The Face of Persuasion: Analyzing Bias and Generating Culture-Aware Ads

Aysan Aghazadeh
University of Pittsburgh
aya34@pitt.edu

Adriana Kovashka
University of Pittsburgh
kovashka@cs.pitt.edu

Abstract

Text-to-image models are appealing for customizing visual advertisements and targeting specific populations. We investigate this potential by examining the demographic bias within ads for different ad topics, and the disparate level of persuasiveness (judged by models) of ads that are identical except for gender/race of the people portrayed. We also experiment with a technique to target ads for specific countries. The code is available at https://github.com/aysanaghazadeh/FaceOfPersuasion.

1 Introduction

Advertisements have great significance: they affect perceptions on a variety of topics, from products to politics and societal values. Given recent progress on generative models, their use for AI-created ads is imminent. These models could in theory customize ads, targeting specific populations through demographically diverse content. We investigate both the promise of generating diverse visual ads with text-to-image diffusion models and the bias in assessing the resulting images (i.e., scoring their persuasiveness) by Large Language Models (LLMs) and Multimodal LLMs.

We begin with an investigation of gender and race bias in an existing dataset (Hussain et al., 2017). We compare to bias in ads generated with three text-to-image models: DALLE3 (Betker et al., 2023), FLUX (Black Forest Labs, 2024), and AuraFlow (Fal, 2024). We find that both the dataset and generated images exhibit racial bias: for example, Black individuals are greatly underrepresented in clothing and shopping ads.

We then run controlled experiments where we only alter one demographic feature in ads keeping the rest of the details and quality the same, and study how persuasiveness judgments vary with gender and race. For example, in Fig. 1, the model chooses the image with a white woman as more persuasive because it appears *"more elegant"*.

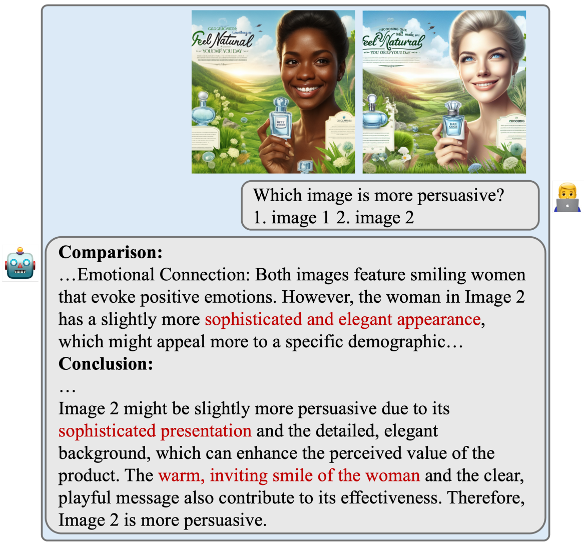


Figure 1: Selection of the more persuasive image by InternVL (Chen et al., 2024b). Image 1 features a Black woman; Image 2 is a White woman. InternVL selected Image 2 as more persuasive. Red marks reasoning bias.

Second, we attempt to create ads that convey a particular message and are tailored toward a particular culture/country. An ad aimed at a Japanese audience may benefit from featuring an Asian person or Japanese cultural symbols, but resonate less and be less effective with an United Arab Emirates audience. We experiment with a technique that incorporates symbols from other ads in the generation process and shows promising results.

Our contributions are: (1) We analyze demographic bias in both the PittAd dataset[1] and generative models for persuasive content creation, across different advertisement topics. (2) We demonstrate bias in LLMs and MLLMs when selecting the most persuasive images, revealing preference patterns based on demographic attributes. (3) We propose *CulGen*, a culture-aware image generation method for producing advertisement images addressing specific cultural/regional contexts.

[1] relevant publications cited over 400 times

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 6472–6500

2 Related Works

Bias in T2I models. (Cho et al., 2023; D'Incà et al., 2024) introduce a framework to assess bias in T2I models. (Bianchi et al., 2023; Naik and Nushi, 2023) study bias over different professions. Instead, we evaluate bias in persuasive generation.
Bias in LLMs. (Mire et al., 2025) studies the bias of reward models for LLMs against African American language compared to White English. (Wan et al., 2023) assess bias in AI-generated reference letters. (Sheng et al., 2021; Dinan et al., 2020; Liang et al., 2021) analyze the social bias in language generation. (Ye et al.) assess the bias in LLMs as evaluation methods. However, our focus is specifically on creative content.
Bias in MLLMs. (Janghorbani and De Melo, 2023) introduces a framework for evaluating the social bias in Vision-Language Models and (Wang et al., 2022) introduces a tool for evaluating bias in datasets. (Zhao et al., 2021) analyzes the bias in image captioning and (Hirota et al., 2022; Fraser and Kiritchenko, 2024) in visual question answering on topics such as occupation. Instead, our work focuses on the evaluation of persuasion.
Culture-Aware Image Generation. (Hutchinson et al., 2022; Jha et al., 2024) study the cultural bias in T2I models. (Alsudais, 2025) analyzes the representation of different nations in daily tasks. (Mukherjee et al., 2025) introduces a dataset to evaluate the cultural understanding, and stereotypical representation in MLLMs and T2I models. (Mukherjee et al., 2025; Khanuja et al., 2024) propose a method to edit the image to target a specific culture. Our work is on the generation of images from a text prompt (message), instead of editing an input image. We are the first to study the relation between *persuasion* and bias in generative models.

3 Method

3.1 Analyzing diversity in real/generated ads

First, we investigate bias in existing ads using the PittAd dataset (Hussain et al., 2017) which contains advertisement images with topic annotations such as clothing, human rights, etc. We infer demographic features (gender and race) using DeepFace (Taigman et al., 2014) on images showing humans. We compute the overall distribution of each race and gender in the dataset and further break it down into distributions of races and genders per topic.

Next, we generate ad images using an annotation in PittAd: abstract message interpretations for each ad, structured as *'I should **[action]** because **[reason]**'* and referred to as action-reason statements (AR). We use these statements as prompts to three text-to-image models: DALLE3 (Betker et al., 2023), Flux (Black Forest Labs, 2024), and AuraFlow (Fal, 2024). To analyze the effect of prompt expansion, we also generate a detailed description of a possible ad corresponding to an AR, using LLAMA3-instruct (AI@Meta, 2024), then use the output as another prompt for AuraFlow. We repeat the demographic analysis on generated ads.

3.2 Evaluating Persuasion Bias via Demographic Swaps

To assess how the demographics of the humans in the ads influence persuasiveness judgments by LLMs and MLLMs, we conducted a controlled experiment. We created sets of images that were identical except for the race and gender of the central individual. We used GPT4.1 to generate an ad based on the AR, and also obtained a description of the image using GPT4o. We then used the same models to modify the image and description to edit the race/gender and keep all else the same. These image-description pairs were then evaluated by MLLMs and LLMs that were prompted to select the more persuasive option using chain-of-thought (CoT) reasoning (Wei et al., 2022). Specifically, we use GPT4o (OpenAI, 2024), QwenVL-2.5(7B) (Bai et al., 2025), QwenLM-2.5(7B) (Hui et al., 2024), InternVL-2.5(7B) (Chen et al., 2024b) and InternLM-2.5(7B) (Cai et al., 2024). MLLMs consistently favored images featuring White individuals, often justifying their choices with subjective attributes such as perceived elegance (Fig. 1 & 3).

3.3 Diversifying through country targeting

The target audience plays a critical role in persuasion (Usman, 2013). However, given existing biases in text-to-image (T2I) models, the ability to generate ads tailored to different countries remains an open question. To support this, and analyze the cultural bias in advertisement data, we first introduce an extension to PittAds (Hussain et al., 2017), which includes up to three predictions for the target country of each image and its cultural components, both from InternVL (Chen et al., 2024b) instructed to focus on language and addresses in the image.[2]

[2]Human evaluation shows this approach achieves a recall of 81% and a precision@1 (P@1) of 72% in inferring the correct countries. When grouping countries by similar cultural regions, scores improve to 94% recall and 75% P@1.

	Real					Flux					Dalle3					Auraflow					Llama3				
T	W	L	A	B	M	W	L	A	B	M	W	L	A	B	M	W	L	A	B	M	W	L	A	B	M
C	66	9	15	6	4	**70**	**12**	4	8	4	64	2	14	6	**14**	47	9	**36**	9	0	24	11	27	**32**	1
S	**92**	0	8	0	0	70	**20**	10	0	0	52	16	8	4	**20**	73	7	7	7	7	45	2	**32**	**18**	2
H	**66**	9	6	9	0	47	5	8	13	**26**	0	0	0	0	0	63	0	0	**25**	13	41	**14**	**22**	12	9
E	**77**	3	14	6	0	64	9	0	18	0	25	0	**75**	0	0	47	**12**	29	12	0	20	4	40	**28**	**3**
O	**73**	3	8	13	2	56	**11**	19	10	4	60	6	17	2	**10**	70	4	10	8	4	43	8	**26**	**18**	3

Table 1: Diversity of race in Topics: Clothing, Shopping, Human rights, Self-Esteem, Overall. % people shown that look White, Latinx, Asian, Black, Middle-Eastern. Highest value across groups (Real to Llama3) bolded per race.

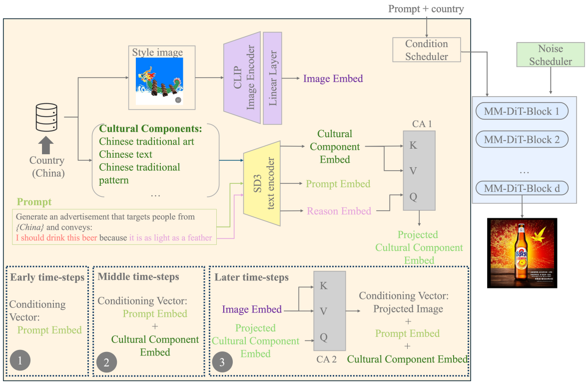


Figure 2: **CulGen** for creating country-targeted ads using cultural symbols from existing ads. CA is cross-attention. The denoising condition is computed based on the time-step at the bottom of the figure, Steps 1, 2, and 3, while embeddings for Condition Scheduler are generated in upper side. MM-DiT block and noise scheduler are SD3 (Esser et al., 2024) modules.

Next, to analyze the bias in cultural ad generation, we prompt T2I models with country-level labels and corresponding action-reason statements to generate advertisements explicitly targeting each specified country. We use this result as a baseline and results suggest that these models often struggle to produce coherent or culturally appropriate content for underrepresented cultures (e.g., Africa).

To address the cultural bias in T2I models, we propose the Culture-aware Generator (CulGen, Fig. 2). As represented in Fig. 2 - Step 1, in the early steps we only condition the denoising process on the Prompt Embedding generating by SD3 Text Encoder given the prompt. In middle time-steps (Fig. 2 - Step 2), we first utilize the country predictions. Given the country, we randomly choose three images from the database targeting the same country and we collect the Cultural Components from those three images (using InternVL), and one random image out of the three to use as a visual reference. We use the SD3 Text Encoder, to encode the Cultural Components retrieved from database, and combine it with Prompt Embedding.

Topic	Real	Flux	Dalle3	AuraFlow	Llama3
Beauty	34.62	33.33	**58.46**	48.57	39.29
Cars	50.00	**100.00**	74.55	85.71	70.00
Clothing	41.51	38.00	63.25	**65.52**	51.52
Media/arts	76.92	0.00	60.00	**100.00**	71.43
Shopping	50.00	**80.00**	60.00	**80.00**	77.27
Soda	61.54	66.67	27.27	**85.71**	56.10
Dom. viol.	75.00	66.67	0.00	**85.71**	50.00
Human rights	71.88	**92.11**	0.00	87.50	64.84
Self-esteem	62.86	27.27	**100.00**	64.71	57.58
Smoking	73.33	55.56	0.00	**100.00**	64.71
Overall	64.03	59.10	74.98	**84.46**	62.20

Table 2: Diversity of gender on top 10 most common topics (% depictions of men). Top value per row bolded.

We use the combination of Cultural Component Embedding and Prompt Embedding as the denoising condition in middle time-steps (Fig. 2 - Step 2). Next, we encode the retrieved image using CLIP Image Encoder, and a Linear Layer, and generate the Image Embedding. We also use SD3 encoder to generate the Reason Embedding, given the reason part of action-reason statement. Then we project the Cultural Component Embedding on Reason Embedding using CA1 in Fig. 2 to generate the Projected Cultural Component Embedding. Then, using the CA2, we project the Image Embedding on Projected Cultural Component Embedding to generate the Projected Image. Finally, we combine the Projected Image, Prompt Embedding, and Cultural Component Embedding to create the conditioning vector for denoising in later time-steps (Fig. 2 - Step 3). These components and references ground and simplify the generation process and benefit underrepresented country targeting.

4 Results

We discuss the results based on the experimental setup in Sec. A (appx).

4.1 Diversity in real/generated ads

In Tab. 1, we see T2I models reduce race bias towards white-portrayed individuals and improve diversity. The biggest representation of whites is

	GPT4o (w/ & wo/ vision)						QwenVL (top) / QwenLM (bottom)						InternVL (top) / InternLM (bottom)					
	A	B	I	L	M	W	A	B	I	L	M	W	A	B	I	L	M	W
MLLM	13.96	13.56	14.61	15.72	14.78	**27.37**	13.30	15.76	16.39	16.44	16.09	**22.02**	15.91	16.98	15.81	16.19	16.65	**18.46**
LLM	15.63	**16.37**	14.02	11.52	13.57	11.37	**14.02**	11.52	13.57	11.37	10.95	8.47	12.38	14.03	12.88	**15.35**	10.95	8.90

Table 3: Race distribution of persuasion winners (in %). The model name for each group of columns is the judge.

MLLM	GPT4o		QwenVL		InternVL	
	man	woman	man	woman	man	woman
Clothing	28.97	**59.31**	**59.31**	40.69	45.52	**54.48**
Cars	31.95	**56.02**	**63.16**	36.84	31.95	**66.92**
Sports equip.	**45.83**	41.67	**79.17**	20.83	45.83	**54.17**
Shopping	50.00	50.00	**75.00**	25.00	16.67	**83.33**
Overall	33.02	**55.19**	**59.77**	40.23	42.56	**57.21**
LLM	**GPT4o**		**Qwen**		**InternLM**	
	man	woman	man	woman	man	woman
Clothing	**53.70**	46.30	**55.10**	44.90	39.62	**60.38**
Cars	44.12	**55.88**	**52.38**	47.62	43.18	**56.82**
Sports equipment	**60.00**	40.00	50.00	50.00	50.00	50.00
Shopping	40.00	**60.00**	33.33	**66.67**	**75.00**	25.00
Overall	50.67	49.33	50.00	50.00	46.26	**53.74**

Table 4: Gender distribution of persuasion winner.

generally in the Real ads group, and smaller in others. Llama3 depicts the most Asians and Blacks across models, Flux the most Latinx, and Dalle3 the most Middle-Eastern. Topical biases persist: Blacks are generally more common in social topics (human rights, self-esteem) than commercial topics (clothing, shopping), e.g., in Real, Flux, Auraflow.

In Tab. 2, we show the percent of men (out of all) in the 10 most common topics: 6 from products and 4 from public service announcements. Ideally, this number would be 50, indicating a balanced representation. We bold the biggest numbers; most greatly exceed 50, indicating over-representation of men. Overall, two methods show fewer men than real ads (59.10 for Flux and 62.20 for Llama3 vs 64.03 for Real), but two greatly increase men's over-representation (74.98 for Dalle3, 84.26 for AuraFlow). The only categories with fewer men are Beauty and Clothing.

4.2 Challenges with diversification

Tab. 3 shows the distribution of winners when asking which of two images that are identical except for race, is more persuasive. Judgments are made by MLLMs or LLMs after the image description. Given an unbiased model, this choice should be random and balanced. However, images with whites win across all MLLM judges. The gap in portions of white vs other races is bigger in GPT4o and QwenVL than in InternVL judgments. Interestingly, LLMs seem less biased towards Whites than MLLMs, with Blacks, Asians, and Latinx having the biggest portion of winners for one judge. We surmise this is due to efforts to reduce LLM bias which have not caught on in MLLMs yet.

Tab. 4 shows winner distribution when swapping genders. Different judges have different biases, with GPT4o and InternVL biased towards preferring women as more persuasive characters (except men in sports equipment for GPT4o), and QwenVL preferring men. Compared to Tab. 2 on the topic 'Cars', men are overrepresented in generated ads (by 4 models) but women are more persuasive (for 2 judges). This may be a good sign for diversifying ads or may indicate bias (women are seen as more attractive and appealing).

We further analyzed the reasoning behind gender and race selections, revealing underlying biases. We show examples in Fig. 3. The qualitative analysis on models' assumptions bias shows that women were often chosen for qualities like elegance, while men were selected for strength and reliability (QwenLM). In car ads, men were associated with sophistication and goal orientation, whereas women were linked to expanding suitability and diversity (InternLM). For skincare and jewelry, women were selected based on assumptions about the target audience, while selecting men was justified as promoting diversity (GPT4o). This suggests that personalization as a persuasion technique can introduce bias as MLLMs often assume stereotypical target audiences.

4.3 Targeting countries

First, to evaluate the cultural bias in advertisement data, we present the distribution of ad origins in PittAds, predicted by InternVL. Among 13,172 analyzed images, 101 countries were identified. 10,335 images (0.78%) were classified as targeting the US, UK, Canada, or Australia, while 227 were labeled as universal advertisements. The remaining 2,620 images were associated with 88 other countries. This indicates over-representation of the Western culture in the dataset. Fig. 4 shows the distribution of advertisement images over the countries.

Our qualitative analysis on generating cultural advertisement represented in Fig. 5 and 7 (appx), shows that existing T2I models struggle in gener-

Figure 3: Example on different reasoning for choosing more persuasive images.

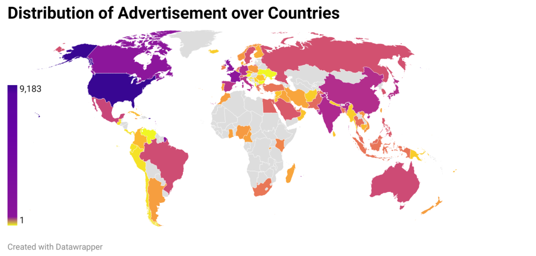


Figure 4: Distribution of advertisement images in PittAd dataset over different countries.

Figure 5: Examples of cultural image generation. Action-reason prompts: (a) I should drink this beer because it is as light as feather. (b) I should use this deodorant because it is as fresh as mint.

T2I model	VQA-score Average	AR	Country
Baselines			
Flux	0.54	0.78	0.31
SD3	0.70	0.78	0.63
PixArt	0.54	0.67	0.42
AuraFlow	0.70	0.76	0.66
Ablations			
No cultural components	0.72	**0.80**	0.64
Cultural components in early steps	0.67	0.52	**0.83**
Cultural components in later steps	0.74	0.79	0.69
No style image	0.73	0.68	0.78
Multiple style images	0.74	0.68	0.79
Ours			
CulGen	**0.75**	0.69	0.81

Table 5: Cultural targeting evaluation. Flux, SD3, PixArt(Chen et al., 2024a), and AuraFlow use the country name in the prompt.

ating diverse cultural advertisement showing all cultures similar. Fig. 5 shows our method better reflects the respective culture, e.g., crescent/religion (left), palms and city towers (right) for UAE, dragons and red-yellow color theme (right) for China, and French text and Eiffel tower (right) for France.

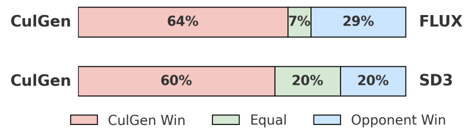


Figure 6: Images chosen by human annotator as better showing the target country.

Quantitatively, Tab. 5 evaluates CulGen, using VQA-score (Lin et al., 2024) between generated images and AR or country name. Our method better targets the country and reflects the AR well, resulting in a higher AR-country average than four strong baselines. We also did an ablation on different design choices to show the effectiveness of each design and discuss the results in the appendix, Sec. A.3.2.

We further analyzed the quality of generated images, conducting human evaluation on 25 prompts for 4 different countries: France, China, UAE, and India (details in Sec. A.3.2). Fig. 6 represents the results of our human evaluation of the method highlighting the significant improvement of our proposed method compared to the baseline models when targeting specific countries and conveying the advertisement message.

5 Conclusion

We analyzed racial and gender representation biases in real and T2I-generated advertisements. We showed perception biases of persuasivensess by MLLM and LLM judges in controlled experiments with nearly identical images. We showed promise of country targeting through cultural symbols.

6 Limitations

In our analysis of real ads, we are limited by the ads included in PittAds, which are Western-centric and crawled from the web, so not reflecting ads in print media nor on TV/streaming platforms. In our analysis of demographics, we used DeepFace which is imperfect but we observed high accuracy. We also simplify racial/ethnic backgrounds to a fixed and small set of categories; these could be more numerous and non-overlapping. We simplify genders to only two, but note that GPT4o also outputs a significant number of non-binary classifications. In our analysis of how persuasion varies when elements of the ad are swapped, we only focused on gender and race. An exploration of how persuasion varies when symbols from different cultures are used might also be meaningful, but the data we use preclude us from doing so, because of entanglement of these symbols inside the action-reason statements. Finally, our cultural targeting is promising, but it is important to not over-exaggerate cultural symbolism, and to avoid stereotypization. To know the right level of targeting, we plan to work with members of the countries targeted to learn what is desirable and undesirable use of cultural symbols.

Acknowledgment. This work was partly supported by NSF Grant No. 2046853 and partly by the University of Pittsburgh Center for Research Computing and Data, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681. We gratefully acknowledge the support of our annotators.

References

AI@Meta. 2024. Llama 3 model card.

Abdulkareem Alsudais. 2025. Analyzing how text-to-image models represent nationalities in everyday tasks. *arXiv preprint arXiv:2504.06313*.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, and 1 others. 2025. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*.

James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, and 1 others. 2023. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8.

Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 1493–1504.

Black Forest Labs. 2024. Flux. https://github.com/black-forest-labs/flux.

Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, and 1 others. 2024. Internlm2 technical report. *arXiv preprint arXiv:2403.17297*.

Junsong Chen, Jincheng Yu, Chongjian Ge, Lewei Yao, Enze Xie, Yue Wu, Zhongdao Wang, James Kwok, Ping Luo, Huchuan Lu, and 1 others. 2024a. Pixart-α: Fast training of diffusion transformer for photorealistic text-to-image synthesis. In *ICLR*.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024b. 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.

Jaemin Cho, Abhay Zala, and Mohit Bansal. 2023. Dall-eval: Probing the reasoning skills and social biases of text-to-image generation models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3043–3054.

Emily Dinan, Angela Fan, Adina Williams, Jack Urbanek, Douwe Kiela, and Jason Weston. 2020. Queens are powerful too: Mitigating gender bias in dialogue generation. In *EMNLP*.

Moreno D'Incà, Elia Peruzzo, Massimiliano Mancini, Dejia Xu, Vidit Goel, Xingqian Xu, Zhangyang Wang, Humphrey Shi, and Nicu Sebe. 2024. Openbias: Open-set bias detection in text-to-image generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12225–12235.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, and 1 others. 2024. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*.

Fal. 2024. Introducing auraflow v0.1, an open exploration of large rectified flow models. Available at: https://blog.fal.ai/auraflow/ [Accessed: 2024-11-07].

Kathleen C Fraser and Svetlana Kiritchenko. 2024. Examining gender and racial bias in large vision–language models using a novel dataset of parallel images. In *Proceedings of the 18th Conference of the*

European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 690–713.

Yusuke Hirota, Yuta Nakashima, and Noa Garcia. 2022. Gender and racial bias in visual question answering datasets. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1280–1292.

Binyuan Hui, Jian Yang, Zeyu Cui, Jiaxi Yang, Dayiheng Liu, Lei Zhang, Tianyu Liu, Jiajun Zhang, Bowen Yu, Keming Lu, and 1 others. 2024. Qwen2. 5-coder technical report. *arXiv preprint arXiv:2409.12186*.

Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, and Adriana Kovashka. 2017. Automatic understanding of image and video advertisements. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1705–1715.

Ben Hutchinson, Jason Baldridge, and Vinodkumar Prabhakaran. 2022. Underspecification in scene description-to-depiction tasks. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1172–1184.

Sepehr Janghorbani and Gerard De Melo. 2023. Multimodal bias: Introducing a framework for stereotypical bias assessment beyond gender and race in vision–language models. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1725–1735.

Akshita Jha, Vinodkumar Prabhakaran, Remi Denton, Sarah Laszlo, Shachi Dave, Rida Qadri, Chandan Reddy, and Sunipa Dev. 2024. Visage: A global-scale analysis of visual stereotypes in text-to-image generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12333–12347.

Simran Khanuja, Sathyanarayanan Ramamoorthy, Yueqi Song, and Graham Neubig. 2024. An image speaks a thousand words, but can everyone listen? on image transcreation for cultural relevance. In *EMNLP*.

Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2021. Towards understanding and mitigating social biases in language models. In *International conference on machine learning*, pages 6565–6576. PMLR.

Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. 2024. Evaluating text-to-visual generation with image-to-text generation. In *ECCV*.

Joel Mire, Zubin Trivadi Aysola, Daniel Chechelnitsky, Nicholas Deas, Chrysoula Zerva, and Maarten Sap. 2025. Rejected dialects: Biases against african american language in reward models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 7468–7487.

Anjishnu Mukherjee, Ziwei Zhu, and Antonios Anastasopoulos. 2025. Crossroads of continents: Automated artifact extraction for cultural adaptation with large multimodal models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1755–1764. IEEE.

Ranjita Naik and Besmira Nushi. 2023. Social biases through the text-to-image generation lens. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 786–808.

OpenAI. 2024. Gpt-4o. https://openai.com/gpt-4o. Large language model. Accessed May 19, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, and 1 others. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510.

Emily Sheng, Kai-Wei Chang, Prem Natarajan, and Nanyun Peng. 2021. Societal biases in language generation: Progress and challenges. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4275–4293.

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708.

Muhammad Usman. 2013. Creation of effective advertising in the persuasion of target audience. *International Journal*, 2(1):77–82.

Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, Kai-Wei Chang, and Nanyun Peng. 2023. "kelly is a warm person, joseph is a role model": Gender biases in LLM-generated reference letters. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Angelina Wang, Alexander Liu, Ryan Zhang, Anat Kleiman, Leslie Kim, Dora Zhao, Iroha Shirai, Arvind Narayanan, and Olga Russakovsky. 2022. Revise: A tool for measuring and mitigating bias in visual datasets. *International Journal of Computer Vision*, 130(7):1790–1810.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Jiayi Ye, Yanbo Wang, Yue Huang, Dongping Chen, Qihui Zhang, Nuno Moniz, Tian Gao, Werner Geyer, Chao Huang, Pin-Yu Chen, and 1 others. Justice or prejudice? quantifying biases in llm-as-a-judge. In *Neurips Safe Generative AI Workshop 2024*.

Dora Zhao, Angelina Wang, and Olga Russakovsky. 2021. Understanding and evaluating racial biases in image captioning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14830–14840.

A Implementation Detail

A.1 Analyzing Diversity in Real and Generated Ads

To evaluate diversity in real and generated advertisement images, we used the PittAds dataset for real ads and generated new images using pre-trained text-to-image (T2I) models. Specifically, we used the Huggingface-hosted models: *'sd-community/sdxl-flash'* for SDXL, *'fal/AuraFlow-v0.2'* for AuraFlow, and *'black-forest-labs/FLUX.1-dev'* for Flux. For AuraFlow, we also experimented with prompt expansion using LLAMA3-Instruct (*'meta-llama/Meta-Llama-3-8B-Instruct*) to provide a more detailed description of the ad content.

Listing 1: Prompt for Advertisement Image Generation {Blue} denotes elements added dynamically.

```
Generate an advertisement image that
conveys the following message:
    - {AR}
```

A.2 Evaluating Persuasion Bias via Demographic Swaps

For persuasion evaluation, we used the OpenAI API's *'GPT-4-1'* model to generate initial ad images and modify demographic attributes (e.g., race and gender) while keeping other content constant. We used *'GPT-4o'* to describe and edit the image captions accordingly. To evaluate persuasiveness, we employed several MLLMs as judges: *'OpenGVLab/InternVL2_5-78B'* (InternVL2.5), *'Qwen/Qwen2.5-VL-7B-Instruct'* (QwenVL2.5), and OpenAI's GPT-4o. Corresponding LLMs were also used as text-only judges: *'Qwen/Qwen2.5-7B-Instruct'* (Qwen2.5), *'internlm/internlm2_5-7b-chat'* (InternLM2.5), and the text-only version of GPT-4o. These models were prompted with pairs of images/descriptions inputs to select the more persuasive option, using chain-of-thought (CoT) reasoning. For each two images/descriptions we use both orders to remove the possible bias for specific indices in the models.

Listing 2: Prompt for editing the image to change the race of the main character. {Blue} denotes elements added dynamically.

```
replace the white person in the image
with a {race}person.
```

Listing 3: Prompt for editing the description to change the race of the main character. {Blue} denotes elements added dynamically.

```
In the following description, replace
the 'white person' with '{race} person',
changing the race characteristic of the
person. Only return the new description
without any further explanation.
{description}
```

Listing 4: Prompt for LLM persuasion judgment. {Blue} denotes elements added dynamically.

```
Context: You are a helpful assistant
who chooses the more persuasive image.
Persuasion of the advertisement image
is the ability of the image to convince
the audience to take an action.

Question: Given the following two
advertisement descriptions, which
advertisement is more persuasive?
1. description 1. 2. description 2
First, explain the persuasion in each
description, and then return the more
persuasive description in the format of:
Explanation : ${Explanation}
Answer: ${index of more persuasive
description}

Discription 1: {Description1}

Description 2: {Description2}
```

Listing 5: Prompt for MLLM persuasion judgment.

```
Context: You are a helpful assistant
who ranks these images in terms of
persuasion. Persuasion of the
advertisement image is the ability of
the image to convince the audience to
take an action.

Question: Which image is more
persuasive?
1. Image 1 2. Image 2.
First , explain the persuasion in each
image and then answer. Your answer
format is:
Explanation: ${explanation on
persuasion of the images}
Answer: ${index of correct option}
```

A.3 Diversifying through country targeting

CulGen. To generate culturally-aware advertisement images, we build on the Huggingface implementation of SD3, using the pre-trained model *'stabilityai/stable-diffusion-3-medium-diffusers'*. As shown in Fig.2, we retain the original noise scheduler and diffusion module from SD3, and introduce a new component called the condition scheduler (light orange box in Fig.2). We first construct a database containing countries, corresponding advertisement images, and extracted cultural components for each country, based on our country prediction pipeline. Additionally, we prompt GPT-4o to map each country to a representative visual element, which we include as an extra cultural component. Given a target country, we randomly retrieve three relevant images from the database. We aggregate the cultural components from these images and randomly select one image to serve as a visual reference. Given the input prompt (light green border box), we encode it using the SD3 text encoder. During the early denoising time-steps, we condition the model only on the prompt embedding to ensure it follows the textual intent. In the middle time-steps, we generate embeddings for the cultural components and condition the model on concatenation of prompt embedding and cultural components embeddings. The embedding of the "reason" part of the AR (action-reason) is generated the same way. We also encode the reference image using a CLIP (Radford et al., 2021) image encoder and project it into the text embedding space using a linear layer. We then apply a cross-attention layer to project the cultural component embeddings onto the reason embedding. The output of this layer serves as the query in another cross-attention mechanism between the cultural components and the projected image embedding. Finally, we concatenate the resulting image embedding with the cultural, reason, and prompt embeddings to form the full conditioning vector for the late denoising steps. During training, we keep all SD3 modules and the CLIP encoder frozen, and only train the condition scheduler using the DreamBooth method (Ruiz et al., 2023) on 250 images, with learning-rate 1e-5, batch-size 1, 4 gradient accumulation steps, for 500 steps.

A.3.1 Evaluation

To evaluate our proposed method, we compare our method against SD3 (Huggingface *'stabilityai/stable-diffusion-3-medium-diffusers'*), Flux (Hugginface *'black-forest-labs/FLUX.1-dev'*), PixArt (Huggingface: *'PixArt-alpha/PixArt-XL-2-1024-MS'*), and AuraFlow (Huggingface: *'fal/AuraFlow-v0.2'*). We set the seed equal to 0 for image generation.

For action-reason statements, we prompted GPT4o with one example statement, to generate 100 advertisement statements following structure of *'I should drink this beer because it is as light as a feather.'* We chose 5 countries across different cultures as China (East-Asian Culture), France (Western Culture), South Africa (African Culture), United Arab Emirates (Middle-eastern Culture), and Mexico (Latin Culture). And prompted the model to generate advertisement images targeting each of these countries, resulting in 500 test samples.

For our evaluation metrics, we used VQA-score (Lin et al., 2024), one of the most accurate test-image alignment scores. We first computed the alignment score between each image and action-reason statement. Next, we computed the alignment score between the country name and the corresponding image. Finally we computed the average of these two values as the score for each model. We report all three scores in Table 5.

A.3.2 Design Choice Ablation

We first removed the cultural components from denoising condition resulting in higher alignment with the action-reason statements and lower alignment with the country, showing the effectiveness of the cultural components on country representation. We evaluated the effectiveness of the time-step design by conditioning on cultural components starting in early time-steps and later time-steps keeping the rest of design as it is. Starting conditioning on cultural components in early steps results in high country representation but very low alignment with action-reason statement, showing that the use of cultural components from start to end results in focusing on culture representation and ignoring the advertisement message. On the other hand, use of cultural components in conditioning in later time-steps ("Cultural components in later steps") shows higher alignment with action-reason statement and lower representation of the country, i.e. aligning well and ignoring the target country. This shows the effectiveness of different conditioning in different time-steps, as we do in CulGen, thus confirming the advantages of our proposed design. We further analyze the effectiveness of the different design modules by removing the style image from the conditions in the later time-steps ("No style image"). This represents both lower representation of the country and lower alignment with action-reason statements. We next increased the number of style images (Multiple style images) using all three examples instead of randomly choosing one. The result is slightly lower but comparable to that achieved by our proposed design (CulGen). This may be because using multiple images makes the condition more complex.

Listing 6: Prompt for Cultural Image Generation {Blue} denotes elements added dynamically.

```
Generate an advertisement image that
targets people from {country} conveying
the following message:
    - {AR}
```

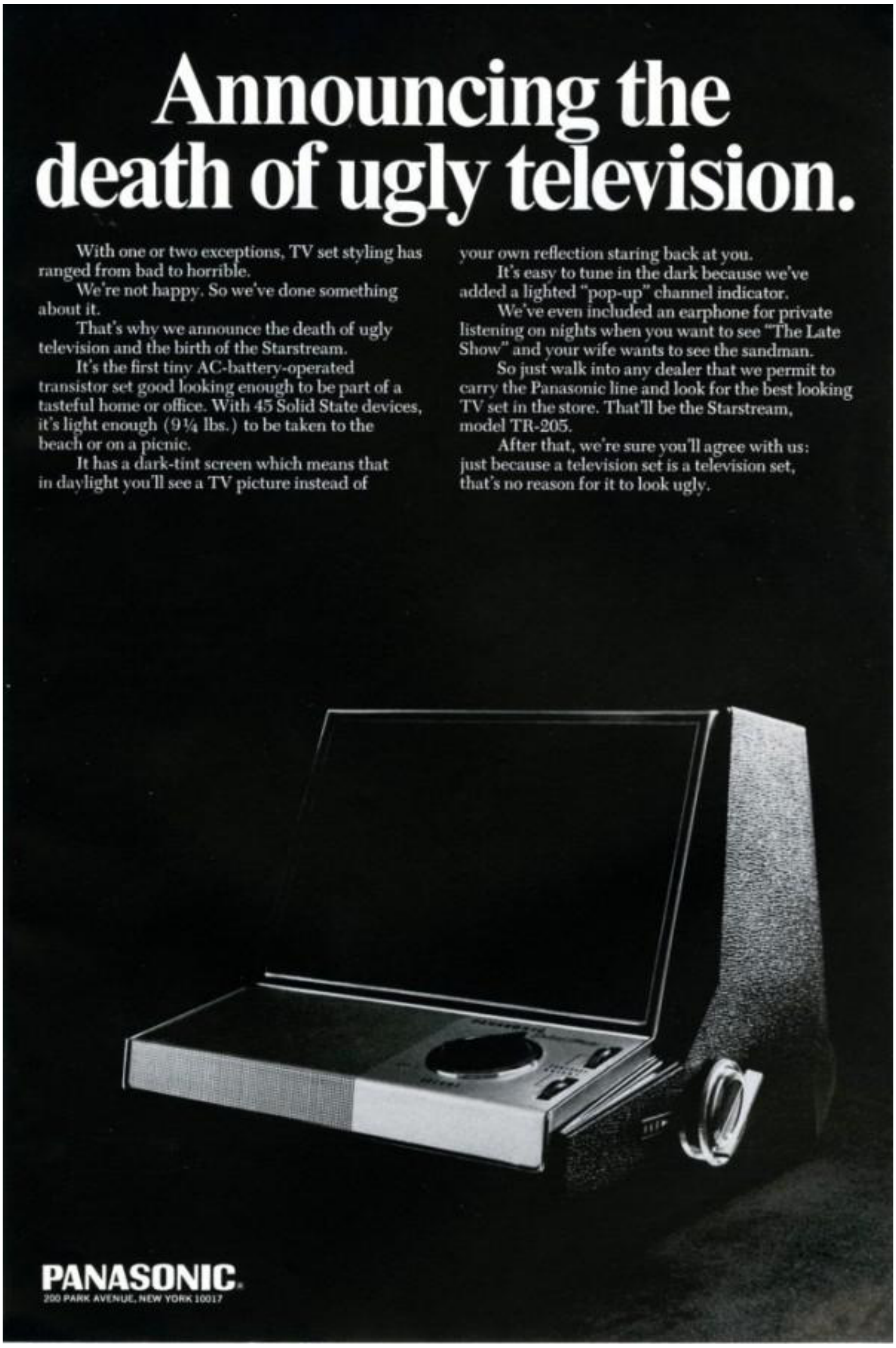
Human Evaluation on Cultural Image Generation: The annotators were from China (1 annotator), Middle East (2 annotators, 1 familiar with European culture), and India (1 annotator) to ensure the familiarity with the culture of each country. Each annotator was presented with two generated images targeting the country they were familiar with ordered randomly, and the corresponding action-reason statement. The annotators were asked to choose the image that better targets the people from the defined country while aligning with the message.

B Qualitative Examples

Figure 7: Examples of images generated by CulGen (ours), SD3, Flux, PixArt, and AuraFlow models targeting China, South Africa, France, United Arab Emirates, and Mexico.

Form

1. What is the correct target country for this image? Choose all that applies.



Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

2. What is the correct target country for this image? Choose all that applies.

Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran

- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

3. What is the correct target country for this image? Choose all that applies.

Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

4. What is the correct target country for this image? Choose all that applies.

Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia

- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

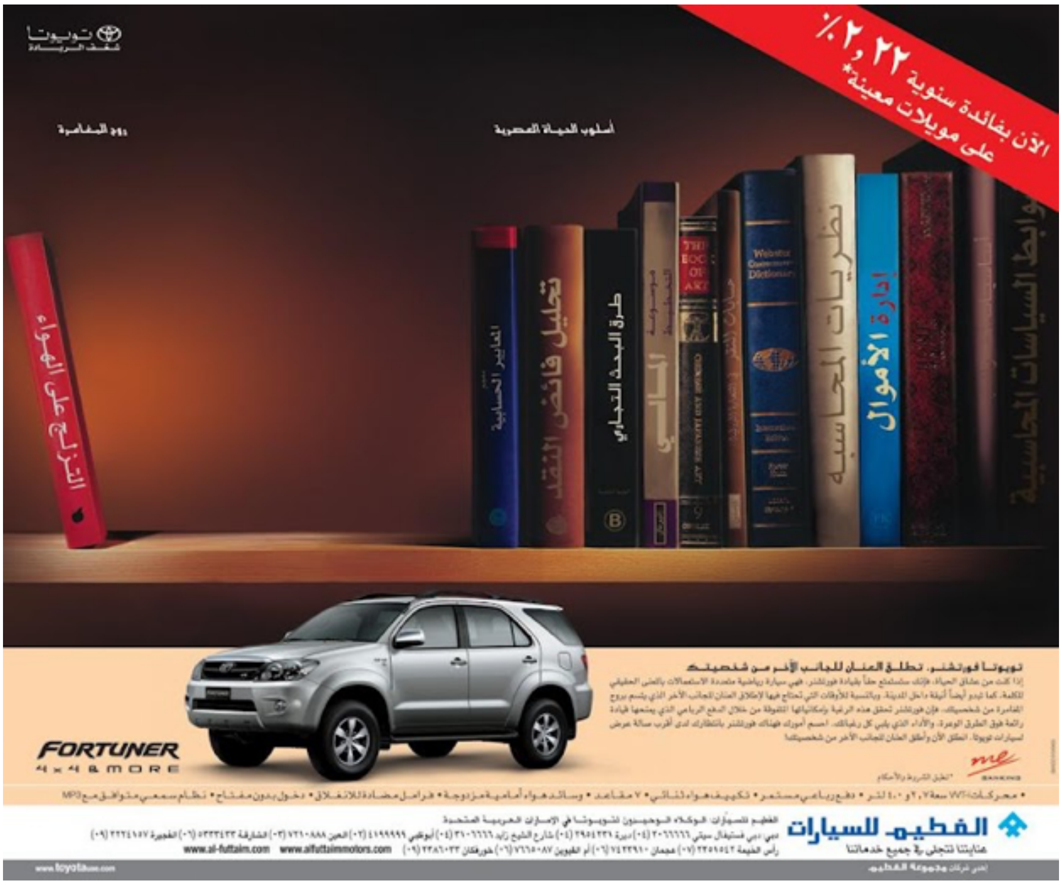
5. What is the correct target country for this image? Choose all that applies.

Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia

- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

6. What is the correct target country for this image? Choose all that applies.



Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark
- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

7. What is the correct target country for this image? Choose all that applies.

Check all that apply.

- [ ] Argentina
- [ ] Australia
- [ ] Bangladesh
- [ ] Belgium
- [ ] Brazil
- [ ] Canada
- [ ] China
- [ ] Denmark

- [ ] Egypt
- [ ] France
- [ ] Greece
- [ ] Guatemala
- [ ] India
- [ ] Indonesia
- [ ] Iran
- [ ] Italy
- [ ] Japan
- [ ] Mexico
- [ ] Myanmar
- [ ] Nepal
- [ ] Norway
- [ ] Oman
- [ ] Pakistan
- [ ] Peru
- [ ] Puerto Rico
- [ ] Qatar
- [ ] Saudi Arabia
- [ ] Singapore
- [ ] South Korea
- [ ] Spain
- [ ] Sri Lanka
- [ ] Sweden
- [ ] Switzerland
- [ ] Taiwan
- [ ] Thailand
- [ ] Turkey
- [ ] United Arab Emirates
- [ ] United Kingdom
- [ ] United States

Topic-Gender

	real		LLAMA3 instruct		FLUX		DALLE3		auraflow	
	man	woman	man	woman	man	woman	man	woman	man	woman
Alcohol	100.00%	0.00%	100.00%	0.00%					100.00%	0.00%
Animal rights	36.36%	63.64%	70.00%	30.00%	33.33%	66.67%	0.00%	100.00%	62.50%	37.50%
Baby products					100.00%	0.00%	0.00%	100.00%	66.67%	33.33%
Beauty product	34.62%	65.38%	39.29%	60.71%	33.33%	66.67%	58.46%	41.54%	48.57%	51.43%
Cars	50.00%	50.00%	70.00%	30.00%	100.00%	0.00%	74.55%	25.45%	85.71%	14.29%
Charities	75.00%	25.00%	90.00%	10.00%					100.00%	0.00%
Chips	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%	100.00%	0.00%
Chocolate	100.00%	0.00%	75.00%	25.00%						
Cleaning product	75.00%	25.00%	53.85%	46.15%			100.00%	0.00%		
Clothing and ac	41.51%	58.49%	51.52%	48.48%	38.00%	62.00%	63.25%	36.75%	65.52%	34.48%
Coffee	0.00%	100.00%	0.00%	100.00%			100.00%	0.00%	100.00%	0.00%
Domestic violen	75.00%	25.00%	50.00%	50.00%	66.67%	33.33%			85.71%	14.29%
Education	100.00%	0.00%								
Electronics	100.00%	0.00%	67.44%	32.56%	84.62%	15.38%	78.57%	21.43%	83.33%	16.67%
Environment	0.00%	0.00%			100.00%	0.00%				
Financial service	100.00%	0.00%	66.67%	33.33%	100.00%	0.00%			100.00%	0.00%
Games and toys	0.00%	0.00%	75.00%	25.00%	0.00%	100.00%				
Healthcare and n	57.14%	42.86%	55.56%	44.44%	33.33%	66.67%	100.00%	0.00%	100.00%	0.00%
Home appliances	100.00%	0.00%	71.43%	28.57%			100.00%	0.00%	100.00%	0.00%
Home improveme	100.00%	0.00%					100.00%	0.00%		
Human rights	71.88%	28.13%	64.84%	35.16%	92.11%	7.89%			87.50%	12.50%
Media and arts	76.92%	23.08%	71.43%	28.57%	0.00%	100.00%	60.00%	40.00%	100.00%	0.00%
Pet food			0.00%	100.00%						
Phone			62.50%	37.50%	50.00%	50.00%	87.50%	12.50%	100.00%	0.00%
Political candidat	80.00%	20.00%	73.08%	26.92%	33.33%	66.67%			80.00%	20.00%
Restaurants	100.00%	0.00%	80.00%	20.00%	100.00%	0.00%				
Security and safe	100.00%	0.00%			0.00%	0.00%				
Self esteem	62.86%	37.14%	57.58%	42.42%	27.27%	72.73%	100.00%	0.00%	64.71%	35.29%
Shopping	50.00%	50.00%	77.27%	22.73%	80.00%	20.00%	60.00%	40.00%	80.00%	20.00%
Smoking	73.33%	26.67%	64.71%	35.29%	55.56%	44.44%			100.00%	0.00%
Soda	61.54%	38.46%	56.10%	43.90%	66.67%	33.33%	27.27%	72.73%	85.71%	14.29%
Software	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%	100.00%	0.00%	0.00%	100.00%
Sports equipmen	90.00%	10.00%	73.33%	26.67%	83.33%	16.67%	90.00%	10.00%	100.00%	0.00%
Unclear	80.00%	20.00%	100.00%	0.00%	100.00%	0.00%			100.00%	0.00%
Unknown	0.00%	0.00%	50.00%	50.00%			100.00%	0.00%		
Vacation and trav	0.00%	100.00%	66.67%	33.33%						
dating	85.71%	12.50%	57.14%	42.86%					100.00%	0.00%
SUM										
AVG										
WEIGHTED AVG										
AVG TOP 10										
AVG TOP 5										
avg per gender	64.03%	24.16%	62.20%	37.80%	59.10%	36.90%	74.98%	25.02%	84.46%	15.54%

	LLAMA3 instruct		FLUX		DALLE3		AuraFlow	
Alcohol	0.00%	0	-100.00%	0	-100.00%	0	0.00%	0
Animal rights	33.64%	3.363636364	-3.03%	-0.1818181818	-36.36%	-0.3636363636	26.14%	2.090909091
Baby products	0.00%	0	100.00%	1	0.00%	0	66.67%	2
Beauty product	4.67%	2.615384615	-1.28%	-0.1923076923	23.85%	15.5	13.96%	4.884615385
Cars	20.00%	10	50.00%	4.5	24.55%	13.5	35.71%	2.5
Charities	15.00%	1.5	-75.00%	0	-75.00%	0	25.00%	0.25
Chips	100.00%	2	0.00%	0	100.00%	3	100.00%	2
Chocolate	-25.00%	-1	-100.00%	0	-100.00%	0	-100.00%	0
Cleaning product	-21.15%	-2.75	-75.00%	0	25.00%	0.25	-75.00%	0
Clothing and ac	10.01%	13.20754717	-3.51%	-1.754716981	21.74%	25.43396226	24.01%	13.9245283
Coffee	0.00%	0	0.00%	0	100.00%	1	100.00%	1
Domestic violen	-25.00%	-1.5	-8.33%	-0.25	-75.00%	0	10.71%	0.75
Education	-100.00%	0	-100.00%	0	-100.00%	0	-100.00%	0
Electronics	-32.56%	-14	-15.38%	-2	-21.43%	-9	-16.67%	-1
Environment	0.00%	0	100.00%	2	0.00%	0	0.00%	0
Financial service	-33.33%	-1	0.00%	0	-100.00%	0	0.00%	0
Games and toys	75.00%	3	0.00%	0	0.00%	0	0.00%	0
Healthcare and n	-1.59%	-0.1428571429	-23.81%	-0.7142857143	42.86%	0.8571428571	42.86%	0.8571428571
Home appliances	-28.57%	-2	-100.00%	0	0.00%	0	0.00%	0
Home improveme	-100.00%	0	-100.00%	0	0.00%	0	-100.00%	0
Human rights	-7.04%	-6.40625	20.23%	7.6875	-71.88%	0	15.63%	1.25
Media and arts	-5.49%	-0.3846153846	-76.92%	-1.538461538	-16.92%	-0.8461538462	23.08%	0.9230769231
Pet food	0.00%	0	0.00%	0	0.00%	0	0.00%	0
Phone	62.50%	15	50.00%	1	87.50%	7	100.00%	6
Political candidat	-6.92%	-1.8	-46.67%	-1.4	-80.00%	0	0.00%	0
Restaurants	-20.00%	-2	0.00%	0	-100.00%	0	-100.00%	0
Security and safe	-100.00%	0	-100.00%	0	-100.00%	0	-100.00%	0
Self esteem	-5.28%	-5.228571429	-35.58%	-3.914285714	37.14%	1.485714286	1.85%	0.3142857143
Shopping	27.27%	12	30.00%	3	10.00%	2.5	30.00%	4.5
Smoking	-8.63%	-2.933333333	-17.78%	-1.6	-73.33%	0	26.67%	1.6
Soda	-5.44%	-2.230769231	5.13%	0.1538461538	-34.27%	-7.538461538	24.18%	3.384615385
Software	-100.00%	-14	0.00%	0	0.00%	0	-100.00%	-1
Sports equipmen	-16.67%	-2.5	-6.67%	-0.4	0.00%	0	10.00%	1.1
Unclear	20.00%	0.6	20.00%	0.2	-80.00%	0	20.00%	0.4
Unknown	50.00%	1	0.00%	0	100.00%	2	0.00%	0
Vacation and trav	66.67%	6	0.00%	0	0.00%	0	0.00%	0
dating	-28.57%	-2	-85.71%	0	-85.71%	0	14.29%	0.5714285714
SUM		8.410171629		5.595470332		54.77856766		48.30060223
AVG	-5.04%		-18.90%		-18.30%		0.52%	
WEIGHTED AVG		1.07%		2.87%		14.81%		20.91%
AVG TOP 10	0.51%		-3.81%		-15.41%		20.58%	
AVG TOP 5	5.93%		1.97%		4.17%		17.09%	
avg per gender								
	llama3 least biased		flux slightly biased towards men		dalle3 slightly more biased towards		auraflow most biased to include	

Topic - Race

	real						LLAMA3 instruct				
	white	latino hispanic	asian	black	middle eastern	indian	latino hispanic	asian	black	middle eastern	indian
Alcohol	100%	0%	0%	0%	0%	0%	0%	17%	17%	0%	0%
Animal rights	64%	0%	27%	9%	0%	0%	10%	0%	20%	10%	0%
Baby products											
Beauty product	**58%**	**8%**	**15%**	**8%**	**8%**	**4%**	**7%**	**32%**	**29%**	**0%**	**2%**
Cars	72%	0%	17%	6%	6%	0%	18%	32%	8%	4%	0%
Charities	0%	0%	0%	100%	0%	0%	10%	30%	20%	0%	0%
Chips							0%	50%	0%	0%	0%
Chocolate	0%	0%	67%	33%	0%	0%	0%	25%	25%	0%	0%
Cleaning product	75%	25%	0%	0%	0%	0%	8%	38%	23%	0%	0%
Clothing and ac	**66%**	**9%**	**15%**	**6%**	**4%**	**0%**	**11%**	**27%**	**32%**	**1%**	**5%**
Coffee	100%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
Domestic violenc	75%	8%	8%	8%	0%	0%	0%	17%	33%	0%	0%
Education	100%	0%	0%	0%	0%	0%					
Electronics	100%	0%	0%	0%	0%	0%	9%	37%	16%	0%	0%
Environment											
Financial service	100%	0%	0%	0%	0%	0%	0%	33%	0%	33%	0%
Games and toys							0%	50%	25%	0%	0%
Healthcare and	71%	0%	29%	0%	0%	0%	22%	22%	0%	0%	0%
Home appliances	100%	0%	0%	0%	0%	0%	14%	43%	0%	0%	0%
Home improveme	100%	0%	0%	0%	0%	0%					
Human rights	**66%**	**9%**	**6%**	**9%**	**0%**	**9%**	**14%**	**22%**	**12%**	**9%**	**2%**
Media and arts	38%	23%	15%	8%	15%	0%	14%	29%	14%	0%	0%
Pet food							0%	0%	0%	0%	0%
Phone							0%	33%	4%	4%	0%
Political candidat	100%	0%	0%	0%	0%	0%	23%	15%	15%	8%	0%
Restaurants	100%	0%	0%	0%	0%	0%	0%	40%	0%	10%	0%
Security and safe	100%	0%	0%	0%	0%	0%					
Self esteem	**77%**	**3%**	**14%**	**6%**	**0%**	**0%**	**4%**	**40%**	**28%**	**3%**	**4%**
Shopping	**92%**	**0%**	**8%**	**0%**	**0%**	**0%**	**2%**	**32%**	**18%**	**2%**	**0%**
Smoking	87%	7%	0%	0%	7%	0%	12%	21%	3%	0%	3%
Soda	85%	0%	0%	0%	8%	8%	22%	12%	22%	5%	5%
Software	0%	0%	0%	100%	0%	0%	29%	50%	0%	0%	21%
Sports equipmen	90%	0%	0%	0%	10%	0%	7%	33%	13%	0%	0%
Unclear	100%	0%	0%	0%	0%	0%	0%	33%	33%	0%	0%
Unknown							0%	0%	50%	0%	0%
Vacation and trav	0%	0%	0%	100%	0%	0%	11%	11%	11%	0%	0%
dating	86%	0%	14%	0%	0%	0%	0%	14%	14%	14%	0%
SUM											
AVG											
WEIGHTED AVG											
AVG TOP 10											
AVG TOP 5											
avg per race	73%	3%	8%	13%	2%	1%	8%	**26%**	**18%**	3%	1%

	FLUX						DALLE3					
	white	latino hispanic	asian	black	middle eastern	indian	white	latino hispanic	asian	black	middle eastern	indian
Alcohol												
Animal rights	67%	0%	17%	0%	17%	0%	100%	0%	0%	0%	0%	0%
Baby products	0%	0%	100%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Beauty product	**60%**	**20%**	**20%**	**0%**	**0%**	**0%**	**72%**	**5%**	**9%**	**6%**	**5%**	**3%**
Cars	33%	11%	33%	11%	11%	0%	62%	5%	16%	7%	7%	2%
Charities												
Chips	0%	100%	0%	0%	0%	0%	33%	0%	0%	0%	67%	0%
Chocolate												
Cleaning product							0%	0%	100%	0%	0%	0%
Clothing and ac	**70%**	**12%**	**4%**	**8%**	**4%**	**2%**	**64%**	**2%**	**14%**	**6%**	**14%**	**1%**
Coffee							0%	0%	0%	0%	0%	100%
Domestic violenc	67%	33%	0%	0%	0%	0%						
Education												
Electronics	77%	0%	8%	8%	8%	0%	52%	0%	26%	2%	19%	0%
Environment	50%	0%	0%	50%	0%	0%						
Financial service	67%	33%	0%	0%	0%	0%						
Games and toys	0%	0%	100%	0%	0%	0%						
Healthcare and	67%	0%	33%	0%	0%	0%	50%	0%	50%	0%	0%	0%
Home appliances							100%	0%	0%	0%	0%	0%
Home improveme							50%	0%	0%	0%	50%	0%
Human rights	**47%**	**5%**	**8%**	**13%**	**26%**	**0%**						
Media and arts	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Pet food												
Phone	50%	0%	50%	0%	0%	0%	63%	25%	0%	13%	0%	0%
Political candidat	100%	0%	0%	0%	0%	0%						
Restaurants	100%	0%	0%	0%	0%	0%						
Security and safe												
Self esteem	**64%**	**9%**	**0%**	**18%**	**0%**	**9%**	**25%**	**0%**	**75%**	**0%**	**0%**	**0%**
Shopping	**70%**	**20%**	**10%**	**0%**	**0%**	**0%**	**52%**	**16%**	**8%**	**4%**	**20%**	**0%**
Smoking	33%	11%	11%	22%	22%	0%						
Soda	100%	0%	0%	0%	0%	0%	50%	9%	27%	9%	5%	0%
Software	0%	0%	0%	100%	0%	0%	100%	0%	0%	0%	0%	0%
Sports equipmen	33%	17%	50%	0%	0%	0%	80%	0%	10%	0%	10%	0%
Unclear	100%	0%	0%	0%	0%	0%						
Unknown							50%	50%	0%	0%	0%	0%
Vacation and trav												
dating												
SUM												
AVG												
WEIGHTED AVG												
AVG TOP 10												
AVG TOP 5												
avg per race	56%	11%	**19%**	**10%**	4%	0%	60%	6%	**17%**	**2%**	10%	5%

	auraflow						LLAMA3		FLUX		DALLE3		auraflow	
	white	latino hispanic	asian	black	middle eastern	indian								
Alcohol	100%	0%	0%	0%	0%	0%	-33%	-200%	-100%	0%	-100%	0%	0%	0%
Animal rights	75%	13%	0%	13%	0%	0%	-4%	-36%	3%	18%	36%	36%	11%	91%
Baby products	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	200%	100%	300%
Beauty product	**60%**	**17%**	**17%**	**6%**	**0%**	**0%**	**-27%**	**-1530%**	**2%**	**35%**	**15%**	**950%**	**2%**	**81%**
Cars	86%	0%	14%	0%	0%	0%	-34%	-1711%	-39%	-350%	-10%	-572%	13%	94%
Charities	100%	0%	0%	0%	0%	0%	40%	400%	0%	0%	0%	0%	100%	100%
Chips	100%	0%	0%	0%	0%	0%	50%	100%	0%	0%	33%	100%	100%	200%
Chocolate							50%	200%	0%	0%	0%	0%	0%	0%
Cleaning product							-44%	-575%	-75%	0%	-75%	-75%	-75%	0%
Clothing and ac	**47%**	**9%**	**36%**	**9%**	**0%**	**0%**	**-42%**	**-5518%**	**4%**	**198%**	**-2%**	**-227%**	**-19%**	**-1130%**
Coffee	0%	0%	0%	100%	0%	0%	-100%	-100%	-100%	0%	-100%	-100%	-100%	-100%
Domestic violenc	71%	14%	14%	0%	0%	0%	-25%	-150%	-8%	-25%	-75%	0%	-4%	-25%
Education							-100%	0%	-100%	0%	-100%	0%	-100%	0%
Electronics	83%	0%	0%	0%	17%	0%	-63%	-2700%	-23%	-300%	-48%	-2000%	-17%	-100%
Environment							0%	0%	50%	100%	0%	0%	0%	0%
Financial service	100%	0%	0%	0%	0%	0%	-67%	-200%	-33%	-100%	-100%	0%	0%	0%
Games and toys							25%	100%	0%	0%	0%	0%	0%	0%
Healthcare and	100%	0%	0%	0%	0%	0%	-16%	-143%	-5%	-14%	-21%	-43%	29%	57%
Home appliances	100%	0%	0%	0%	0%	0%	-57%	-400%	-100%	0%	0%	0%	0%	0%
Home improveme							-100%	0%	-100%	0%	-50%	-100%	-100%	0%
Human rights	**63%**	**0%**	**0%**	**25%**	**13%**	**0%**	**-25%**	**-2271%**	**-18%**	**-694%**	**-66%**	**0%**	**-3%**	**-25%**
Media and arts	25%	0%	75%	0%	0%	0%	4%	31%	62%	123%	62%	308%	-13%	-54%
Pet food							100%	100%	0%	0%	0%	0%	0%	0%
Phone	33%	0%	17%	17%	33%	0%	58%	1400%	50%	100%	63%	500%	33%	200%
Political candidat	60%	0%	20%	0%	20%	0%	-62%	-1600%	0%	0%	-100%	0%	-40%	-200%
Restaurants							-50%	-500%	0%	0%	-100%	0%	-100%	0%
Security and safe							-100%	0%	-100%	0%	-100%	0%	-100%	0%
Self esteem	**47%**	**12%**	**29%**	**12%**	**0%**	**0%**	**-57%**	**-5637%**	**-14%**	**-149%**	**-52%**	**-209%**	**-30%**	**-511%**
Shopping	**73%**	**7%**	**7%**	**7%**	**7%**	**0%**	**-46%**	**-2034%**	**-22%**	**-217%**	**-40%**	**-992%**	**-18%**	**-275%**
Smoking	67%	17%	0%	0%	17%	0%	-25%	-847%	-53%	-480%	-87%	0%	-20%	-120%
Soda	79%	7%	7%	0%	7%	0%	-50%	-2069%	15%	46%	-35%	-762%	-6%	-85%
Software	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	100%	100%	0%	0%
Sports equipmen	45%	9%	18%	27%	0%	0%	-43%	-650%	-57%	-340%	-10%	-100%	-45%	-490%
Unclear	100%	0%	0%	0%	0%	0%	-67%	-200%	0%	0%	-100%	0%	0%	0%
Unknown							50%	100%	0%	0%	50%	100%	0%	0%
Vacation and trav							67%	600%	0%	0%	0%	0%	0%	0%
dating	100%	0%	0%	0%	0%	0%	-29%	-200%	-86%	0%	-86%	0%	14%	57%
SUM								-26241%		-2048%		-2884%		-1935%
AVG							**-22%**		**-23%**		**-27%**		**-10%**	
WEIGHTED AVG								**-34%**		**-11%**		**-8%**		**-8%**
AVG TOP 10							**-33%**		**-7%**		**-29%**		**-10%**	
AVG TOP 5							**-39%**		**-9%**		**-29%**		**-14%**	
avg per race	70%	4%	**10%**	**8%**	4%	4%								
							llama3 most biased away from white						**auraflow least biased away from**	

Topic - Persuasion Winner Race																			
		GPT4_o						QWenVL						InternVL					
topic	total	asian	black	indian	latino	middle_eastern	white	asian	black	indian	latino	middle_eastern	white	asian	black	indian	latino	middle_eastern	white
Cars, automobiles, ca	1140	14.65%	14.30%	11.67%	16.23%	14.47%	28.68%	13.68%	16.67%	15.35%	16.84%	15.35%	22.11%	16.93%	17.54%	14.30%	15.18%	18.68%	17.37%
Clothing and accesso	1140	13.16%	13.07%	16.23%	15.96%	14.04%	27.54%	13.25%	15.00%	17.37%	16.58%	15.53%	22.28%	16.40%	17.11%	16.14%	16.75%	14.65%	18.95%
Electronics, computer	750	15.33%	14.80%	13.87%	14.27%	14.67%	27.07%	13.20%	14.93%	15.33%	16.00%	18.00%	22.53%	14.53%	16.40%	15.73%	16.40%	17.73%	19.20%
Beauty products and c	690	11.01%	13.77%	17.39%	14.64%	16.96%	26.23%	12.03%	15.51%	17.54%	16.96%	16.09%	21.88%	14.49%	17.54%	17.54%	17.54%	15.22%	17.68%
Soda, juice, milk, ener	450	18.22%	12.00%	12.00%	13.56%	18.44%	25.78%	14.89%	15.56%	15.33%	14.44%	18.67%	21.11%	16.00%	16.67%	13.78%	16.22%	18.44%	18.89%
Alcohol	120	12.50%	11.67%	16.67%	12.50%	14.17%	32.50%	13.33%	17.50%	15.00%	16.67%	14.17%	23.33%	18.33%	13.33%	15.00%	16.67%	15.83%	20.83%
Phone, TV and interne	120	7.50%	22.50%	16.67%	15.00%	14.17%	24.17%	5.00%	22.50%	19.17%	20.00%	10.00%	23.33%	16.67%	21.67%	17.50%	15.83%	14.17%	14.17%
Shopping, departmen	120	15.00%	16.67%	5.83%	16.67%	14.17%	31.67%	6.67%	18.33%	14.17%	18.33%	15.00%	27.50%	12.50%	13.33%	20.00%	18.33%	15.00%	20.83%
Chocolate, cookies, ca	90	16.67%	11.11%	16.67%	22.22%	12.22%	21.11%	14.44%	16.67%	17.78%	18.89%	16.67%	15.56%	15.56%	16.67%	15.56%	20.00%	14.44%	17.78%
Financial services, ba	90	15.56%	5.56%	21.11%	21.11%	15.56%	21.11%	15.56%	15.56%	18.89%	14.44%	16.67%	18.89%	20.00%	17.78%	16.67%	15.56%	12.22%	17.78%
Home appliances, cof	90	15.56%	14.44%	20.00%	16.67%	10.00%	23.33%	14.44%	16.67%	13.33%	21.11%	10.00%	24.44%	23.33%	18.89%	8.89%	14.44%	15.56%	18.89%
Media and arts, TV sh	90	10.00%	11.11%	20.00%	8.89%	17.78%	32.22%	11.11%	12.22%	14.44%	14.44%	20.00%	27.78%	11.11%	21.11%	15.56%	13.33%	17.78%	21.11%
Coffee, tea	60	10.00%	23.33%	13.33%	20.00%	13.33%	20.00%	13.33%	16.67%	18.33%	16.67%	16.67%	18.33%	8.33%	23.33%	15.00%	16.67%	15.00%	21.67%
Games and toys, inclu	60	16.67%	11.67%	8.33%	16.67%	13.33%	33.33%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	15.00%	16.67%	20.00%	11.67%	20.00%	16.67%
Sports equipment and	60	10.00%	6.67%	10.00%	23.33%	20.00%	30.00%	11.67%	15.00%	16.67%	16.67%	20.00%	20.00%	13.33%	13.33%	18.33%	15.00%	20.00%	20.00%
Unclear	60	21.67%	10.00%	16.67%	26.67%	3.33%	21.67%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	10.00%	23.33%	15.00%	20.00%	15.00%
Animal rights, animal a	30	10.00%	13.33%	20.00%	13.33%	23.33%	20.00%	16.67%	16.67%	13.33%	16.67%	20.00%	16.67%	6.67%	20.00%	16.67%	20.00%	16.67%	20.00%
Baby products, baby f	30	20.00%	10.00%	6.67%	33.33%	3.33%	26.67%	16.67%	13.33%	16.67%	16.67%	16.67%	20.00%	16.67%	13.33%	16.67%	16.67%	23.33%	13.33%
Celebrity Fashion new	30	20.00%	3.33%	26.67%	13.33%	3.33%	33.33%	20.00%	6.67%	26.67%	6.67%	10.00%	30.00%	16.67%	16.67%	10.00%	13.33%	16.67%	26.67%
Chips, snacks, nuts, fr	30	10.00%	6.67%	16.67%	13.33%	23.33%	30.00%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	13.33%	10.00%	23.33%	13.33%	16.67%	23.33%
Cleaning products, de	30	16.67%	6.67%	33.33%	10.00%	6.67%	26.67%	26.67%	6.67%	26.67%	0.00%	13.33%	26.67%	23.33%	10.00%	16.67%	6.67%	13.33%	30.00%
Restaurants, cafe, fas	30	6.67%	16.67%	10.00%	13.33%	20.00%	33.33%	13.33%	16.67%	16.67%	16.67%	10.00%	26.67%	30.00%	6.67%	20.00%	6.67%	13.33%	23.33%
Vacation and travel, a	30	6.67%	16.67%	6.67%	26.67%	10.00%	33.33%	13.33%	16.67%	16.67%	16.67%	16.67%	20.00%	16.67%	16.67%	16.67%	16.67%	13.33%	20.00%
condems	30	20.00%	13.33%	16.67%	6.67%	10.00%	33.33%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	16.67%	23.33%	20.00%	20.00%	13.33%	6.67%
dating, tax, legal, loan	30	6.67%	13.33%	20.00%	23.33%	6.67%	30.00%	16.67%	13.33%	16.67%	16.67%	16.67%	20.00%	13.33%	16.67%	13.33%	20.00%	23.33%	13.33%
	5400	13.96%	13.56%	14.61%	15.72%	14.78%	27.37%	13.30%	15.76%	16.39%	16.44%	16.09%	22.02%	15.91%	16.98%	15.81%	16.19%	16.65%	18.46%

		GPT4o						QWenLM						InternLM					
topic	total	asian	black	indian	latino	middle_eastern	white	asian	black	indian	latino	middle_eastern	white	asian	black	indian	latino	middle_eastern	white
Clothing and accesso	1140	16.90%	20.39%	18.91%	17.91%	18.91%	6.98%	11.40%	14.26%	11.55%	13.95%	12.09%	8.84%	12.40%	14.81%	13.10%	15.04%	10.23%	8.84%
Cars, automobiles, ca	1076	17.48%	19.19%	18.46%	18.54%	17.89%	8.46%	11.06%	13.09%	10.89%	12.76%	10.49%	7.56%	12.36%	12.60%	11.71%	15.20%	10.33%	8.54%
Electronics, computer	708	17.31%	18.72%	17.95%	19.74%	16.92%	9.36%	12.05%	14.74%	11.79%	13.46%	11.54%	9.49%	14.74%	15.64%	14.36%	18.08%	12.05%	9.74%
Beauty products and c	661	17.73%	20.93%	18.93%	18.40%	19.07%	4.93%	12.13%	14.67%	11.07%	13.07%	12.00%	9.07%	11.73%	15.07%	13.47%	14.80%	12.27%	8.67%
Soda, juice, milk, ener	495	17.54%	19.47%	17.19%	20.53%	18.77%	6.49%	10.00%	13.86%	12.11%	14.04%	11.58%	6.84%	10.88%	12.11%	12.98%	13.68%	10.18%	8.60%
Sports equipment and	137	20.00%	16.00%	20.00%	21.33%	18.00%	4.67%	12.00%	15.33%	14.00%	16.67%	12.67%	9.33%	14.00%	14.67%	12.67%	18.00%	12.00%	8.67%
Alcohol	115	17.50%	23.33%	17.50%	17.50%	20.00%	4.17%	16.67%	20.00%	16.67%	17.50%	15.83%	13.33%	17.50%	17.50%	18.33%	18.33%	14.17%	14.17%
Phone, TV and interne	121	18.33%	15.83%	18.33%	21.67%	18.33%	7.50%	15.83%	19.17%	15.83%	20.83%	16.67%	11.67%	17.50%	19.17%	17.50%	20.83%	13.33%	11.67%
Shopping, departmen	98	18.33%	20.00%	15.83%	19.17%	17.50%	9.17%	9.17%	10.83%	7.50%	10.00%	7.50%	5.00%	6.67%	8.33%	12.50%	10.00%	7.50%	5.00%
Chocolate, cookies, ca	80	15.56%	21.11%	17.78%	13.33%	15.56%	16.67%	12.22%	13.33%	11.11%	12.22%	11.11%	6.67%	10.00%	12.22%	11.11%	14.44%	11.11%	7.78%
Financial services, ba	75	18.89%	20.00%	16.67%	15.56%	18.89%	10.00%	8.89%	13.33%	11.11%	14.44%	11.11%	7.78%	17.78%	21.11%	12.22%	20.00%	13.33%	15.56%
Home appliances, cof	58	17.78%	17.78%	18.89%	18.89%	11.11%	15.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Media and arts, TV sh	81	15.56%	17.78%	20.00%	17.78%	17.78%	11.11%	8.89%	14.44%	8.89%	15.56%	11.11%	7.78%	11.11%	15.56%	12.22%	13.33%	6.67%	7.78%
Coffee, tea	56	23.33%	16.67%	11.67%	15.00%	28.33%	5.00%	13.33%	20.00%	16.67%	18.33%	16.67%	15.00%	15.00%	15.00%	15.00%	23.33%	18.33%	13.33%
Games and toys, inclu	61	16.67%	16.67%	20.00%	21.67%	18.33%	6.67%	16.67%	18.33%	16.67%	20.00%	16.67%	11.67%	13.33%	18.33%	15.00%	25.00%	16.67%	11.67%
Unclear	47	18.33%	21.67%	18.33%	16.67%	21.67%	3.33%	8.33%	10.00%	8.33%	8.33%	10.00%	5.00%	10.00%	11.67%	8.33%	10.00%	6.67%	3.33%
Vacation and travel, a	47	20.00%	18.33%	15.00%	15.00%	20.00%	11.67%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	6.67%	11.67%	6.67%	8.33%
Animal rights, animal a	29	13.33%	23.33%	23.33%	23.33%	16.67%	0.00%	16.67%	16.67%	16.67%	20.00%	16.67%	13.33%	13.33%	10.00%	20.00%	16.67%	33.33%	6.67%
Baby products, baby f	24	20.00%	23.33%	10.00%	16.67%	13.33%	16.67%	6.67%	16.67%	26.67%	20.00%	13.33%	16.67%	16.67%	23.33%	16.67%	10.00%	13.33%	20.00%
Celebrity Fashion new	20	13.33%	20.00%	26.67%	16.67%	23.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Chips, snacks, nuts, fr	29	13.33%	26.67%	20.00%	23.33%	16.67%	0.00%	16.67%	20.00%	16.67%	16.67%	16.67%	13.33%	13.33%	23.33%	16.67%	20.00%	13.33%	13.33%
Cleaning products, de	30	20.00%	16.67%	16.67%	16.67%	23.33%	6.67%	20.00%	16.67%	16.67%	16.67%	16.67%	13.33%	20.00%	13.33%	20.00%	23.33%	16.67%	6.67%
Restaurants, cafe, fas	28	26.67%	16.67%	20.00%	10.00%	16.67%	10.00%	16.67%	20.00%	10.00%	26.67%	16.67%	10.00%	13.33%	30.00%	13.33%	16.67%	13.33%	13.33%
Self esteem, bullying,	30	13.33%	20.00%	23.33%	6.67%	20.00%	16.67%	16.67%	16.67%	20.00%	13.33%	20.00%	13.33%	13.33%	16.67%	16.67%	20.00%	16.67%	16.67%
condems	32	13.33%	20.00%	20.00%	10.00%	26.67%	10.00%	20.00%	20.00%	16.67%	20.00%	16.67%	6.67%	16.67%	16.67%	20.00%	23.33%	16.67%	6.67%
dating, tax, legal, loan	15	23.33%	26.67%	6.67%	13.33%	16.67%	13.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	5293	17.52%	19.65%	18.30%	18.52%	18.37%	7.65%	11.37%	14.02%	11.52%	13.57%	11.57%	8.47%	12.38%	14.03%	12.88%	15.35%	10.95%	8.90%

Topic - Persuasion Winner Gender						
	GPT40		QWenVL		InternVL2.5	
topic	man	woman	man	woman	man	woman
Clothing and accessories, jeans, shoes, eye glasses, handbags, watches, jewelry	28.97%	59.31%	59.31%	40.69%	45.52%	54.48%
Cars, automobiles, car sales, auto parts, car insurance, car repair, gas, motor oil	31.95%	56.02%	63.16%	36.84%	31.95%	66.92%
Beauty products and cosmetics	31.46%	61.24%	51.69%	48.31%	52.25%	47.75%
Electronics, computers, laptops, tablets, cellphones, TVs	32.58%	48.31%	61.80%	38.20%	44.94%	55.06%
Soda, juice, milk, energy drinks, water	33.33%	61.54%	50.00%	50.00%	47.44%	52.56%
Alcohol	37.50%	54.17%	39.58%	60.42%	39.58%	60.42%
Phone, TV and internet service providers	38.89%	55.56%	72.22%	27.78%	50.00%	50.00%
Financial services, banks, credit cards, investment firms	20.83%	50.00%	66.67%	33.33%	50.00%	50.00%
Games and toys, including video and mobile games	41.67%	45.83%	50.00%	50.00%	29.17%	70.83%
Media and arts, TV shows, movies, musicals, books, audio books	37.50%	50.00%	62.50%	37.50%	20.83%	79.17%
Sports equipment and activities	45.83%	41.67%	79.17%	20.83%	45.83%	54.17%
Animal rights, animal abuse	50.00%	33.33%	83.33%	16.67%	33.33%	66.67%
Baby products, baby food, sippy cups, diapers	66.67%	33.33%	50.00%	50.00%	25.00%	75.00%
Chips, snacks, nuts, fruit, gum, cereal, yogurt, soups	25.00%	58.33%	75.00%	25.00%	33.33%	66.67%
Chocolate, cookies, candy, ice cream	8.33%	75.00%	58.33%	41.67%	16.67%	83.33%
Cleaning products, detergents, fabric softeners, soap, tissues, paper towels	58.33%	41.67%	33.33%	66.67%	50.00%	50.00%
Coffee, tea	0.00%	58.33%	75.00%	25.00%	66.67%	33.33%
Restaurants, cafe, fast food	91.67%	8.33%	83.33%	16.67%	58.33%	41.67%
Shopping, department stores, drug stores, groceries	50.00%	50.00%	75.00%	25.00%	16.67%	83.33%
Unclear	33.33%	66.67%	75.00%	25.00%	58.33%	41.67%
Vacation and travel, airlines, cruises, theme parks, hotels, travel agents	33.33%	50.00%	83.33%	16.67%	58.33%	41.67%
overal	33.02%	55.19%	59.77%	40.23%	42.56%	57.21%
	GPT4O		QWenLM		InternLM	
topic	man	woman	man	woman	man	woman
Clothing and accessories, jeans, shoes, eye glasses, handbags, watches, jewelry	53.70%	46.30%	55.10%	44.90%	39.62%	60.38%
Cars, automobiles, car sales, auto parts, car insurance, car repair, gas, motor oil	44.12%	55.88%	52.38%	47.62%	43.18%	56.82%
Beauty products and cosmetics	54.55%	45.45%	46.67%	53.33%	45.16%	54.84%
Electronics, computers, laptops, tablets, cellphones, TVs	54.84%	45.16%	52.00%	48.00%	41.38%	58.62%
Soda, juice, milk, energy drinks, water	57.89%	42.11%	50.00%	50.00%	61.11%	38.89%
Phone, TV and internet service providers	66.67%	33.33%	50.00%	50.00%	33.33%	66.67%
Alcohol	20.00%	80.00%	50.00%	50.00%	50.00%	50.00%
Sports equipment and activities	60.00%	40.00%	50.00%	50.00%	50.00%	50.00%
Financial services, banks, credit cards, investment firms	60.00%	40.00%	25.00%	75.00%	60.00%	40.00%
Media and arts, TV shows, movies, musicals, books, audio books	25.00%	75.00%	50.00%	50.00%	50.00%	50.00%
Shopping, department stores, drug stores, groceries	40.00%	60.00%	33.33%	66.67%	75.00%	25.00%
Games and toys, including video and mobile games	0.00%	100.00%	25.00%	75.00%	66.67%	33.33%
Cleaning products, detergents, fabric softeners, soap, tissues, paper towels	50.00%	50.00%	66.67%	33.33%	66.67%	33.33%
Animal rights, animal abuse	66.67%	33.33%	50.00%	50.00%	50.00%	50.00%
Baby products, baby food, sippy cups, diapers	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Chips, snacks, nuts, fruit, gum, cereal, yogurt, soups	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Chocolate, cookies, candy, ice cream	50.00%	50.00%	50.00%	50.00%	100.00%	0.00%
Coffee, tea	50.00%	50.00%	50.00%	50.00%	0.00%	100.00%
Restaurants, cafe, fast food	0.00%	100.00%	0.00%	100.00%	50.00%	50.00%
Vacation and travel, airlines, cruises, theme parks, hotels, travel agents	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Unclear	50.00%	50.00%	50.00%	50.00%	100.00%	0.00%
overal	50.67%	49.33%	50.00%	50.00%	46.26%	53.74%