

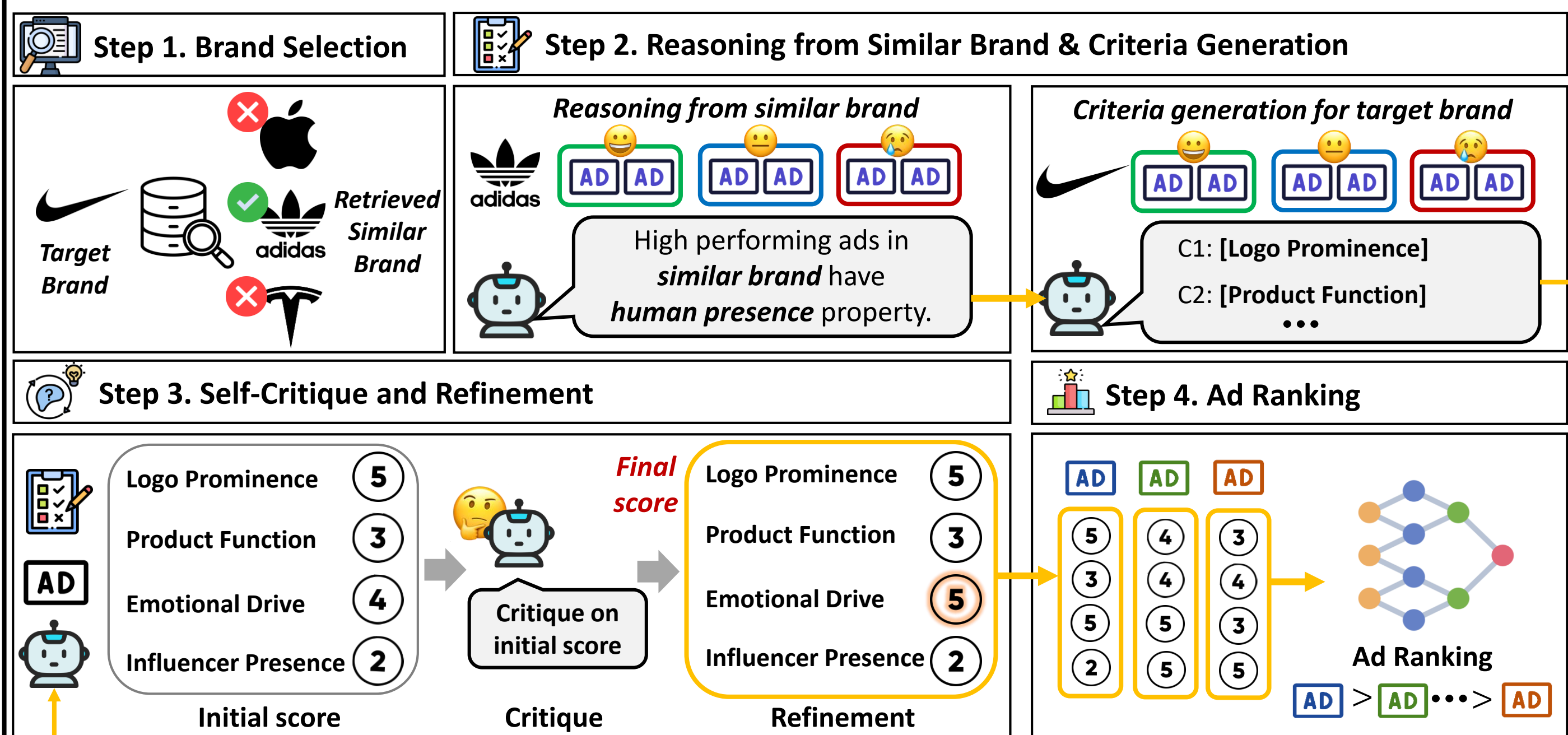
Problem Definition

- New Problem:** Brand-specific advertisement ranking
- Target Brand:** Nike
- Input:** Nike ads



AD #1 AD #2 AD #3
- Output:** Ranking of ads AD #2 > AD #3 > AD #1
Reference: Nike Official Instagram (@nike)

Overall Framework



Background & Problem Definition

- Background:**
 - VLMs struggle with real-world business decisions which requires task-specific reasoning under limited data.
- Problem Definition: Brand-specific advertisement ranking**
 - Goal:** Rank new ads according to their performance metrics before actual ad deployment.
 - Input:** For each brand,
 - Brand description
 - Ads with visual content, textual content
 - Output:** Ranking of ads of the target brand.
- Challenges:**
 - (1) Data scarcity:** Limited ads and performance labels for each brand.
 - (2) Lack of decision criteria:** No clear definition of an effective ad for specific brands.
- Practical Relevance:**
 - Represents real-world tasks of advertising agencies.
 - Enables **effective pre-deployment ad selection** under limited data, reducing costly trial-and-error in real-world ad campaigns.

Proposed Method: ADvisor

- (C1) Brand-specific criteria generation with cross-brand context augmentation**
 - Goal:** To address lack of decision criteria under data scarcity.
 - Construct contrastive few-shot examples (high-medium-low performance ads are grouped).
 - If similar brand exists, augment with similar-brand insights.
 - Use VLM to generate brand-specific evaluation criteria.
- (C2) Reflection-based scoring**
 - Goal:** To improve reliability and consistency of VLM scoring.
 - [Sub-step 1] Initial scoring:** Assign scores on each generated criterion with rationales.
 - [Sub-step 2] Self-critique:** Examine inconsistencies and insufficient grounding.
 - [Sub-step 3] Refinement:** Revise scores based on critique feedback.
- (C3) Brand-specific ranking**
 - Goal:** To learn brand-specific ranking under limited labeled data.
 - Use refined scores as input features and train a brand-specific lightweight ranker (MLP).
 - Optimize the ranker with pairwise ranking loss.

Experimental Results

- RQ1. Accuracy:** ADvisor achieves the best overall ranking performance.

Measure	Metric	MLP			VLM (Zero-shot)			VLM (Few-shot)			DEVL	ECSF	MMF	ADVISOR (Ours)
		T	V	T+V	T	V	T+V	T	V	T+V				
NDCG @1	CTR	42.78	52.56	41.98	31.89	36.09	41.76	36.90	35.15	39.36	39.17	24.54	39.98	52.32
	CPC	55.68	61.33	50.72	49.39	62.12	62.80	55.11	66.05	65.02	64.28	48.53	63.20	68.55
	CPM	57.93	60.96	66.22	63.88	62.83	63.00	68.11	69.00	65.62	72.83	75.43	69.26	70.79
	Avg	52.13	58.29	52.97	48.38	53.68	55.85	53.37	56.74	56.67	58.76	49.50	57.48	63.89
NDCG @3	CTR	55.95	52.58	51.64	44.88	46.88	50.30	45.51	48.14	49.26	47.59	43.43	48.38	56.00
	CPC	60.94	63.03	57.80	56.30	66.93	65.19	59.20	65.25	67.68	62.86	64.50	66.85	67.23
	CPM	66.16	66.58	69.19	70.68	65.85	64.62	71.08	73.55	69.96	72.85	76.70	70.23	73.21
	Avg	61.01	60.73	59.55	57.29	59.89	60.04	58.59	62.31	62.30	61.10	61.54	61.82	65.48
NDCG @5	CTR	63.50	56.92	56.66	53.91	58.09	54.46	53.94	57.19	53.56	54.37	52.14	57.86	63.84
	CPC	63.58	66.68	64.31	63.16	68.70	67.67	64.68	67.70	67.40	69.76	67.12	70.03	70.23
	CPM	73.58	74.50	73.29	71.93	69.98	67.88	72.16	73.57	73.02	73.22	79.40	74.91	76.45
	Avg	66.89	66.03	64.75	63.00	65.56	63.33	63.60	66.16	64.66	65.78	66.22	67.60	70.17
Total	Avg	60.01	61.68	59.09	56.22	59.71	59.74	58.02	61.74	61.21	61.88	59.09	62.03	66.51

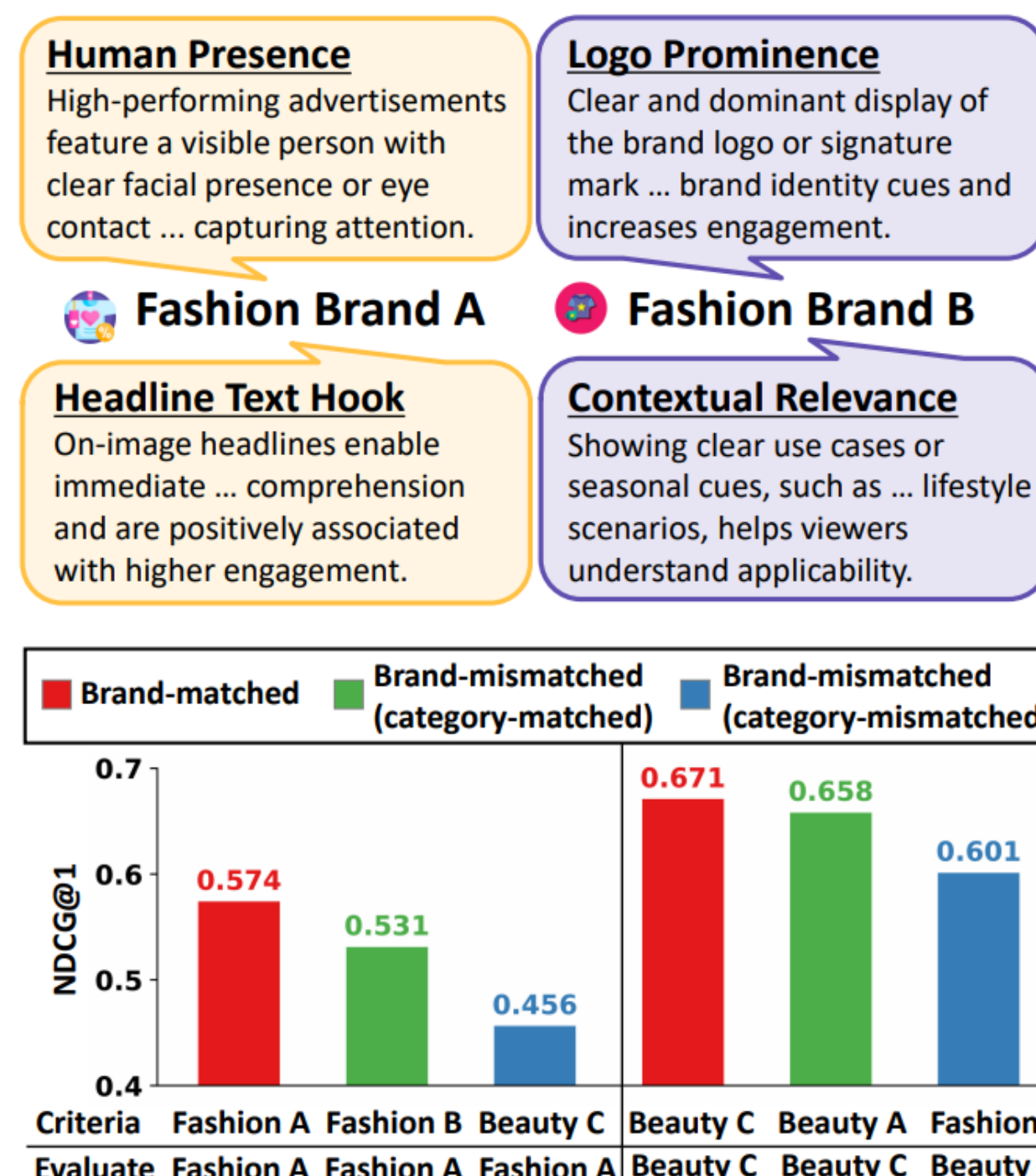
CTR: Click-Through-Rate / CPC: Cost Per Click / CPM: Cost Per Mille

- RQ2. Ablation study:** All key components of ADvisor are effective.

Method	NDCG@1	NDCG@3	NDCG@5
ADVISOR	63.89	65.48	70.17
-AB	54.10	60.32	65.13
-CB	56.25	62.02	67.82
-PR	57.78	64.20	68.49
-RE	44.22	51.37	58.58
-RA	51.29	58.56	63.35

- CB:** without cross-brand context augmentation.
- PR:** without private performance labels.
- AB:** using all brands for context augmentation without selection.
- RE:** without self-critique and refinement.
- RA:** without a trainable ranker (MLP).

- RQ3. Case study:** Generated brand-specific criteria are both meaningful and effective.



- RQ4. Online A/B testing:** ADvisor demonstrates real-world effectiveness against human experts.

Method	CTR ↑	CPC ↓	ROAS ↑
Human Marketers	8.37%	428	1,070%
ADVISOR	10.14%	231	1,219%
Improvement (%)	21.15%	46.03%	13.93%

