GUESS THE UNIFIED MODEL: DOMAIN AND LINGUISTIC EFFECTS IN GENERATED IMAGES

Jasin Cekinmez* **Ryo Mitsuhashi*** **Yida Yin**
Princeton University
{jasincekinmez, rm4411}@princeton.edu

ABSTRACT

With unified model generated images now widespread online, attributing their model of origin offers a path toward transparency and deeper insight into the characteristic behaviors of individual models. Prior work has explored provenance in LLM generated text, diffusion model images, and datasets, but the separability of unified model-generated images remains an underexplored area. We address this gap by examining separability across corruption, domains, and prompt languages using images generated by seven unified models. We show that model attribution is highly feasible as our model achieves near perfect accuracy with around 20K images per model. Corruptions and structural perturbations have only a modest effect on attribution performance, and cross-domain generalization reveals that semantic content contributes to separability but is not the dominant signal. Finally, we observe that for most models, prompt language attribution is around chance levels, suggesting minimal language specific visual signatures. These findings highlight consistent model-specific visual characteristics in unified models outputs and open new directions for tracing and auditing generative image pipelines. Our code is available here [1].

1 INTRODUCTION

With the rapid advancement of large language models (LLMs), recent years have seen these systems expand beyond text into images, audio, and video, giving rise to unified multimodal models capable of jointly understanding and generating diverse forms of data. Models such as ChatGPT (OpenAI et al., 2024) and Gemini (Team et al., 2025) now produce high quality images that are increasingly difficult to distinguish from real content and are widely deployed across online platforms. As a result, model-generated images have become pervasive in real-world settings, raising urgent questions about transparency, accountability, and the ability to trace synthetic content back to its source.

Understanding the origin of synthetic images is essential for characterizing model specific behaviors and failure modes, as well as for responding to misuse such as misinformation or deceptive content. Prior work has explored provenance and attribution, including dataset attribution for analyzing visual bias (Zeng et al., 2024; Liu & He, 2025), classification of LLM generated text (Sun et al., 2025), and attribution of diffusion model generated images (Xu et al., 2025). However, images generated by unified models differ fundamentally from those produced by diffusion pipelines, as they are conditioned on language, vision, and internal cross-modal representations. Despite their deployment, the separability of unified model generated images remains underexplored.

We find that a classifier based on five open-source unified models achieves high accuracy on this task even with limited training images, indicating clear idiosyncrasies in the unified models. Notably, with 500 training images per model, the classifier achieves 59.8% accuracy on a held-out test set, compared to the random chance of 20%. With 25K images per model, the accuracy reaches 99.9%.

We analyze what factors influence classification accuracy by applying corruptions and structural perturbations. We find that the images are separable in high-level features, spatial structure, object-

*Equal contribution. Author order was determined by a coin flip.
[1]Project code: https://github.com/rm-3284/Unified-model-image-classification

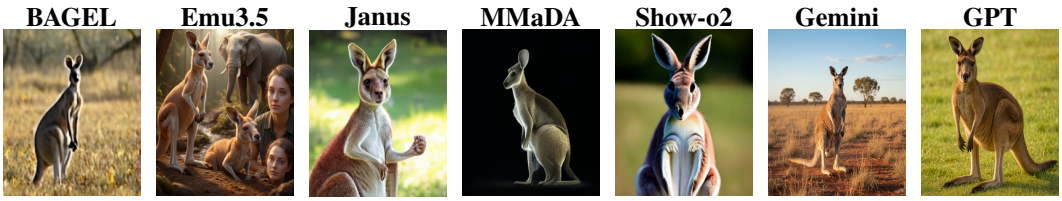


Figure 1: Images generated by each of the seven unified models using the prompt "kangaroo"

level structure, and color distribution. We also observe that image attribution is not driven by the semantic content of the images by conducting out-of-distribution experiments.

Lastly, we investigate whether the prompt language affects the distribution of the generated images and find that the distribution of generated images do not depend on the prompt languages for many state of the art models. Specifically, we conduct image attribution tasks across five languages for five open-source and two closed-source unified models and find that the classification accuracy stays at the chance-level 20% for two of the open-source models and both of the closed-source models.

2 Related Work

Dataset Classification. Prior work on dataset classification has shown that visual data sources often leave strong, identifiable signatures. Torralba & Efros (2011) demonstrated that image datasets from the 2010s could be reliably distinguished by a classifier, revealing substantial dataset, specific biases. More recently, Liu & He (2025) scaled this analysis to larger and more diverse datasets and found that, despite increased diversity, datasets remain highly separable indicating that systematic biases persist. Similar conclusions have been drawn in large-scale visual settings (Zeng et al., 2024). Unlike dataset classification, which captures biases arising from data collection and curation, unified model classification focuses on biases introduced during model creation and training. This helps us better understand how design and training choices shape the biases seen in generated images.

Attribution of Machine-Generated Content. Prior work demonstrates that generated content can often be attributed to the specific model that produced it. In the text domain, Sun et al. (2025) demonstrate that text can be attributed to its originating LLM, indicating that generation introduces stable, model-specific biases. Similarly, in the image domain, Xu et al. (2025) show that images generated by diffusion models can be reliably attributed to their source model. A systematic study of image attribution in unified models is warranted, as it remains underexplored and are central to understanding whether generated images can be reliably traced back to their source model.

Prompt Language Effects. Beyond model identity, recent work raises questions about what linguistic and cultural information is preserved or homogenized during image generation. Shi et al. (2025) show that text-to-image models often produce culturally Westernized imagery even when prompted in non-Western languages, suggesting a collapse of culturally specific conditioning in the generation process. Other studies demonstrate that properties of the prompt language, such as whether it is grammatically gendered and the gender stereotypes embedded in that language or culture can significantly influence generated images, particularly with respect to gender presentation and social roles (Friedrich et al., 2024; Saeed et al., 2025). These findings present a tension, while some cultural signals appear to be erased, others are strongly preserved and propagated through generation. This raises a broader question of whether prompt language leaves identifiable traces in the output of unified models, and whether such traces could be leveraged to infer properties of the input language. Understanding whether and how language-specific information remains separable in these models has important implications for robustness, bias, and security sensitive applications.

3 METHODOLOGY

3.1 CLASSIFICATION TASK

Following the dataset classification formulation introduced in Liu & He (2025), we treat attribution as a multi-class classification problem over data sources. In that work, dataset classification is used to evaluate separability between datasets, serving as a diagnostic for systematic biases or statistically meaningful differences in their underlying distributions.

We adapt this formulation to unified model attribution by treating each of the N unified models as a distinct class. Given an image, the task is to predict which unified model generated the image, resulting in an N-way classification problem. This reframing allows us to quantify attribution fidelity by measuring how well model-specific signals are preserved and separable across unified models.

3.2 CLASSIFIER AND IMAGE GENERATION SETUP

We use ConvNeXT (Liu et al., 2022) as our classifier since it offers strong accuracy and computational efficiency for this task. We use the ConvNeXT-Tiny model because we use a relatively small amount of training data and observe no improvement from larger ConvNeXT variants.

We generate images using the same prompts across all unified models and languages, so that the classifier learns model-specific visual cues rather than differences in subject matter. We then train a ConvNeXT classifier from scratch on the resulting training set and evaluate performance on a held-out test set. In all the experiments, we train a classifier until 200 epochs and measure the accuracy on a held-out test set at the last iteration. We used a learning rate of 1e-3 and warmup epochs of 2. Batch size ranges from 32 to 256 depending on the number of training data. All the images from unified models are generated with 1:1 aspect ratio with default resolution and resized into 224×224 before given to the classifier.

3.3 OVERVIEW OF EXPERIMENTS

We conduct five experiments. The first three use images generated by open-source models, while the latter two include images generated by both open- and closed-source models. The open-source models are BAGEL (Deng et al., 2025), MMaDA (Yang et al., 2025), Emu-3.5 (Cui et al., 2025), DeepSeek Janus-Pro-7B (Chen et al., 2025), and Show-o2 (Xie et al., 2025). The closed-source models are Gemini 2.5 Flash Image (Nano Banana) and GPT-Image-1. See Figure 1 for example images from each model.

The first experiment explores the effects of scaling, where we vary the number of training examples to measure how separability changes with size. The second experiment examines corruption, where we apply corruptions to the images prior to training to test how separability depends on low-level features. The third experiment examines how the structural bias influences separability by applying structural perturbations to the images. The fourth experiment studies Out-of-Distribution (OOD) generalization, where we train on a specific domain and evaluate on all domains. Lastly, the language experiment tests whether we can identify the prompt language conditioned on knowing the unified model, in order to determine whether language provides any visual cues.

4 RESULTS

4.1 SCALING TRENDS

We utilize the MJHQ-30K dataset (Li et al., 2024) to generate images for the 5 open-source models for this experiment since it provides a large collection of complex and descriptive prompts, allowing us to examine separability under controlled and realistic prompting conditions. We trained a 5-way classifier (BAGEL, Emu3.5, Janus, MMaDA, Show-o2) with different number of training images, ranging from 100 images per model to 25K images per model. For every training run, we used a held-out test set containing 5K images per model. With 100 images per model, the classifier achieved 36% accuracy over the 5 models, which surpasses a random guess of 20%. With 3K training images

per model, the classifier achieved an accuracy of >90% and with 25K training images per model, the accuracy reached 99.9% as shown in table 1.

# Training images	Accuracy (%)
100	36.1
500	59.8
1000	69.7
2000	84.2
3000	93.9
5000	96.2
10000	97.8
20000	99.8
25000	99.9

Table 1: Scaling effect on accuracy by the number of training images per model

While discerning models from images is not challenging for the classifier, some models are easier to discern than others. Table 2 reveals asymmetries in how different models are identified. Janus and Show-o2 have the highest recall rates (57.3% and 61.2%) when the number of training images is 100 per model indicating that their outputs are more separable than the remaining models. In contrast, Emu achieves a relatively high precision rate but one of the lowest recall rates showing that though the classifier does not identify Emu correctly very often, when it does make an Emu prediction, it is usually correct. This suggests that Emu's outputs may contain some distinct visual characteristics, but these cues are not consistently present. Overall, this result highlights that separability is not uniform across models. Qualitatively, we found that the images generated by Emu are the most consistent and rarely disfigured, and the images by MMaDA often have simpler background than others as shown in figure 2, which could explain the high precision for Emu and MMaDA.

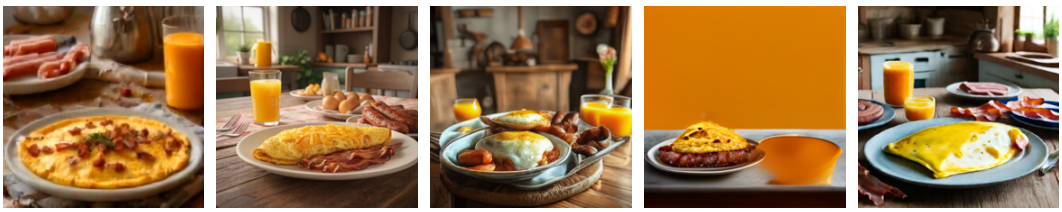
True \ Pred	BAGEL	Emu	Janus	MMaDA	Show-o2
BAGEL	1.6	4.8	30.4	18.1	45.1
Emu	1.4	11.0	25.0	26.7	35.9
Janus	1.4	2.6	57.3	20.0	18.7
MMaDA	1.1	5.7	28.8	47.0	17.4
Show-o2	1.9	3.8	23.4	9.7	61.2

(a) Recall (Row-normalized)

True \ Pred	BAGEL	Emu	Janus	MMaDA	Show-o2
BAGEL	19.7	17.1	18.8	13.8	25.6
Emu	22.0	38.9	14.7	22.1	20.5
Janus	17.3	8.2	35.4	16.9	9.9
MMaDA	18.5	20.5	17.5	39.4	9.6
Show-o2	22.5	15.3	13.5	7.8	34.4

(b) Precision (Col-normalized)

Table 2: **Recall and Precision of the 5-way classifier trained with 100 images per model.** True is the ground-truth label of the generated images and Pred is the prediction the classifier made.



(a) BAGEL (b) Emu3.5 (c) Janus (d) MMaDA (e) Show-o2

Figure 2: Images generated by different models with a prompt "american breakfast, photography, rustic farm house kitchen and dining room. omelette cheesey photography 4k, sausage links and bacon breakfast meat. served dishes. orange juice background"

4.2 CORRUPTION

We use the same dataset of MJHQ-30K (Li et al., 2024) for this experiment. This experiment evaluates the effect of low level features of the images on the classification accuracy. We apply the corruptions to both the training and test data. We use 25K images for training and 5K images for test in all the corruption experiments. To investigate the impact of low level features, we perform ablations using color jittering, Gaussian noise, Gaussian blur, and resizing. Despite the corruption of low level features, accuracy remained at approximately 95% (table 3), suggesting that the model relies on high level feature representations that remain robust to these perturbations.

Corruption	Parameter	ACC (%)
no corruption	n/a	99.9
Color jitter	strength 1	94.4
Color jitter	strength 2	95.2
Gaussian noise	std 0.2	96.3
Gaussian noise	std 0.3	95.2
Gaussian blur	radius 3	99.4
Gaussian blur	radius 5	98.7
Resize	64×64	96.6
Resize	32×32	85.2

Table 3: Accuracy with corruption

Transformation	ACC (%)
no transformation	99.9
Depth	83.2
SAM	79.2
Random pixel shuffle	72.7

Table 4: Accuracy under structural transformations

4.3 STRUCTURAL PERTURBATION

We use the same dataset of MJHQ-30K (Li et al., 2024) for this experiment. This experiment evaluates the effect of structural bias of the images on the classification accuracy. For all the perturbations, we applied them to training data and test data. We used 25K images for training and 5K images for test in all the experiments. We transformed images using Depth-Anything-V2 (Yang et al., 2024) to make a depth map of images and Segment-Anything Model (Kirillov et al., 2023) to apply image segmentation on images. We make a depth map to isolate the 3D structural information while removing color/texture and segmentation to isolate object-level structure and boundaries. We also apply randomized pixel shuffle to see the specificity of color distribution in each model.

With depth and segment transformation, the accuracy drops to around 80% (table 4). This means that the images from each model have some characteristics in depth-map and segments but these elements alone are not enough to classify the images. Similarly, the accuracy of 72% with random pixel shuffle suggests that the color distribution of images are different from one model to another but that cannot explain all the differences between models.

4.4 OUT-OF-DISTRIBUTION EXPERIMENT

4.4.1 10D DATASET CREATION

We construct prompts that are mutually exclusive across domains and collectively exhaustive within each domain, so we define ten semantic domains: animals, vehicles, arts and works, landscapes, foods and drinks, clothing, interior spaces, household items, buildings, and people, chosen to represent broad, realistic visual concepts. Within each domain, we generate 300 prompts, yielding 3K prompts in total. We call this set of prompts the 10D (10 Domains) Dataset. These prompts are minimal, capturing the concept rather than subjective or stylistic phrases. We avoid adjectives, viewpoints, artistic styles, and narrative elements to eliminate subjective variance and ensure that prompts capture only the core concept being represented. Some examples are shown in table 5.

Prompt generation is first done by asking an LLM to propose subcategories within each domain, and then we query the LLM for concepts within each subcategory. All concepts are reviewed to confirm that they fit their respective domain, do not overlap with other domains, and remain as objective as possible. This construction allows us to yield an OOD evaluation setting in which separability is driven by domain structure, enabling us to isolate model behavior at the conceptual level.

Example	Animals	Vehicles	Arts and Works	Landscapes	Food and Drinks
1	Platypus	Asphalt Paver	Oil Impasto	Geyser	Wheat
2	Hare	Limousine	Trading Card Illustration	Bayou	Corn
3	Camel	Aircraft carrier	Architectural Photo Study	Levee	Strawberry
4	Tern	Horse cart	Glassblown Vessel Form	Atoll	Edamame
5	Stingray	Mobile crane	Quilted Art Wall Piece	River Canyon	Feta
6	Turtle	Harbor tug	Album Cover Artwork	Lush Forest	Brazil Nut
7	Firefly	Gas balloon	Surreal Collage Poster	Light Fog	Cinnamon
8	Tarantula	Bicycle	Pixel Art Tileset Map	Bog	Hummus
9	Clam	Tow Truck	Spray Paint Street Mural	Butte	Turkish Coffee
10	Starfish	Rowboat	Ad Campaign Storyboard	Seamount	Kombucha

Ex.	Clothing	Interior Spaces	Household Items	Buildings	People
1	Jumpsuit	Basement	Freezer	Cabin	Doctor
2	Fanny Pack	Studio	Air Conditioner	Skyscraper	Professor
3	Jeans	Arcade Room	Printer	Convention Center	Sheriff
4	Midi Skirt	X-Ray Room	Rug	Data Center Building	Musician
5	Cardigan	Altar Area	Clock	Power Plant	Journalist
6	Baseball Cap	Concert Hall	Mug	Water Treatment Plant	Engineer
7	Hijab	Indoor Tennis Court	Fork	City Hall	DJ
8	Sandals	Bowling Alley	Toolbox	Fire Station	Court Clerk
9	Hazmat Suit	Baggage Claim	Broom	Hotel	Ventriloquist
10	Utility Vest	Steam Room	Napkins	Supermarket	Bookkeeper

Table 5: Ten representative examples for each domain in the 10D dataset.

4.4.2 Out-of-Distribution Experiment Trends

This experiment evaluates whether the classifier's performance is driven by the underlying semantic content of the images or by spurious distributional trends and stylistic biases. To run it, we treat each of the 10 semantic domains in 10D as a separate OOD setting. For each domain, we train a classifier to predict the generating unified model using 200 randomly sampled images per unified model from that domain. We then evaluate each domain classifier on a held-out test set of 100 images per unified model from every domain. This yields a 10×10 accuracy matrix, where rows correspond to the domain used for training and columns to the domain used for evaluation.

The cross-domain accuracy is shown on table 6. In order to establish a baseline, we trained 10 classifiers mixing the 10 domains with the same number of training images, and the average accuracy is 46.7%. As expected, the classifier trained on one domain and evaluated on the same domain (the diagonal of table 6) achieve higher accuracies, except for arts and works. It achieved 45.7% accuracy which is lower than the baseline, and this could be explained by a wide variety of the art objects, such as landscapes, portraits, and logos. Also, the matrix is not symmetric. For example, the accuracy of a classifier trained on vehicles and evaluated on food and drinks is 24.4%, but the accuracy of a classifier trained on food and drinks and evaluated on vehicles is 39.1%.

Although the prompts we provided are contained in a single domain and independent from each other, it is possible that the images produced by the models are more likely to contain some domains than others depending on the prompts. Therefore, we gave images to an image understanding model, Qwen3-VL (Bai et al., 2025), and asked "In the image, do you see {question_domain}? Answer the question with just yes or no." The result is shown in table 7. Surprisingly, there is not much correlation between the frequency and accuracy, which thus points toward the idea that the classification is not leveraging semantic content. Rather, it is likely relying on non-semantic, model-specific visual cues that are independent of the domain semantics.

4.5 Language Experiment

We use the same MJHQ-30K dataset (Li et al., 2024) to evaluate the linguistic separability of each unified model, examining whether visual representations are language dependent. We begin by translating 1000 randomly sampled prompts using Google Translate into 4 additional languages: Spanish, Turkish, Japanese, and Simplified Chinese, chosen to cover diverse linguistic families.

Train \ Eval	animals	arts and works	buildings	clothing	food and drinks	household items	interior spaces	landscapes	people	vehicles
animals	58.6	37.5	39.6	39.0	43.3	42.3	45.0	36.4	41.7	35.1
arts and works	36.6	45.7	38.6	41.1	32.1	34.9	42.7	29.0	45.3	32.6
buildings	39.1	36.6	71.1	37.1	30.2	37.4	53.0	44.7	52.1	49.6
clothing	41.6	35.9	31.3	61.1	41.9	45.1	37.3	34.0	41.0	31.3
food and drinks	50.0	36.0	41.1	48.6	65.1	43.1	41.1	42.1	48.6	39.1
household items	45.4	38.6	40.0	52.6	46.7	55.7	48.3	31.3	48.0	38.4
interior spaces	40.6	42.7	53.4	36.3	39.6	44.4	66.4	40.6	52.0	47.6
landscapes	33.4	35.3	45.0	34.3	35.2	33.6	41.4	55.7	43.3	41.4
people	37.1	37.3	45.0	36.9	37.5	37.7	51.7	36.0	72.3	42.4
vehicles	29.4	36.0	52.1	33.7	24.4	27.9	42.0	41.0	46.9	64.1

Table 6: **Cross-Domain Accuracy Heatmap of 7-way Classifier**. Cell (X, Y) where X is the training domain and Y is the evaluation domain shows the accuracy of a classifier trained on images from domain X and evaluated on images from domain Y, across the 7 models. For example, a classifier trained on interior spaces and evaluated on buildings achieves an accuracy of 53.4%, exceeding random chance ($100/7 \approx 14.3\%$).

Using these 5 languages (including English), we train a classifier to identify the prompt language from the generated images for each unified model, and then evaluate overall accuracy on a held-out test set. We used 700 images for each language for training and 300 images for evaluation.

The accuracies for each model are shown in table 8. As for Bagel, Emu, Gemini, and ChatGPT, the images produced by different languages are inseparable as the accuracy is the same as random guess of 20%. In Janus, we observed that while images generated by English, Spanish, and Chinese are pretty high-quality, images for Japanese and Turkish prompts are often scenery that is unrelated to the prompts with most of them being a high tower or a mountain. In particular, in many of the images for Japanese prompts and some of the images for Chinese prompts, there is a temple-like house on the left-bottom (figure 3). As for MMaDA, we observed that the images generated by Japanese and Turkish prompts often do not follow the prompts and are just some abstract objects in the middle of the images (figure 4). This explains the high accuracy for Japanese and Turkish prompts and moderately high confusion of Japanese and Turkish prompts (table 9). We observed that Show-o2 generates a picture of Asian women with high likelihood when the prompt language is Japanese or Chinese regardless of what the prompt is saying (figure 5).

5 Conclusion

As synthetic imagery increasingly circulates, practical mechanisms for distinguishing and attributing model-generated images are needed to support downstream analysis, monitoring, and measurement of generative ecosystems. Our results show that such attribution is feasible, a classifier trained on images from five open-source unified models substantially outperforms chance, achieving 93.9% accuracy with only 3K images per model and approaching perfect performance at larger scales.

Orig. \ Quest.	animals	arts and works	buildings	clothing	food and drinks	household items	interior spaces	landscapes	people	vehicles
animals	96.7	5.7	4.0	3.0	6.3	3.0	6.7	63.3	2.0	0.7
arts and works	18.3	82.4	23.9	24.9	5.3	24.6	32.9	29.9	25.2	7.6
buildings	0.3	11.0	95.3	43.7	2.7	6.7	25.0	59.0	51.7	41.3
clothing	4.0	10.3	27.0	77.3	7.3	28.7	43.3	33.7	47.7	5.3
food and drinks	3.0	0.3	4.3	3.0	90.0	57.8	49.2	16.3	4.3	0.7
household items	3.0	15.7	12.3	12.7	26.3	89.3	83.7	15.7	6.0	1.0
interior spaces	1.3	29.0	45.3	66.3	8.7	19.3	94.3	13.0	63.3	10.0
landscapes	11.0	11.7	12.7	4.0	0.0	0.0	3.0	98.0	5.3	3.0
people	4.3	26.7	44.3	99.0	20.3	20.3	79.7	12.7	99.0	13.3
vehicles	5.0	5.3	35.7	46.0	2.3	1.3	2.3	73.0	50.0	71.7

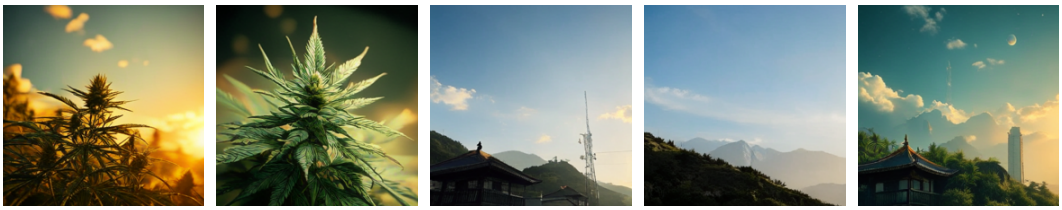
Table 7: **Domain Frequency Heatmap of the generated images**. Orig. refers to the prompt's domain, while Quest. refers to the domain queried for presence in the image. Each domain comprises 300 prompts evaluated across 7 models (2.1K images total). Cell values indicate the percentage of images from the Orig. domain that contain elements of the Quest. domain (e.g., (people, animals) = 4.3%).

Model	Accuracy (%)
BAGEL	21.2
Emu	21.9
Janus	52.9
MMaDA	34.2
Show-o2	54.0
Gemini	20.1
ChatGPT	21.9

Table 8: Accuracy on prompt language classification from images

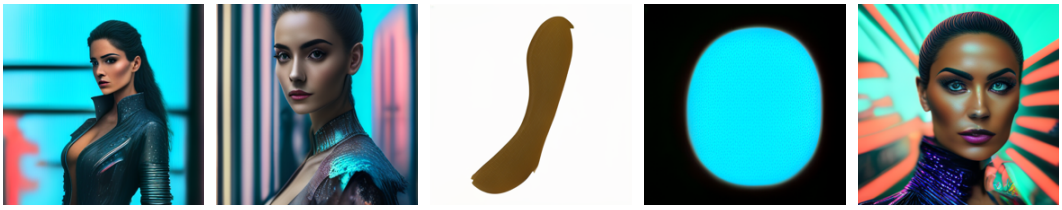
Attribution remains robust under corruptions and structural perturbations, indicating that separability is driven by high-level visual characteristics rather than low-level artifacts or semantic content.

At the same time, our findings reveal clear limits to what can be inferred from generated images alone. While model identity can often be recovered with high confidence, prompt language does not reliably influence the distribution of generated images once the generating model is fixed. Across five languages and both open- and closed-source models, language attribution remains near chance, suggesting that linguistic variation leaves weak visual traces in current unified systems. Together, these results indicate that unified multimodal models exhibit stable, distinguishable generation behavior at the model level, while finer-grained contextual signals remain difficult to recover. This delineates the scope of current image attribution methods and highlights important directions for future work on understanding, characterizing, and monitoring large-scale synthetic image generation.



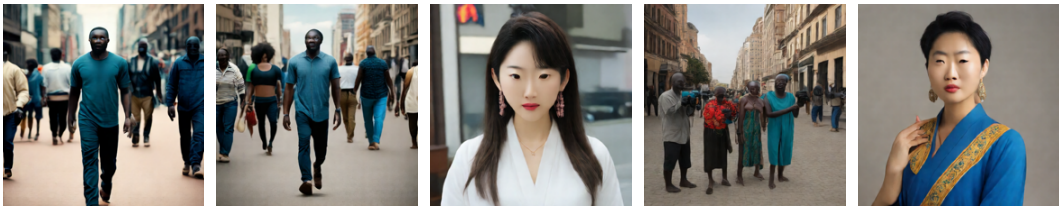
(a) English (b) Spanish (c) Japanese (d) Turkish (e) Chinese

Figure 3: Images generated by Janus with a prompt ”Golden Hour, cannabis, hyper realistic, futuristic optics, highly detailed” in different languages



(a) English prompt (b) Spanish prompt (c) Japanese prompt (d) Turkish prompt (e) Chinese prompt

Figure 4: Images generated by MMaDA with a prompt ”the most beautiful woman in the world beautiful colombian woman wearing cyberpunk clothes, standing in the rain on a cyberpunk city street, high angle, neon lights, Use a Nikon D850 DSLR camera with a 200mm lens at F 1.2 aperture setting to isolate the subject, full body shot ” in different languages



(a) English prompt (b) Spanish prompt (c) Japanese prompt (d) Turkish prompt (e) Chinese prompt

Figure 5: Images generated by Show-o2 with a prompt ”a photorealistic photo of an african male walking in the city with other people in the background” in different languages

There are several promising directions for future work. One is to apply mechanistic interpretability techniques to better understand which visual features drive separability, and why certain factors make attribution more difficult. Another avenue is to investigate how prompt complexity influences separability, particularly whether richer prompts reduce the degree to which model-specific biases and artifacts appear in the generated images. Finally exploring additional sources of variation such as prompt style, output resolution, and changes in model parameters may further help characterize what governs separability in unified model generated images.

6 Acknowledgments

We would like to thank Professor Zhuang Liu for his invaluable support and resources throughout this project. This research was supported by Princeton University’s Office of Undergraduate Research Undergraduate Fund for Academic Conferences through the ”Hewlett Foundation”, ”Marboe and Hewson Family Research” Funds, and an Anonymous Fund as well as the Center for Statistics and Machine Learning for supporting our work through the Scholarly Travel Fund.

True \ Pred	en	es	ja	tr	zh
en	30.0	10.7	27.3	26.7	5.3
es	21.3	18.0	32.7	25.3	2.7
ja	9.3	4.7	60.0	20.0	6.0
tr	4.7	6.0	30.7	58.0	0.7
zh	22.0	8.7	26.0	34.0	9.3

Table 9: **Recall Heatmap (Row-normalized) for MMaDA.** True is the ground-truth label of the generated images and Pred is the prediction the classifier made. For each row (ground-truth), the numbers indicate the probability of each prediction.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-vl technical report, 2025. URL https://arxiv.org/abs/2511.21631.

Xiaokang Chen, Zhiyu Wu, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, and Chong Ruan. Janus-pro: Unified multimodal understanding and generation with data and model scaling. *arXiv preprint arXiv:2501.17811*, 2025.

Yufeng Cui, Honghao Chen, Haoge Deng, Xu Huang, Xinghang Li, Jirong Liu, Yang Liu, Zhuoyan Luo, Jinsheng Wang, Wenxuan Wang, et al. Emu3. 5: Native multimodal models are world learners. *arXiv preprint arXiv:2510.26583*, 2025.

Chaorui Deng, Deyao Zhu, Kunchang Li, Chenhui Gou, Feng Li, Zeyu Wang, Shu Zhong, Weihao Yu, Xiaonan Nie, Ziang Song, et al. Emerging properties in unified multimodal pretraining. *arXiv preprint arXiv:2505.14683*, 2025.

Felix Friedrich, Katharina Hämmerl, Patrick Schramowski, Manuel Brack, Jindrich Libovicky, Kristian Kersting, and Alexander Fraser. Multilingual text-to-image generation magnifies gender stereotypes and prompt engineering may not help you. *arXiv preprint arXiv:2401.16092*, 2024.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023. URL https://arxiv.org/abs/2304.02643.

Daiqing Li, Aleks Kamko, Ehsan Akhgari, Ali Sabet, Linmiao Xu, and Suhail Doshi. Playground v2. 5: Three insights towards enhancing aesthetic quality in text-to-image generation. *arXiv preprint arXiv:2402.17245*, 2024.

Zhuang Liu and Kaiming He. A decade's battle on dataset bias: Are we there yet?, 2025. URL https://arxiv.org/abs/2403.08632.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun

Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. Gpt-4 technical report, 2024. URL https://arxiv.org/abs/2303.08774.

Muhammed Saeed, Shaina Raza, Ashmal Vayani, Muhammad Abdul-Mageed, Ali Emami, and Shady Shehata. Beyond content: How grammatical gender shapes visual representation in text-to-image models. *arXiv preprint arXiv:2508.03199*, 2025.

Chuancheng Shi, Shangze Li, Shiming Guo, Simiao Xie, Wenhua Wu, Jingtong Dou, Chao Wu, Canran Xiao, Cong Wang, Zifeng Cheng, et al. Where culture fades: Revealing the cultural gap in text-to-image generation. *arXiv preprint arXiv:2511.17282*, 2025.

Mingjie Sun, Yida Yin, Zhiqiu Xu, J. Zico Kolter, and Zhuang Liu. Idiosyncrasies in large language models, 2025. URL https://arxiv.org/abs/2502.12150.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, Paul R. Barham, Tom Hennigan, Benjamin Lee, Fabio Viola, Malcolm Reynolds, Yuanzhong Xu, Ryan Doherty, Eli Collins, Clemens Meyer, Eliza Rutherford, Erica Moreira, Kareem Ayoub, Megha Goel, Jack Krawczyk, Cosmo Du, Ed Chi, Heng-Tze Cheng, Eric Ni, Purvi Shah, Patrick Kane, Betty Chan, Manaal Faruqui, Aliaksei Severyn, Hanzhao Lin, YaGuang Li, Yong Cheng, Abe Ittycheriah, Mahdis Mahdieh, Mia Chen, Pei Sun, Dustin Tran, Sumit Bagri, Balaji Lakshminarayanan, Jeremiah Liu, Andras Orban, Fabian Güra, Hao Zhou, Xinying Song, Aurelien Boffy, Harish Ganapathy, Steven Zheng, HyunJeong Choe, Ágoston Weisz, Tao Zhu, Yifeng Lu, Siddharth Gopal, Jarrod Kahn, Maciej Kula, Jeff Pitman, Rushin Shah, Emanuel Taropa, Majd Al Merey, Martin Baeuml, Zhifeng Chen, Laurent El Shafey, Yujing Zhang, Olcan Sercinoglu, George Tucker, Enrique Piqueras, Maxim Krikun, Iain Barr, Nikolay Savinov, Ivo Danihelka, Becca Roelofs, Anaïs White, Anders Andreassen, Tamara von Glehn, Lakshman Yagati, Mehran Kazemi, Lucas Gonzalez, Misha Khalman, Jakub Sygnowski, Alexandre Frechette, Charlotte Smith, Laura Culp, Lev Proleev, Yi Luan, Xi Chen, James Lottes, Nathan Schucher, Federico Lebron, Alban

Rrustemi, Natalie Clay, Phil Crone, Tomas Kocisky, Jeffrey Zhao, Bartek Perz, Dian Yu, Heidi Howard, Adam Bloniarz, Jack W. Rae, Han Lu, Laurent Sifre, Marcello Maggioni, Fred Alcober, Dan Garrette, Megan Barnes, Shantanu Thakoor, Jacob Austin, Gabriel Barth-Maron, William Wong, Rishabh Joshi, Rahma Chaabouni, Deeni Fatiha, Arun Ahuja, Gaurav Singh Tomar, Evan Senter, Martin Chadwick, Ilya Kornakov, Nithya Attaluri, Iñaki Iturrate, Ruibo Liu, Yunxuan Li, Sarah Cogan, Jeremy Chen, Chao Jia, Chenjie Gu, Qiao Zhang, Jordan Grimstad, Ale Jakse Hartman, Xavier Garcia, Thanumalayan Sankaranarayana Pillai, Jacob Devlin, Michael Laskin, Diego de Las Casas, Dasha Valter, Connie Tao, Lorenzo Blanco, Adrià Puigdomènech Badia, David Reitter, Mianna Chen, Jenny Brennan, Clara Rivera, Sergey Brin, Shariq Iqbal, Gabriela Surita, Jane Labanowski, Abhi Rao, Stephanie Winkler, Emilio Parisotto, Yiming Gu, Kate Olszewska, Ravi Addanki, Antoine Miech, Annie Louis, Denis Teplyashin, Geoff Brown, Elliot Catt, Jan Balaguer, Jackie Xiang, Pidong Wang, Zoe Ashwood, Anton Briukhov, Albert Webson, Sanjay Ganapathy, Smit Sanghavi, Ajay Kannan, Ming-Wei Chang, Axel Stjerngren, Josip Djolonga, Yuting Sun, Ankur Bapna, Matthew Aitchison, Pedram Pejman, Henryk Michalewski, Tianhe Yu, Cindy Wang, Juliette Love, Junwhan Ahn, Dawn Bloxwich, Kehang Han, Peter Humphreys, Thibault Sellam, James Bradbury, Varun Godbole, Sina Samangooei, Bogdan Damoc, Alex Kaskasoli, Sébastien M. R. Arnold, Vijay Vasudevan, Shubham Agrawal, Jason Riesa, Dmitry Lepikhin, Richard Tanburn, Srivatsan Srinivasan, Hyeontaek Lim, Sarah Hodkinson, Pranav Shyam, Johan Ferret, Steven Hand, Ankush Garg, Tom Le Paine, Jian Li, Yujia Li, Minh Giang, Alexander Neitz, Zaheer Abbas, Sarah York, Machel Reid, Elizabeth Cole, Aakanksha Chowdhery, Dipanjan Das, Dominika Rogozińska, Vitaliy Nikolaev, Pablo Sprechmann, Zachary Nado, Lukas Zilka, Flavien Prost, Luheng He, Marianne Monteiro, Gaurav Mishra, Chris Welty, Josh Newlan, Dawei Jia, Miltiadis Allamanis, Clara Huiyi Hu, Raoul de Liedekerke, Justin Gilmer, Carl Saroufim, Shruti Rijhwani, Shaobo Hou, Disha Shrivastava, Anirudh Baddepudi, Alex Goldin, Adnan Ozturel, Albin Cassirer, Yunhan Xu, Daniel Sohn, Devendra Sachan, Reinald Kim Amplayo, Craig Swanson, Dessie Petrova, Shashi Narayan, Arthur Guez, Siddhartha Brahma, Jessica Landon, Miteyan Patel, Ruizhe Zhao, Kevin Villela, Luyu Wang, Wenhao Jia, Matthew Rahtz, Mai Giménez, Legg Yeung, James Keeling, Petko Georgiev, Diana Mincu, Boxi Wu, Salem Haykal, Rachel Saputro, Kiran Vodrahalli, James Qin, Zeynep Cankara, Abhanshu Sharma, Nick Fernando, Will Hawkins, Behnam Neyshabur, Solomon Kim, Adrian Hutter, Priyanka Agrawal, Alex Castro-Ros, George van den Driessche, Tao Wang, Fan Yang, Shuo yiin Chang, Paul Komarek, Ross McIlroy, Mario Lučić, Guodong Zhang, Wael Farhan, Michael Sharman, Paul Natsev, Paul Michel, Yamini Bansal, Siyuan Qiao, Kris Cao, Siamak Shakeri, Christina Butterfield, Justin Chung, Paul Kishan Rubenstein, Shivani Agrawal, Arthur Mensch, Kedar Soparkar, Karel Lenc, Timothy Chung, Aedan Pope, Loren Maggiore, Jackie Kay, Priya Jhakra, Shibo Wang, Joshua Maynez, Mary Phuong, Taylor Tobin, Andrea Tacchetti, Maja Trebacz, Kevin Robinson, Yash Katariya, Sebastian Riedel, Paige Bailey, Kefan Xiao, Nimesh Ghelani, Lora Aroyo, Ambrose Slone, Neil Houlsby, Xuehan Xiong, Zhen Yang, Elena Gribovskaya, Jonas Adler, Mateo Wirth, Lisa Lee, Music Li, Thais Kagohara, Jay Pavagadhi, Sophie Bridgers, Anna Bortsova, Sanjay Ghemawat, Zafarali Ahmed, Tianqi Liu, Richard Powell, Vijay Bolina, Mariko Iinuma, Polina Zablotskaia, James Besley, Da-Woon Chung, Timothy Dozat, Ramona Comanescu, Xiance Si, Jeremy Greer, Guolong Su, Martin Polacek, Raphaël Lopez Kaufman, Simon Tokumine, Hexiang Hu, Elena Buchatskaya, Yingjie Miao, Mohamed Elhawaty, Aditya Siddhant, Nenad Tomasev, Jinwei Xing, Christina Greer, Helen Miller, Shereen Ashraf, Aurko Roy, Zizhao Zhang, Ada Ma, Angelos Filos, Milos Besta, Rory Blevins, Ted Klimenko, Chih-Kuan Yeh, Soravit Changpinyo, Jiaqi Mu, Oscar Chang, Mantas Pajarskas, Carrie Muir, Vered Cohen, Charline Le Lan, Krishna Haridasan, Amit Marathe, Steven Hansen, Sholto Douglas, Rajkumar Samuel, Mingqiu Wang, Sophia Austin, Chang Lan, Jiepu Jiang, Justin Chiu, Jaime Alonso Lorenzo, Lars Lowe Sjösund, Sébastien Cevey, Zach Gleicher, Thi Avrahami, Anudhyan Boral, Hansa Srinivasan, Vittorio Selo, Rhys May, Konstantinos Aisopos, Léonard Hussenot, Livio Baldini Soares, Kate Baumli, Michael B. Chang, Adrià Recasens, Ben Caine, Alexander Pritzel, Filip Pavetic, Fabio Pardo, Anita Gergely, Justin Frye, Vinay Ramasesh, Dan Horgan, Kartikeya Badola, Nora Kassner, Subhrajit Roy, Ethan Dyer, Víctor Campos Campos, Alex Tomala, Yunhao Tang, Dalia El Badawy, Elspeth White, Basil Mustafa, Oran Lang, Abhishek Jindal, Sharad Vikram, Zhitao Gong, Sergi Caelles, Ross Hemsley, Gregory Thornton, Fangxiaoyu Feng, Wojciech Stokowiec, Ce Zheng, Phoebe Thacker, Çağlar Ünlü, Zhishuai Zhang, Mohammad Saleh, James Svensson, Max Bileschi, Piyush Patil, Ankesh Anand, Roman Ring, Katerina Tsihlas, Arpi Vezer, Marco Selvi, Toby Shevlane, Mikel Rodriguez, Tom Kwiatkowski, Samira Daruki, Keran Rong, Allan Dafoe, Nicholas FitzGerald, Keren Gu-Lemberg, Mina Khan, Lisa Anne Hendricks, Marie Pel-

lat, Vladimir Feinberg, James Cobon-Kerr, Tara Sainath, Maribeth Rauh, Sayed Hadi Hashemi, Richard Ives, Yana Hasson, Eric Noland, Yuan Cao, Nathan Byrd, Le Hou, Qingze Wang, Thibault Sottiaux, Michela Paganini, Jean-Baptiste Lespiau, Alexandre Moufarek, Samer Hassan, Kaushik Shivakumar, Joost van Amersfoort, Amol Mandhane, Pratik Joshi, Anirudh Goyal, Matthew Tung, Andrew Brock, Hannah Sheahan, Vedant Misra, Cheng Li, Nemanja Rakićević, Mostafa Dehghani, Fangyu Liu, Sid Mittal, Junhyuk Oh, Seb Noury, Eren Sezener, Fantine Huot, Matthew Lamm, Nicola De Cao, Charlie Chen, Sidharth Mudgal, Romina Stella, Kevin Brooks, Gautam Vasudevan, Chenxi Liu, Mainak Chain, Nivedita Melinkeri, Aaron Cohen, Venus Wang, Kristie Seymore, Sergey Zubkov, Rahul Goel, Summer Yue, Sai Krishnakumaran, Brian Albert, Nate Hurley, Motoki Sano, Anhad Mohananey, Jonah Joughin, Egor Filonov, Tomasz Kepa, Yomna Eldawy, Jiawern Lim, Rahul Rishi, Shirin Badiezadegan, Taylor Bos, Jerry Chang, Sanil Jain, Sri Gayatri Sundara Padmanabhan, Subha Puttagunta, Kalpesh Krishna, Leslie Baker, Norbert Kalb, Vamsi Bedapudi, Adam Kurzrok, Shuntong Lei, Anthony Yu, Oren Litvin, Xiang Zhou, Zhichun Wu, Sam Sobell, Andrea Siciliano, Alan Papir, Robby Neale, Jonas Bragagnolo, Tej Toor, Tina Chen, Valentin Anklin, Feiran Wang, Richie Feng, Milad Gholami, Kevin Ling, Lijuan Liu, Jules Walter, Hamid Moghaddam, Arun Kishore, Jakub Adamek, Tyler Mercado, Jonathan Mallinson, Siddhinita Wandekar, Stephen Cagle, Eran Ofek, Guillermo Garrido, Clemens Lombriser, Maksim Mukha, Botu Sun, Hafeezul Rahman Mohammad, Josip Matak, Yadi Qian, Vikas Peswani, Pawel Janus, Quan Yuan, Leif Schelin, Oana David, Ankur Garg, Yifan He, Oleksii Duzhyi, Anton Älgmyr, Timothée Lottaz, Qi Li, Vikas Yadav, Luyao Xu, Alex Chinien, Rakesh Shivanna, Aleksandr Chuklin, Josie Li, Carrie Spadine, Travis Wolfe, Kareem Mohamed, Subhabrata Das, Zihang Dai, Kyle He, Daniel von Dincklage, Shyam Upadhyay, Akanksha Maurya, Luyan Chi, Sebastian Krause, Khalid Salama, Pam G Rabinovitch, Pavan Kumar Reddy M, Aarush Selvan, Mikhail Dektiarev, Golnaz Ghiasi, Erdem Guven, Himanshu Gupta, Boyi Liu, Deepak Sharma, Idan Heimlich Shtacher, Shachi Paul, Oscar Akerlund, François-Xavier Aubet, Terry Huang, Chen Zhu, Eric Zhu, Elico Teixeira, Matthew Fritze, Francesco Bertolini, Liana-Eleonora Marinescu, Martin Bölle, Dominik Paulus, Khyatti Gupta, Tejasi Latkar, Max Chang, Jason Sanders, Roopa Wilson, Xuewei Wu, Yi-Xuan Tan, Lam Nguyen Thiet, Tulsee Doshi, Sid Lall, Swaroop Mishra, Wanming Chen, Thang Luong, Seth Benjamin, Jasmine Lee, Ewa Andrejczuk, Dominik Rabiej, Vipul Ranjan, Krzysztof Styrc, Pengcheng Yin, Jon Simon, Malcolm Rose Harriott, Mudit Bansal, Alexei Robsky, Geoff Bacon, David Greene, Daniil Mirylenka, Chen Zhou, Obaid Sarvana, Abhimanyu Goyal, Samuel Andermatt, Patrick Siegler, Ben Horn, Assaf Israel, Francesco Pongetti, Chih-Wei "Louis" Chen, Marco Selvatici, Pedro Silva, Kathie Wang, Jackson Tolins, Kelvin Guu, Roey Yogev, Xiaochen Cai, Alessandro Agostini, Maulik Shah, Hung Nguyen, Noah Ó Donnaile, Sébastien Pereira, Linda Friso, Adam Stambler, Adam Kurzrok, Chenkai Kuang, Yan Romanikhin, Mark Geller, ZJ Yan, Kane Jang, Cheng-Chun Lee, Wojciech Fica, Eric Malmi, Qijun Tan, Dan Banica, Daniel Balle, Ryan Pham, Yanping Huang, Diana Avram, Hongzhi Shi, Jasjot Singh, Chris Hidey, Niharika Ahuja, Pranab Saxena, Dan Dooley, Srividya Pranavi Potharaju, Eileen O'Neill, Anand Gokulchandran, Ryan Foley, Kai Zhao, Mike Dusenberry, Yuan Liu, Pulkit Mehta, Ragha Kotikalapudi, Chalence Safranek-Shrader, Andrew Goodman, Joshua Kessinger, Eran Globen, Prateek Kolhar, Chris Gorgolewski, Ali Ibrahim, Yang Song, Ali Eichenbaum, Thomas Brovelli, Sahitya Potluri, Preethi Lahoti, Cip Baetu, Ali Ghorbani, Charles Chen, Andy Crawford, Shalini Pal, Mukund Sridhar, Petru Gurita, Asier Mujika, Igor Petrovski, Pierre-Louis Cedoz, Chenmei Li, Shiyuan Chen, Niccolò Dal Santo, Siddharth Goyal, Jitesh Punjabi, Karthik Kappaganthu, Chester Kwak, Pallavi LV, Sarmishta Velury, Himadri Choudhury, Jamie Hall, Premal Shah, Ricardo Figueira, Matt Thomas, Minjie Lu, Ting Zhou, Chintu Kumar, Thomas Jurdi, Sharat Chikkerur, Yenai Ma, Adams Yu, Soo Kwak, Victor Ähdel, Sujeevan Rajayogam, Travis Choma, Fei Liu, Aditya Barua, Colin Ji, Ji Ho Park, Vincent Hellendoorn, Alex Bailey, Taylan Bilal, Huanjie Zhou, Mehrdad Khatir, Charles Sutton, Wojciech Rzadkowski, Fiona Macintosh, Roopali Vij, Konstantin Shagin, Paul Medina, Chen Liang, Jinjing Zhou, Pararth Shah, Yingying Bi, Attila Dankovics, Shipra Banga, Sabine Lehmann, Marissa Bredesen, Zifan Lin, John Eric Hoffmann, Jonathan Lai, Raynald Chung, Kai Yang, Nihal Balani, Arthur Bražinskas, Andrei Sozanschi, Matthew Hayes, Héctor Fernández Alcalde, Peter Makarov, Will Chen, Antonio Stella, Liselotte Snijders, Michael Mandl, Ante Kärrman, Paweł Nowak, Xinyi Wu, Alex Dyck, Krishnan Vaidyanathan, Raghavender R, Jessica Mallet, Mitch Rudominer, Eric Johnston, Sushil Mittal, Akhil Udathu, Janara Christensen, Vishal Verma, Zach Irving, Andreas Santucci, Gamaleldin Elsayed, Elnaz Davoodi, Marin Georgiev, Ian Tenney, Nan Hua, Geoffrey Cideron, Edouard Leurent, Mahmoud Alnahlawi, Ionut Georgescu, Nan Wei, Ivy Zheng, Dylan Scandinaro, Heinrich Jiang, Jasper Snoek, Mukund Sundararajan,

Xuezhi Wang, Zack Ontiveros, Itay Karo, Jeremy Cole, Vinu Rajashekhar, Lara Tumeh, Eyal Ben-David, Rishub Jain, Jonathan Uesato, Romina Datta, Oskar Bunyan, Shimu Wu, John Zhang, Piotr Stanczyk, Ye Zhang, David Steiner, Subhajit Naskar, Michael Azzam, Matthew Johnson, Adam Paszke, Chung-Cheng Chiu, Jaume Sanchez Elias, Afroz Mohiuddin, Faizan Muhammad, Jin Miao, Andrew Lee, Nino Vieillard, Jane Park, Jiageng Zhang, Jeff Stanway, Drew Garmon, Abhijit Karmarkar, Zhe Dong, Jong Lee, Aviral Kumar, Luowei Zhou, Jonathan Evens, William Isaac, Geoffrey Irving, Edward Loper, Michael Fink, Isha Arkatkar, Nanxin Chen, Izhak Shafran, Ivan Petrychenko, Zhe Chen, Johnson Jia, Anselm Levskaya, Zhenkai Zhu, Peter Grabowski, Yu Mao, Alberto Magni, Kaisheng Yao, Javier Snaider, Norman Casagrande, Evan Palmer, Paul Suganthan, Alfonso Castaño, Irene Giannoumis, Wooyeol Kim, Mikołaj Rybiński, Ashwin Sreevatsa, Jennifer Prendki, David Soergel, Adrian Goedeckemeyer, Willi Gierke, Mohsen Jafari, Meenu Gaba, Jeremy Wiesner, Diana Gage Wright, Yawen Wei, Harsha Vashisht, Yana Kulizhskaya, Jay Hoover, Maigo Le, Lu Li, Chimezie Iwuanyanwu, Lu Liu, Kevin Ramirez, Andrey Khorlin, Albert Cui, Tian LIN, Marcus Wu, Ricardo Aguilar, Keith Pallo, Abhishek Chakladar, Ginger Perng, Elena Allica Abellan, Mingyang Zhang, Ishita Dasgupta, Nate Kushman, Ivo Penchev, Alena Repina, Xihui Wu, Tom van der Weide, Priya Ponnapalli, Caroline Kaplan, Jiri Simsa, Shuangfeng Li, Olivier Dousse, Fan Yang, Jeff Piper, Nathan Ie, Rama Pasumarthi, Nathan Lintz, Anitha Vijayakumar, Daniel Andor, Pedro Valenzuela, Minnie Lui, Cosmin Paduraru, Daiyi Peng, Katherine Lee, Shuyuan Zhang, Somer Greene, Duc Dung Nguyen, Paula Kurylowicz, Cassidy Hardin, Lucas Dixon, Lili Janzer, Kiam Choo, Ziqiang Feng, Biao Zhang, Achintya Singhal, Dayou Du, Dan McKinnon, Natasha Antropova, Tolga Bolukbasi, Orgad Keller, David Reid, Daniel Finchelstein, Maria Abi Raad, Remi Crocker, Peter Hawkins, Robert Dadashi, Colin Gaffney, Ken Franko, Anna Bulanova, Rémi Leblond, Shirley Chung, Harry Askham, Luis C. Cobo, Kelvin Xu, Felix Fischer, Jun Xu, Christina Sorokin, Chris Alberti, Chu-Cheng Lin, Colin Evans, Alek Dimitriev, Hannah Forbes, Dylan Banarse, Zora Tung, Mark Omernick, Colton Bishop, Rachel Sterneck, Rohan Jain, Jiawei Xia, Ehsan Amid, Francesco Piccinno, Xingyu Wang, Praseem Banzal, Daniel J. Mankowitz, Alex Polozov, Victoria Krakovna, Sasha Brown, MohammadHossein Bateni, Dennis Duan, Vlad Firoiu, Meghana Thotakuri, Tom Natan, Matthieu Geist, Ser tan Girgin, Hui Li, Jiayu Ye, Ofir Roval, Reiko Tojo, Michael Kwong, James Lee-Thorp, Christopher Yew, Danila Sinopalnikov, Sabela Ramos, John Mellor, Abhishek Sharma, Kathy Wu, David Miller, Nicolas Sonnerat, Denis Vnukov, Rory Greig, Jennifer Beattie, Emily Caveness, Libin Bai, Julian Eisenschlos, Alex Korchemniy, Tomy Tsai, Mimi Jasarevic, Weize Kong, Phuong Dao, Zeyu Zheng, Frederick Liu, Fan Yang, Rui Zhu, Tian Huey Teh, Jason Sanmiya, Evgeny Gladchenko, Nejc Trdin, Daniel Toyama, Evan Rosen, Sasan Tavakkol, Linting Xue, Chen Elkind, Oliver Woodman, John Carpenter, George Papamakarios, Rupert Kemp, Sushant Kafle, Tanya Grunina, Rishika Sinha, Alice Talbert, Diane Wu, Denese Owusu-Afriyie, Cosmo Du, Chloe Thornton, Jordi Pont-Tuset, Pradyumna Narayana, Jing Li, Saaber Fatehi, John Wieting, Omar Ajmeri, Benigno Uria, Yeongil Ko, Laura Knight, Amélie Héliou, Ning Niu, Shane Gu, Chenxi Pang, Yeqing Li, Nir Levine, Ariel Stolovich, Rebeca Santamaria-Fernandez, Sonam Goenka, Wenny Yustalim, Robin Strudel, Ali Elqursh, Charlie Deck, Hyo Lee, Zonglin Li, Kyle Levin, Raphael Hoffmann, Dan Holtmann-Rice, Olivier Bachem, Sho Arora, Christy Koh, Soheil Hassas Yeganeh, Siim Põder, Mukarram Tariq, Yanhua Sun, Lucian Ionita, Mojtaba Seyedhosseini, Pouya Tafti, Zhiyu Liu, Anmol Gulati, Jasmine Liu, Xinyu Ye, Bart Chrzaszcz, Lily Wang, Nikhil Sethi, Tianrun Li, Ben Brown, Shreya Singh, Wei Fan, Aaron Parisi, Joe Stanton, Vinod Koverkathu, Christopher A. Choquette-Choo, Yunjie Li, TJ Lu, Abe Ittycheriah, Prakash Shroff, Mani Varadarajan, Sanaz Bahargam, Rob Willoughby, David Gaddy, Guillaume Desjardins, Marco Cornero, Brona Robenek, Bhavishya Mittal, Ben Albrecht, Ashish Shenoy, Fedor Moiseev, Henrik Jacobsson, Alireza Ghaffarkhah, Morgane Rivière, Alanna Walton, Clément Crepy, Alicia Parrish, Zongwei Zhou, Clement Farabet, Carey Radebaugh, Praveen Srinivasan, Claudia van der Salm, Andreas Fidjeland, Salvatore Scellato, Eri Latorre-Chimoto, Hanna Klimczak-Plucińska, David Bridson, Dario de Cesare, Tom Hudson, Piermaria Mendolicchio, Lexi Walker, Alex Morris, Matthew Mauger, Alexey Guseynov, Alison Reid, Seth Odoom, Lucia Loher, Victor Cotruta, Madhavi Yenugula, Dominik Grewe, Anastasia Petrushkina, Tom Duerig, Antonio Sanchez, Steve Yadlowsky, Amy Shen, Amir Globerson, Lynette Webb, Sahil Dua, Dong Li, Surya Bhupatiraju, Dan Hurt, Haroon Qureshi, Ananth Agarwal, Tomer Shani, Matan Eyal, Anuj Khare, Shreyas Rammohan Belle, Lei Wang, Chetan Tekur, Mihir Sanjay Kale, Jinliang Wei, Ruoxin Sang, Brennan Saeta, Tyler Liechty, Yi Sun, Yao Zhao, Stephan Lee, Pandu Nayak, Doug Fritz, Manish Reddy Vuyyuru, John Aslanides, Nidhi Vyas, Martin Wicke, Xiao Ma, Evgenii Eltyshev, Nina Martin, Hardie Cate, James Manyika, Keyvan Amiri, Yelin Kim,

Xi Xiong, Kai Kang, Florian Luisier, Nilesh Tripuraneni, David Madras, Mandy Guo, Austin Waters, Oliver Wang, Joshua Ainslie, Jason Baldridge, Han Zhang, Garima Pruthi, Jakob Bauer, Feng Yang, Riham Mansour, Jason Gelman, Yang Xu, George Polovets, Ji Liu, Honglong Cai, Warren Chen, XiangHai Sheng, Emily Xue, Sherjil Ozair, Christof Angermueller, Xiaowei Li, Anoop Sinha, Weiren Wang, Julia Wiesinger, Emmanouil Koukoumidis, Yuan Tian, Anand Iyer, Madhu Gurumurthy, Mark Goldenson, Parashar Shah, MK Blake, Hongkun Yu, Anthony Urbanowicz, Jennimaria Palomaki, Chrisantha Fernando, Ken Durden, Harsh Mehta, Nikola Momchev, Elahe Rahimtoroghi, Maria Georgaki, Amit Raul, Sebastian Ruder, Morgan Redshaw, Jinhyuk Lee, Denny Zhou, Komal Jalan, Dinghua Li, Blake Hechtman, Parker Schuh, Milad Nasr, Kieran Milan, Vladimir Mikulik, Juliana Franco, Tim Green, Nam Nguyen, Joe Kelley, Aroma Mahendru, Andrea Hu, Joshua Howland, Ben Vargas, Jeffrey Hui, Kshitij Bansal, Vikram Rao, Rakesh Ghiya, Emma Wang, Ke Ye, Jean Michel Sarr, Melanie Moranski Preston, Madeleine Elish, Steve Li, Aakash Kaku, Jigar Gupta, Ice Pasupat, Da-Cheng Juan, Milan Someswar, Tejvi M., Xinyun Chen, Aida Amini, Alex Fabrikant, Eric Chu, Xuanyi Dong, Amruta Muthal, Senaka Buthpitiya, Sarthak Jauhari, Nan Hua, Urvashi Khandelwal, Ayal Hitron, Jie Ren, Larissa Rinaldi, Shahar Drath, Avigail Dabush, Nan-Jiang Jiang, Harshal Godhia, Uli Sachs, Anthony Chen, Yicheng Fan, Hagai Taitelbaum, Hila Noga, Zhuyun Dai, James Wang, Chen Liang, Jenny Hamer, Chun-Sung Ferng, Chenel Elkind, Aviel Atias, Paulina Lee, Vít Listík, Mathias Carlen, Jan van de Kerkhof, Marcin Pikus, Krunoslav Zaher, Paul Müller, Sasha Zykova, Richard Stefanec, Vitaly Gatsko, Christoph Hirnschall, Ashwin Sethi, Xingyu Federico Xu, Chetan Ahuja, Beth Tsai, Anca Stefanoiu, Bo Feng, Keshav Dhandhania, Manish Katyal, Akshay Gupta, Atharva Parulekar, Divya Pitta, Jing Zhao, Vivaan Bhatia, Yashodha Bhavnani, Omar Alhadlaq, Xiaolin Li, Peter Danenberg, Dennis Tu, Alex Pine, Vera Filippova, Abhipso Ghosh, Ben Limonchik, Bhargava Urala, Chaitanya Krishna Lanka, Derik Clive, Yi Sun, Edward Li, Hao Wu, Kevin Hongtongsak, Ianna Li, Kalind Thakkar, Kuanysh Omarov, Kushal Majmundar, Michael Alverson, Michael Kucharski, Mohak Patel, Mudit Jain, Maksim Zabelin, Paolo Pelagatti, Rohan Kohli, Saurabh Kumar, Joseph Kim, Swetha Sankar, Vineet Shah, Lakshmi Ramachandruni, Xiangkai Zeng, Ben Bariach, Laura Weidinger, Tu Vu, Alek Andreev, Antoine He, Kevin Hui, Sheleem Kashem, Amar Subramanya, Sissie Hsiao, Demis Hassabis, Koray Kavukcuoglu, Adam Sadovsky, Quoc Le, Trevor Strohman, Yonghui Wu, Slav Petrov, Jeffrey Dean, and Oriol Vinyals. Gemini: A family of highly capable multimodal models, 2025. URL https://arxiv.org/abs/2312.11805.

Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pp. 1521–1528. IEEE, 2011.

Jinheng Xie, Zhenheng Yang, and Mike Zheng Shou. Show-o2: Improved native unified multimodal models. *arXiv preprint arXiv:2506.15564*, 2025.

Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Detecting origin attribution for text-to-image diffusion models, 2025. URL https://arxiv.org/abs/2403.19653.

Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2, 2024. URL https://arxiv.org/abs/2406.09414.

Ling Yang, Ye Tian, Bowen Li, Xinchen Zhang, Ke Shen, Yunhai Tong, and Mengdi Wang. Mmada: Multimodal large diffusion language models. *arXiv preprint arXiv:2505.15809*, 2025.

Boya Zeng, Yida Yin, and Zhuang Liu. Understanding bias in large-scale visual datasets, 2024. URL https://arxiv.org/abs/2412.01876.