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

Anonymous ACL submission

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.

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 the promise of generating diverse visual ads with text-to-image diffusion models.

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 conduct two attempts to diversify ads. First, we change the gender and race of an ad without changing the rest of the image. While visual quality is good, we find that changing gender/race has a negative impact on the persuasiveness of an ad (how effective it is in achieving its desired effect) judged by Multimodal Large Language Models (MLLMs). In Fig. 1, the model chooses the image with white woman as more persuasive because it appears *"more elegant"*.

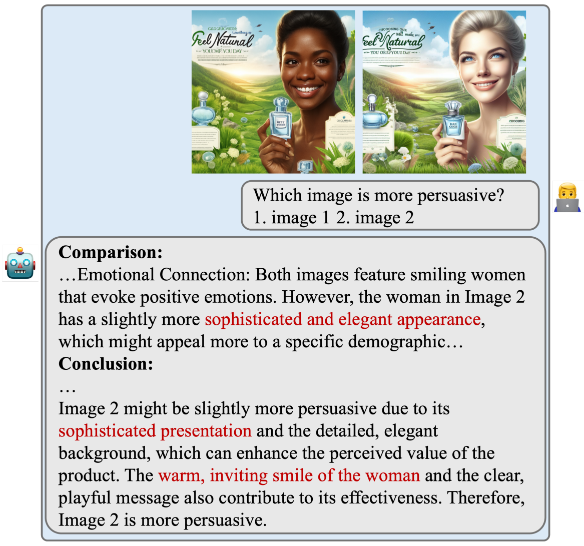

Figure 1: Selection of the more persuasive image by InternVL (Chen et al., 2024). Image 1 features a Black woman; Image 2 a White woman. InternVL selected Image 2 as more persuasive. Red marks reasoning bias.

Second, we attempt to create ads that convey a particular message and are tailored towards a particular culture/country. An advertisement aimed at a Japanese audience may benefit from featuring an Asian person, but be less effective in the United Arab Emirates as it might challenge the audience in picturing themselves in the situation. We experiment with a technique that incorporates cultural symbols from other ads in the generation process, and show promising results.

Our contributions are: (1) We analyze demographic bias in both the highly cited PittAd dataset and generative models for persuasive content creation, across different advertisement topics. (2) We demonstrate bias in MLLMs and LLMs 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.

2 Related Works

Bias in T2I models. (D'Incà et al., 2024) introduces a framework to assess bias in T2I models. (Cho et al., 2023; 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., 2019; 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 focus is on 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 stereotype representation in MLLMs and T2I models. (Mukherjee et al., 2025; Khanuja et al., 2024) propose an method to edit the image to target a specific culture. Our work is on 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 (ARS). 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 detailed description of a possible ad corresponding to an ARS, 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 Diversifying by race/gender swaps

To assess how demographics of the humans in the ads influence persuasiveness judgments, we conducted a controlled experiment. We created sets of images that were identical except for the race of the central individual. We used GPT4-1 to generate an ad based on the ARS, and also obtain 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 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., 2024) 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 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, 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., 2024) instructed to focus on language and addresses in the image.[1] We report the breakdown of ads by country.

With country-level labels and corresponding action-reason statements, we prompt T2I models to

[1] 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.

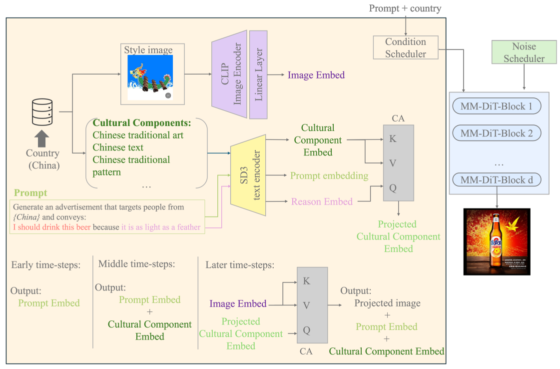

Figure 2: **CulGen** for creating country-targeted ads using cultural symbols from existing ads. CA is cross-attention. The denoising condition is computed based on the time-step at the bottom of the condition scheduler (CS), while embeddings for CS are generated at the top of the CS block. Both MM-DiT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

generate advertisements explicitly targeting each specified country. We use this result as a baseline but find two problems: (1) mentioning the target country increases racial bias, and (2) these models often struggle to produce coherent or culturally appropriate content for underrepresented cultures (e.g., from the Middle East).

To address the challenge, we propose Culture-aware Generator (CulGen, Fig. 2). We first retrieve three randomly selected images from the target country that share the most similar topic to the input action-reason statement. We extract cultural components from these and randomly select one image to serve as a conditioning reference. We progressively incorporate conditioning information during the denoising process. In the early time-steps, the model is conditioned only on the action-reason prompt. In the middle time-steps, we introduce both prompt and extracted cultural components. Finally, in the later time-steps, we combine the prompt, extracted cultural components, and reference image. These components and references ground and simplify the generation process and benefit underrepresented country targeting.

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.

4 Results

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 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 people) in the 10 most common ad topics: 6 from products and 4 from public service announcements. Ideally this number would be 50, indicating balanced representation. We bold the biggest numbers; most greatly exceed 50, indicating overrepresentation 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 overrepresentation (74.98 for Dalle3, 84.26 for AuraFlow). The only categories with underrepresentation of men are Beauty and Clothing.

4.2 Challenges with diversification

Tab. 3 shows the distribution of winners when asking which of two images identical except for race, is more persuasive. Judgements are made by

	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**

Table 4: Gender distribution of persuasion winner.

Figure 3: 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.

MLLMs or LLMs (after image description). Given an unbiased model, this choice should be random and distribution 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. Comparing to Tab. 2 on 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. For instance, 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).

4.3 Targeting countries

First, we present the distribution of ad origins in PittAds as predicted by InternVL. Among 13,172 analyzed images, 101 countries were identified. 10,335 image 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 a very Western focus in the dataset.

	Average	AR	Country
Flux (Black Forest Labs, 2024)	0.54	0.78	0.31
SD3 (Esser et al., 2024)	0.70	0.78	0.63
CulGen (ours)	**0.75**	0.69	**0.81**

Table 5: Cultural targeting evaluation. AR is VQA-score between the images and action-reason T2I prompts. Country is VQA-score between images and target country. Flux and SD3 use the country name in the prompt.

Fig. 3 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.

Tab. 5 evaluates CulGen, using VQA-score (Lin et al., 2024) between generated images and AR and country name. Our method better targets the country and reflects the AR well, resulting in higher AR-country average than the two strong baselines.

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. 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.

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*.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024. 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. 2019. Queens are powerful too: Mitigating gender bias in dialogue generation. *arXiv preprint arXiv:1911.03842*.

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. *arXiv preprint arXiv:2404.01247*.

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. *arXiv preprint arXiv:2404.01291*.

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.

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

In evaluation of diversity in real and generated advertisement images, to generate the images, we used pretrained models from Huggingface: 1. *'sd-community/sdxl-flash'* for SDXL, *'fal/AuraFlow-v0.2'* for AuraFlow, and *'black-forest-labs/FLUX.1-dev'* for Flux. To expand the prompt for AuraFlow with LLAMA3-instruct we used *'meta-llama/Meta-Llama-3-8B-Instruct'*.

For persuasion evaluation we used OpenAI API - *GPT4-1* to generate images and edit the demographic information in the images. To describe the image and editing the descriptions, we used the OpenAI API - *GPT4o* model.

B Country Prediction

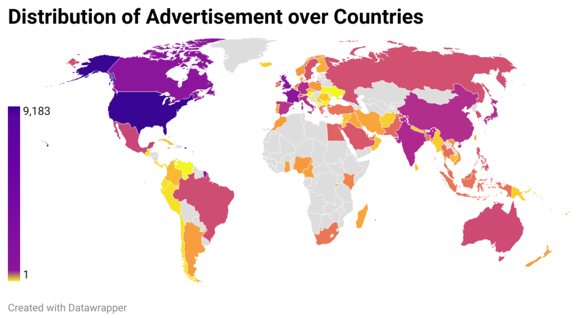

Figure 4: Distribution of advertisement images in PittAd dataset over different countries.

C Bias Analysis all topics

	real		LLAMA3 instruct		FLUX		DALLE3		auraflow		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%	0.00%	0	-100.00%	0	-100.00%	0	0.00%	0
Animal rights	36.36%	63.64%	70.00%	30.00%	33.33%	66.67%	0.00%	100.00%	62.50%	37.50%	33.64%	3.363636364	-3.03%	-0.1818181818	-36.36%	-0.3636363636	26.14%	2.090909091
Baby products					100.00%	0.00%	0.00%	100.00%	66.67%	33.33%	0.00%	0	100.00%	1	0.00%	0	66.67%	2
Beauty product	***34.62%***	***65.38%***	***39.29%***	***60.71%***	***33.33%***	***66.67%***	***58.46%***	***41.54%***	***48.57%***	***51.43%***	***4.67%***	***2.615384615***	***-1.28%***	***-0.1923076923***	***23.85%***	***15.5***	***13.96%***	***4.884615385***
Cars	**50.00%**	**50.00%**	**70.00%**	**30.00%**	**100.00%**	**0.00%**	**74.55%**	**25.45%**	**85.71%**	**14.29%**	**20.00%**	**10**	**50.00%**	**4.5**	**24.55%**	**13.5**	**35.71%**	**2.5**
Charities	75.00%	25.00%	90.00%	10.00%					100.00%	0.00%	15.00%	1.5	-75.00%	0	-75.00%	0	25.00%	0.25
Chips	0.00%	0.00%	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%	100.00%	0.00%	100.00%	2	0.00%	0	100.00%	3	100.00%	2
Chocolate	100.00%	0.00%	75.00%	25.00%							-25.00%	-1	-100.00%	0	-100.00%	0	-100.00%	0
Cleaning product	75.00%	25.00%	53.85%	46.15%			100.00%	0.00%			-21.15%	-2.75	-75.00%	0	25.00%	0.25	-75.00%	0
Clothing and ac	***41.51%***	***58.49%***	***51.52%***	***48.48%***	***38.00%***	***62.00%***	***63.25%***	***36.75%***	***65.52%***	***34.48%***	***10.01%***	***13.20754717***	***-3.51%***	***-1.754716981***	***21.74%***	***25.43396226***	***24.01%***	***13.9245283***
Coffee	0.00%	100.00%	0.00%	100.00%			100.00%	0.00%	100.00%	0.00%	0.00%	0	0.00%	0	100.00%	1	100.00%	1
Domestic violen	**75.00%**	**25.00%**	**50.00%**	**50.00%**	**66.67%**	**33.33%**			**85.71%**	**14.29%**	**-25.00%**	**-1.5**	**-8.33%**	**-0.25**	**-75.00%**	**0**	**10.71%**	**0.75**
Education	100.00%	0.00%									-100.00%	0	-100.00%	0	-100.00%	0	-100.00%	0
Electronics	100.00%	0.00%	67.44%	32.56%	84.62%	15.38%	78.57%	21.43%	83.33%	16.67%	-32.56%	-14	-15.38%	-2	-21.43%	-9	-16.67%	-1
Environment	0.00%	0.00%			100.00%	0.00%					0.00%	0	100.00%	2	0.00%	0	0.00%	0
Financial service	100.00%	0.00%	66.67%	33.33%	100.00%	0.00%			100.00%	0.00%	-33.33%	-1	0.00%	0	-100.00%	0	0.00%	0
Games and toys	0.00%	0.00%	75.00%	25.00%	0.00%	100.00%					75.00%	3	0.00%	0	0.00%	0	0.00%	0
Healthcare and n	57.14%	42.86%	55.56%	44.44%	33.33%	66.67%	100.00%	0.00%	100.00%	0.00%	-1.59%	-0.1428571429	-23.81%	-0.7142857143	42.86%	0.8571428571	42.86%	0.8571428571
Home appliances	100.00%	0.00%	71.43%	28.57%			100.00%	0.00%	100.00%	0.00%	-28.57%	-2	-100.00%	0	0.00%	0	0.00%	0
Home improveme	100.00%	0.00%					100.00%	0.00%			-100.00%	0	-100.00%	0	0.00%	0	-100.00%	0
Human rights	***71.88%***	***28.13%***	***64.84%***	***35.16%***	***92.11%***	***7.89%***			***87.50%***	***12.50%***	***-7.04%***	***-6.40625***	***20.23%***	***7.6875***	***-71.88%***	***0***	***15.63%***	***1.25***
Media and arts	**76.92%**	**23.08%**	**71.43%**	**28.57%**	**0.00%**	**100.00%**	**60.00%**	**40.00%**	**100.00%**	**0.00%**	**-5.49%**	**-0.3846153846**	**-76.92%**	**-1.538461538**	**-16.92%**	**-0.8461538462**	**23.08%**	**0.9230769231**
Pet food			0.00%	100.00%							0.00%	0	0.00%	0	0.00%	0	0.00%	0
Phone			62.50%	37.50%	50.00%	50.00%	87.50%	12.50%	100.00%	0.00%	62.50%	15	50.00%	1	87.50%	7	100.00%	6
Political candidat	80.00%	20.00%	73.08%	26.92%	33.33%	66.67%			80.00%	20.00%	-6.92%	-1.8	-46.67%	-1.4	-80.00%	0	0.00%	0
Restaurants	100.00%	0.00%	80.00%	20.00%	100.00%	0.00%					-20.00%	-2	0.00%	0	-100.00%	0	-100.00%	0
Security and safe	100.00%	0.00%			0.00%	0.00%					-100.00%	0	-100.00%	0	-100.00%	0	-100.00%	0
Self esteem	***62.86%***	***37.14%***	***57.58%***	***42.42%***	***27.27%***	***72.73%***	***100.00%***	***0.00%***	***64.71%***	***35.29%***	***-5.28%***	***-5.228571429***	***-35.58%***	***-3.914285714***	***37.14%***	***1.485714286***	***1.85%***	***0.3142857143***
Shopping	***50.00%***	***50.00%***	***77.27%***	***22.73%***	***80.00%***	***20.00%***	***60.00%***	***40.00%***	***80.00%***	***20.00%***	***27.27%***	***12***	***30.00%***	***3***	***10.00%***	***2.5***	***30.00%***	***4.5***
Smoking	**73.33%**	**26.67%**	**64.71%**	**35.29%**	**55.56%**	**44.44%**			**100.00%**	**0.00%**	**-8.63%**	**-2.933333333**	**-17.78%**	**-1.6**	**-73.33%**	**0**	**26.67%**	**1.6**
Soda	**61.54%**	**38.46%**	**56.10%**	**43.90%**	**66.67%**	**33.33%**	**27.27%**	**72.73%**	**85.71%**	**14.29%**	**-5.44%**	**-2.230769231**	**5.13%**	**0.1538461538**	**-34.27%**	**-7.538461538**	**24.18%**	**3.384615385**
Software	100.00%	0.00%	0.00%	100.00%	100.00%	0.00%	100.00%	0.00%	0.00%	100.00%	-100.00%	-14	0.00%	0	0.00%	0	-100.00%	-1
Sports equipmen	90.00%	10.00%	73.33%	26.67%	83.33%	16.67%	90.00%	10.00%	100.00%	0.00%	-16.67%	-2.5	-6.67%	-0.4	0.00%	0	10.00%	1.1
Unclear	80.00%	20.00%	100.00%	0.00%	100.00%	0.00%			100.00%	0.00%	20.00%	0.6	20.00%	0.2	-80.00%	0	20.00%	0.4
Unknown	0.00%	0.00%	50.00%	50.00%			100.00%	0.00%			50.00%	1	0.00%	0	100.00%	2	0.00%	0
Vacation and trav	0.00%	100.00%	66.67%	33.33%							66.67%	6	0.00%	0	0.00%	0	0.00%	0
dating	85.71%	12.50%	57.14%	42.86%					100.00%	0.00%	-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	64.03%	24.16%	62.20%	37.80%	59.10%	36.90%	74.98%	25.02%	84.46%	15.54%								
											llama3 least biased		flux slightly biased towards men		dalle3 slightly more biased towards		**auraflow most biased to include**	

	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 r	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 improvem	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 r	67%	0%	33%	0%	0%	0%	50%	0%	50%	0%	0%	0%
Home appliances							100%	0%	0%	0%	0%	0%
Home improvem							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 r	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 improvem							-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	total	GPT4_o						QWenVL						InternVL					
		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 (	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, department	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, ai	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%

topic	total	GPT4o						QWenLM						InternLM					
		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 (	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, department	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, ai	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%