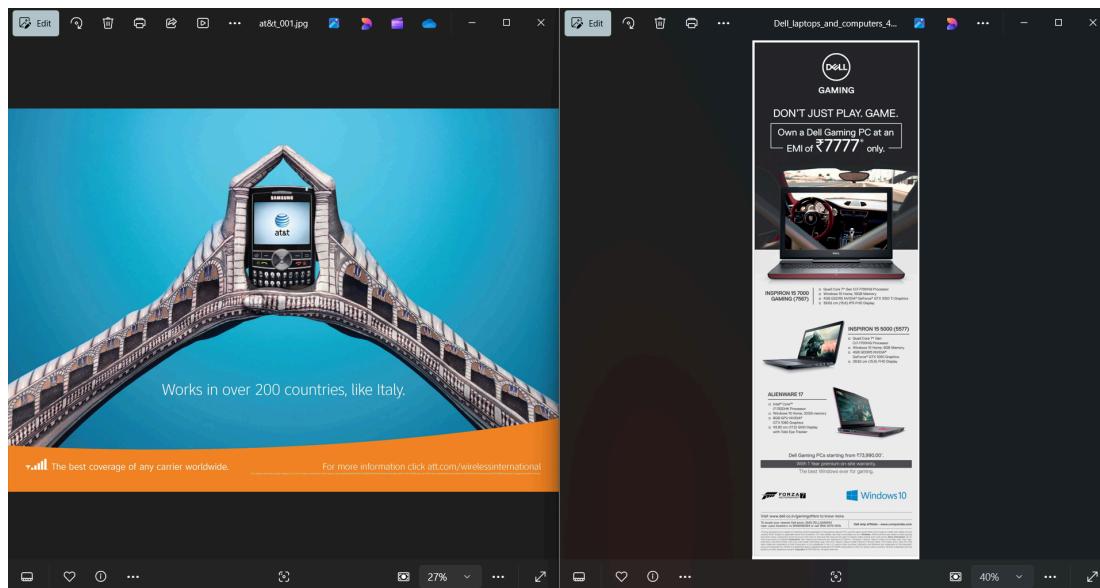


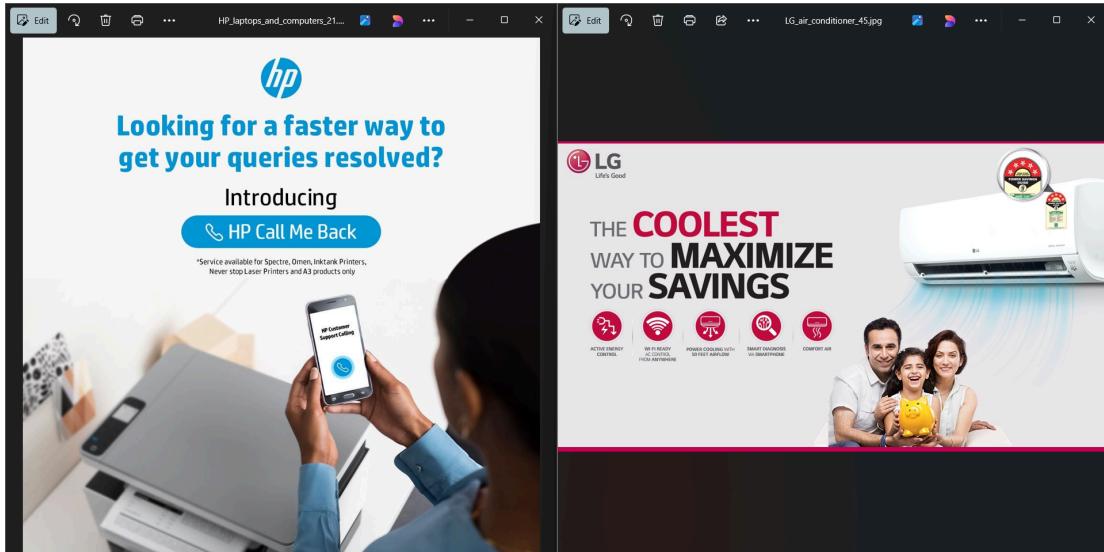
Image examples - ones on the left are the actual images, ones on the right are the advertisements which have the selling points the vlm picked

Different categories

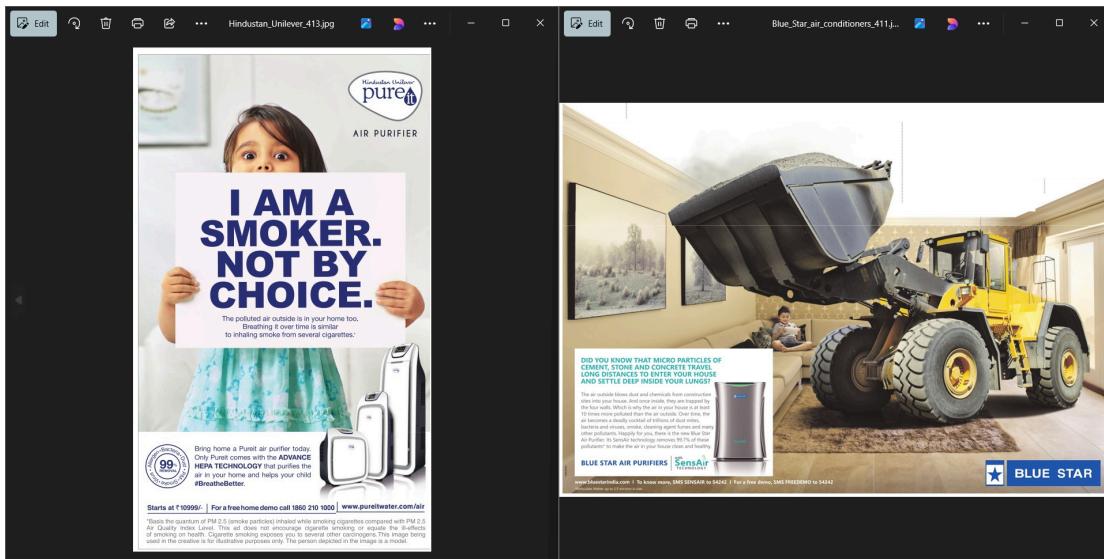
Services & Electronics



Services & Home Appliances

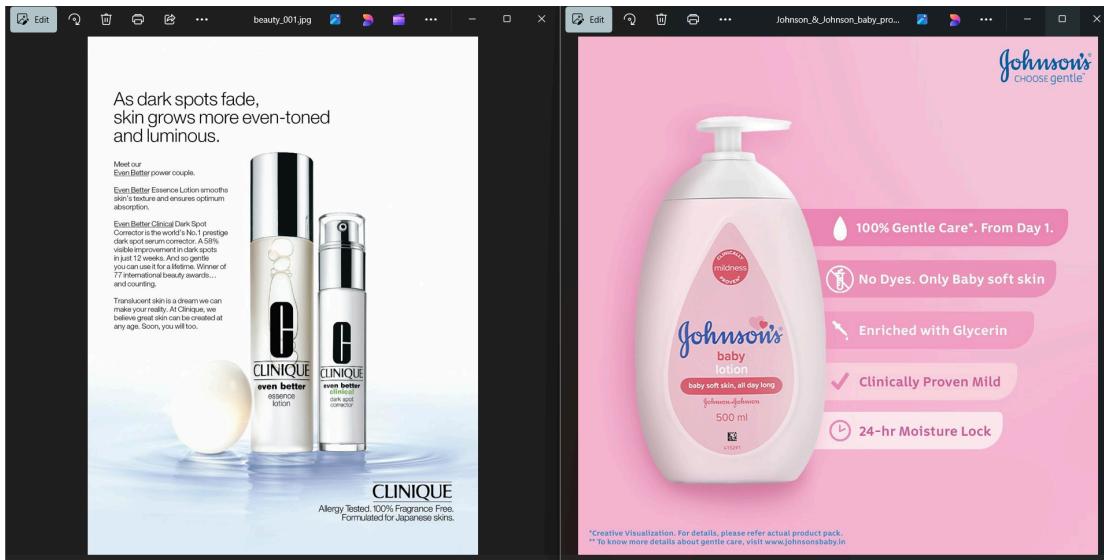


Same category, same product (Home Appliances, air purifier)



Same category : Self Care And Wellness but different product

skin care & baby care



Summary for case 1

Although similarity based on selling points is less the wrong option was picked -

- model might be prioritizing visual elements in the advertisement over textual similarity between selling points
- model may understand individual concepts but struggle to connect them properly in context
- There is confusion between similar products, such as cars and bikes, mobile devices and laptops, and air conditioners and air purifiers.
- Difficulty in matching correct selling points even when the product is correctly identified

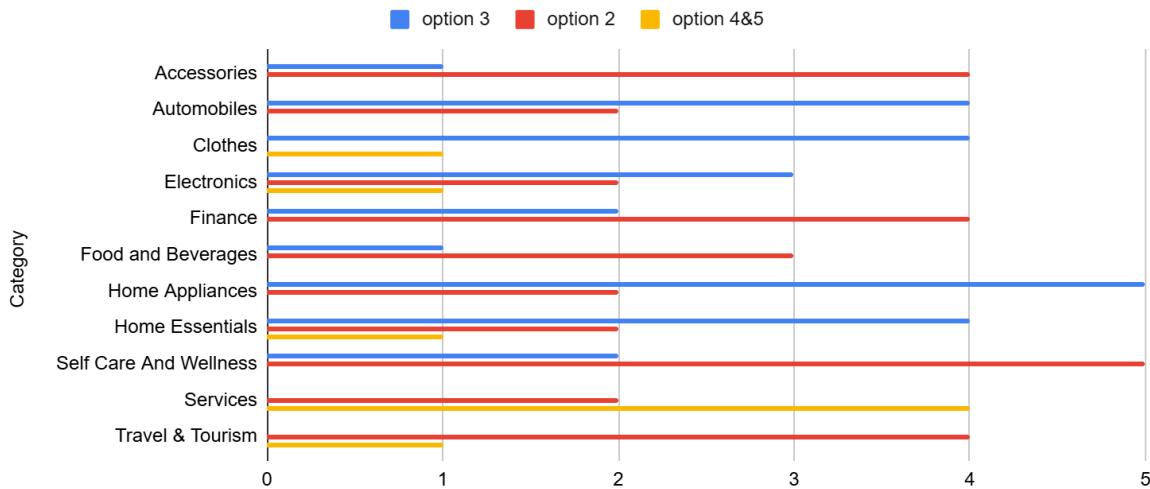
Qwen2.5 (64 mismatches)

There are no cases in the 350 images where a product is part of two different categories

Count for each option

- **Option 2 (Same Category, Different Product):** 26 cases (40.6%)
- **Option 3 (Same Category, Same Product):** 30 cases (46.9%)
- **Option 4/5 (Different Category):** 8 cases (12.5%)

frequency of picking the wrong option, category-wise (qwen2.5, case 1)



Different Category Mismatches

- Services** (4 cases) {2 mismatched with automobiles (car); 1 electronics (laptop); 1 home appliances (AC)}
- Clothes** (1 case) {Self care and wellness, baby care}
- Home Essentials** (1 case) {mismatched fabric softener with shampoo}
- Electronics** (1 case) {mobile to air conditioner}
- Travel & Tourism** (1 case) {travel to air conditioner}

Same Category, Different Product Mismatches

Confusion Pattern	Cases
(Car → Bike)	4
(AC → Purifier)	3
(Mobile → Laptop)	2
(Clothes → Sports Shorts)	2
(Purifier → AC)	1

Categories with misidentified products

Category	Count
Home Appliances	5
Automobiles	4
Clothes	4

Home Essentials	4
Electronics	3
Self Care And Wellness	2
Finance	2
Food and Beverages	1
Accessories	1

- car advertisements are confused with motorcycle advertisements
- confusion between mobile devices and laptops
- difficulty in distinguishing between clothing subcategories

Same Category, Same Product

- Baby Care products - difficulty in matching correct selling points even when the product is correctly identified
- Products with potentially similar visual appearances in advertisements show high confusion rates: Different clothing types, Jewelry and watches

Categories with misidentified products

Category	Cases
Self Care And Wellness	5
Accessories	4
Finance	4
Travel & Tourism	4
Food and Beverages	3
Electronics	2
Services	2
Automobiles	2
Home Appliances	2
Home Essentials	2

Similarity Scores of selling points

Word Overlap similarity is consistently low across both options:

- **Option 2:** 16 of 26 cases (61.5%) have Word Overlap scores of 0-0.018

- **Option 3:** 17 of 30 cases (56.7%) have Word Overlap scores of 0-0.017

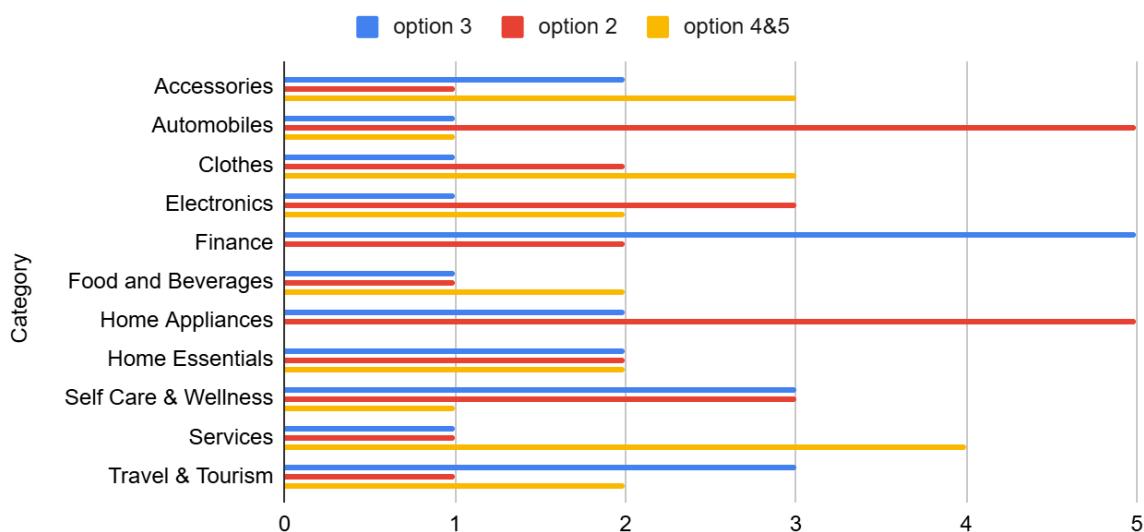
Similarity Metric	Option 2	Option 3	Option 4/5
Similarity_Combined	0.1146	0.1078	0.098
Similarity_Jaccard	0	0	0
Similarity_TFIDF	0.0491	0.0451	0.0252
Similarity_Fuzzy	0.3282	0.3078	0.2992
Similarity_WordOverlap	0.0193	0.0206	0.0097

Llava_onevision (68 mismatches)

Count for each option

- **Option 2 (Same Category, Different Product):** 26 cases (38.2%)
- **Option 3 (Same Category, Same Product):** 22 cases (32.4%)
- **Option 4/5 (Different Category):** 20 cases (29.4%)

frequency of picking the wrong option, category-wise
(llava_onevision, case 1)



Different Category Mismatches

Category	Count

Services	4
Accessories	3
Clothes	3
Electronics	2
Food and Beverages	2
Home Essentials	2
Travel & Tourism	2
Automobiles	1
Self Care & Wellness	1

Same Category, Different Product Mismatches

Mismatched Pairs (product)	Cases
Car → Bike	4
Air Conditioner → Air Purifier	3
Mobile → Laptop	2
Regular Clothes → Sports Shorts	2
Skin Care → Baby Care	1

Categories with misidentified products

Category	Cases
Automobiles	4
Home Essentials	4
Home Appliances	3
Electronics	2
Clothes	2
Self Care And Wellness	2
Finance	2
Food and Beverages	1
Travel & Tourism	1
Services	1

Accessories	1
-------------	---

Same Category, Same Product

Categories with misidentified products

Category	Count
Finance	5
Self Care And Wellness	3
Travel & Tourism	3
Home Appliances	2
Accessories	2
Home Essentials	2
Electronics	1
Clothes	1
Automobiles	1
Food and Beverages	1
Services	1

Similarity Scores of selling points

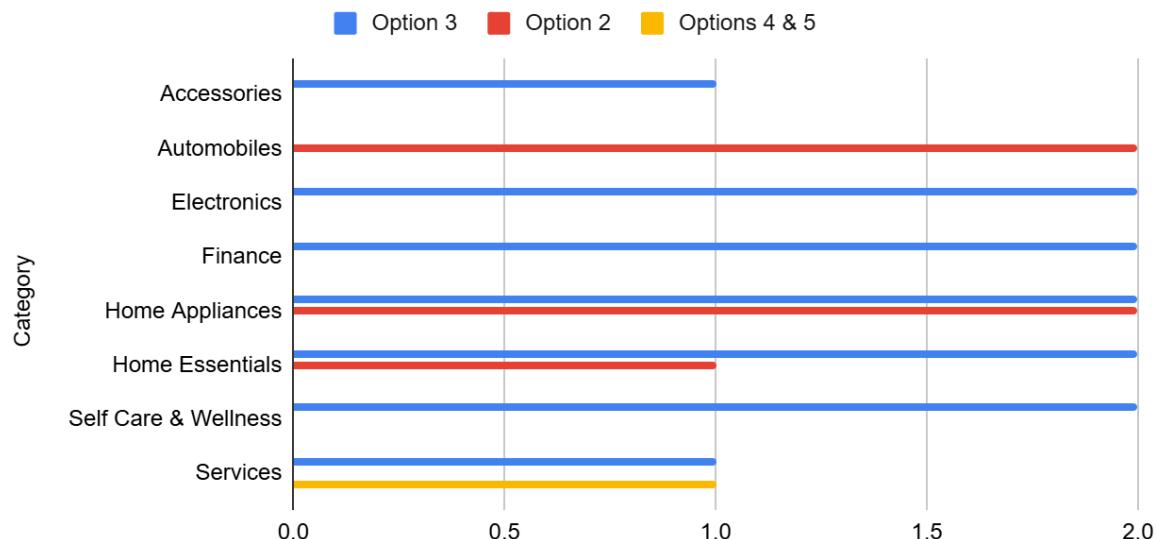
Metric	Option 4/5	Option 2	Option 3
Similarity_Combined	0.0991	0.1194	0.1224
Similarity_Jaccard	0	0	0
Similarity_TFIDF	0.0207	0.0575	0.0746
Similarity_Fuzzy	0.3082	0.3334	0.3231
Similarity_WordOverlap	0.0074	0.0251	0.0342

Llava_next (18 mismatches)

Count for each option

- **Option 4/5 (Different Category):** 1 cases (5.56%)
- **Option 2 (Same Category, Different Product):** 5 cases (27.78%)
- **Option 3 (Same Category, Same Product):** 12 cases (66.67%)

frequency of picking the wrong option, category-wise (llava_next, case 1)



Different Category Mismatches

Services → Automobiles

Same Category, Different Product Mismatches

Mismatched Pairs (product)	Cases
Car → Bike	2
Air Purifier → Air Conditioner	1
Facial tissues → Detergent	1

Categories with misidentified products

Category	Cases
Automobiles	2
Home Essentials	1
Home Appliances	1

Same Category, Same Product

Categories with misidentified products

Category	Count
Electronics	2
Home Essentials	2
Self Care And Wellness	2
Finance	2
Home Appliances	2
Services	1
Accessories	1

Once case where the similarity score is 1 in the selling points. Two different images with same selling point

Case 2

Input : Original Advertisement, Product, Category, Selling Point Options

- **Option 1:** Correct selling point - The actual selling points for the specific product in the advertisement
- **Option 2:** Same category but different product - Selling points from a different product in the same category
- **Option 3, 4 & 5:** Different categories - Selling points from products in completely different categories

Common Files

There are not common files across all three models

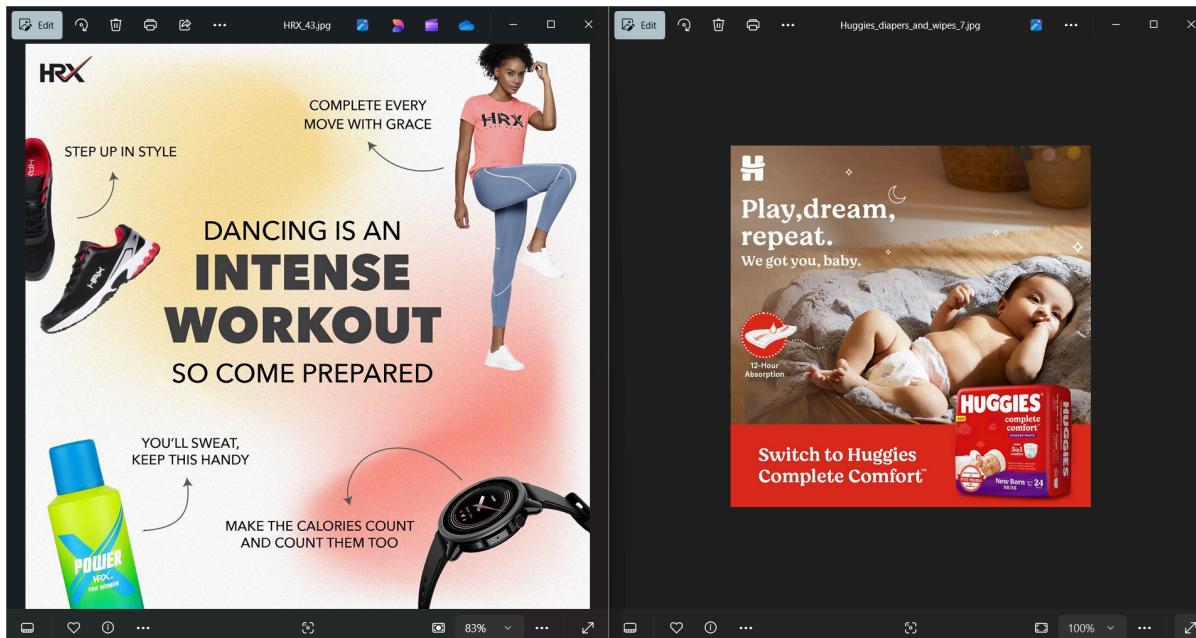
For llava_1onevision & qwen2.5 - 5 files have picked the same wrong selling points

amazon_001, Apple_mobiles_414, Blue_Star_air_conditioners_404,
Blue_Star_air_conditioners_411, ford_400, hollister_001, home_essential_001,
honda_civic_001, hotel_001, hotel_005, HRX_43, Joyalukkas_jewellery_418,

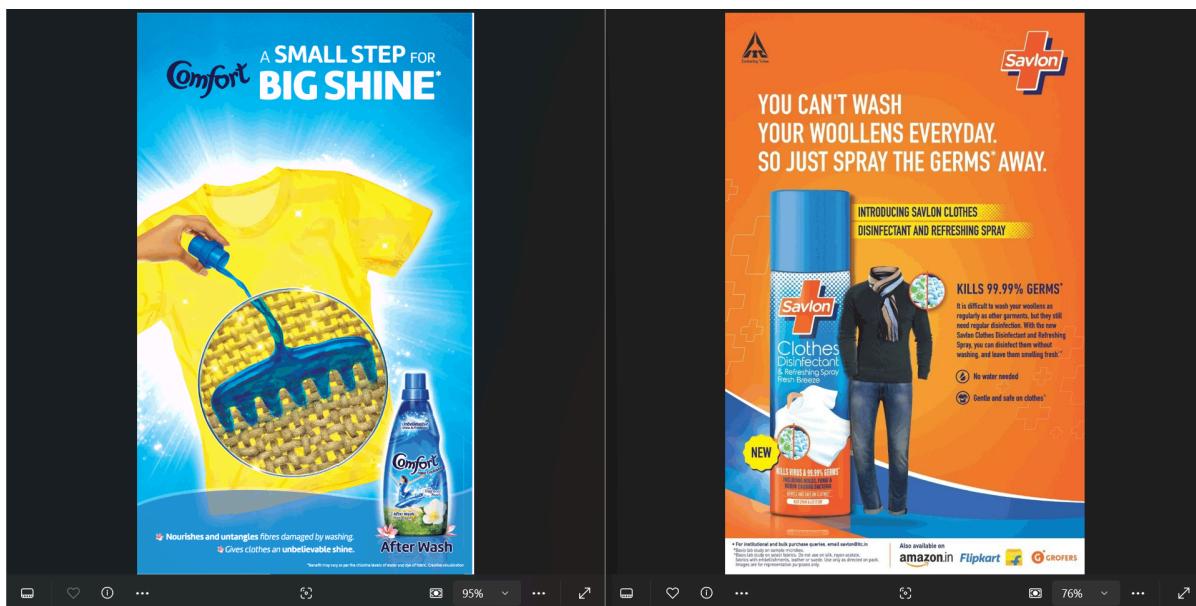
LG_air_conditioner_45, mitsubishi_air_conditioners_703, Rin_detergent_22, skoda_401, State_Bank_Of_India_406, tata_420, Tyka_7, watches_006

Examples images - ones on the left are the actual images, ones on the right are the advertisements which have the selling points the vlm picked

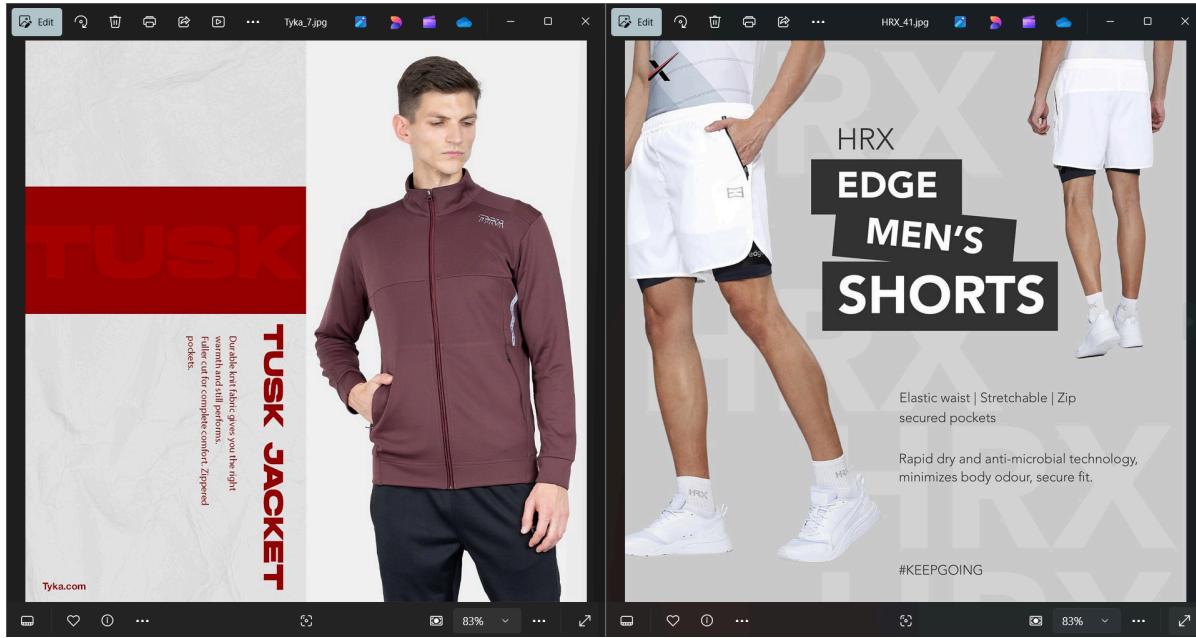
Same category : Self Care And Wellness
Different products : exercise & baby care



Same Category: Home Essentials



Same category : Clothes



Summary for case 2

- Llava_onevision and Qwen2.5 selected the same wrong selling points for 20 files
- Automobiles: Persistent car → bike confusion (highest in Qwen2.5 with 8 cases)
- Electronics: Mobile → laptop confusion
- Home Appliances: Air conditioner → air purifier confusion
- Services and Travel & Tourism had high rates of complete category misidentification

Qwen2.5 (55 mismatches)

Count for each option

- **Option 2 (Same Category, Different Product):** 38 cases (69.1%)
- **Option 3/4/5 (Different Category):** 17 cases (30.9%)

Different Category

Category	Cases
Finance	3
Services	3

Accessories	3
Food and Beverages	2
Travel & Tourism	2
Clothes	1
Electronics	1
Home Appliances	1
Home Essentials	1

Automobiles and Self Care And Wellness had no cases where the VLM selected a completely different category

Same Category

Category	Number of Cases	Notable Confusion
Automobiles	8	Car → Bike
Clothes	6	Shorts → Sports shorts
Electronics	5	Mobile → Laptop, Smart speaker → Laptop
Travel & Tourism	3	Airline → Hotel
Services	3	Delivery services → Logistics
Home Appliances	3	Air conditioner → Air purifier
Self Care And Wellness	3	Various wellness products
Home Essentials	3	Facial tissues → Disinfectant
Food and Beverages	2	Various food products
Finance	2	Bank services → Payment platform

Automobiles, Clothes, and Electronics categories had the highest number of within-category confusions.

Similarity Scores

Similarity Metric	Same Category	Different Category
Combined	0.1118	0.0923

Fuzzy	0.3156	0.2948
TFIDF	0.0498	0.014
Jaccard	0.0013	0
Word Overlap	0.0215	0.0019

Home Appliances and Automobiles had the highest combined similarity scores, indicating selling points in these categories were more textually similar. Electronics had relatively low similarity scores despite having many same-category mismatches, suggesting the model may have been confused by visual features

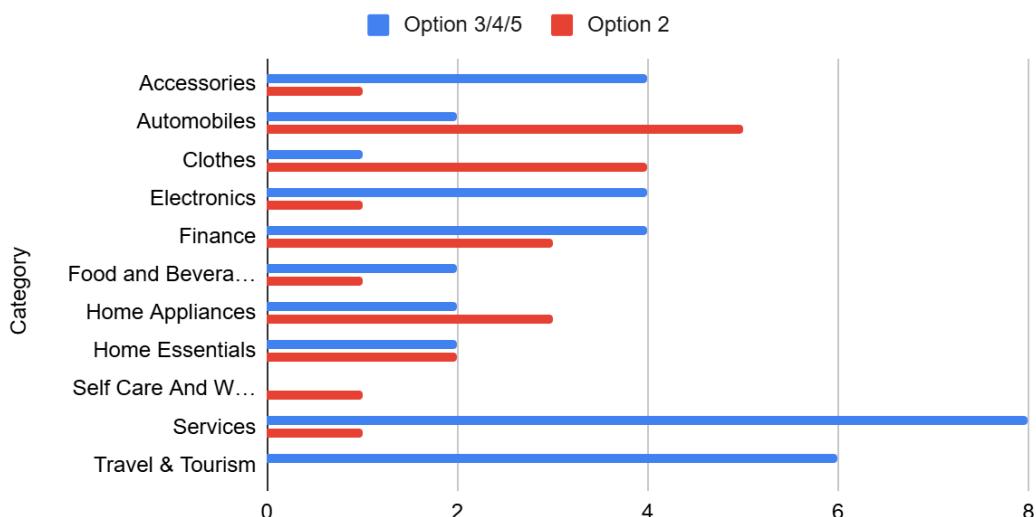
llava_onevision (57 mismatches)

Count for each option

- **Option 2 (Same Category, Different Product):** 22 cases (69.1%)
- **Option 3/4/5 (Different Category):** 35 cases (30.9%)

The model has more cases where it did not even pick the right category

Category Distribution Option 3/4/5 and Option 2,
llava_onevision, case 2



Different Category

Services and Travel & Tourism have very high counts in Option 3/4/5, indicating these categories frequently had both incorrect selections and incorrect

categorization. The VLM seems to struggle significantly with properly recognizing and categorizing items in these domains.

Same Category

Category	Confusion Details
Automobiles	Car → Bike
Clothes	Clothes (general) → Clothes (specific)
Home Appliances	Air conditioner → Air purifier
Finance	Insurance ↔ Bank services
Home Essentials	Disinfectant ↔ Washing (fabric softener)
Services	Delivery services → Logistics
Electronics	Mobile → Laptop
Food and Beverages	Milk → Coffee
Self Care & Wellness	Exercise → Baby care
Accessories	Watch → Jewelry

Several categories show systematic patterns of confusion:

- In Automobiles, car brands were consistently misclassified as Hero Motocorp (motorcycles)
- In Home Appliances, air conditioners were consistently misclassified as air purifiers

Some product types were confused in both directions:

- Insurance ↔ Bank services
- Different types of clothes
- Disinfectant ↔ Washing products

Similarity Scores

Similarity Metric	Same Category	Different Category
Similarity_Combined	0.1103	0.0985
Similarity_Jaccard	0	0
Similarity_TFIDF	0.0427	0.0272
Similarity_Fuzzy	0.3161	0.2997
Similarity_WordOverlap	0.024	0.0089

lava_next (6 mismatches)

Count for each option

- **Option 2 (Same Category, Different Product):** 3 cases(50%)
- **Option 3/4/5 (Different Category):** 3 cases (50%)

Different Category

Services : 2 cases

Accessories : 1 case

Same Category

Home Appliances: 1 occurrence {air conditioner -> fridge}

Electronics: 1 occurrence {laptop -> mobile}

Clothes: 1 occurrence {clothes (sports shorts) -> shorts}

Similarity Scores

Lease similarity across the models in case 2