






Navigating Text-to-Image Generative Bias across Indic Languages (Supplementary)

Surbhi Mittal¹, Arnav Sudan¹, Mayank Vatsa¹, Richa Singh¹,
Tamar Glaser², and Tal Hassner³

¹ Department of CSE, IIT Jodhpur, Rajasthan, India

² Meta, Menlo Park, California, USA

³ Weir P.B.C., Alameda, California, USA

1 Introduction

The contributions of this work are summarized below.

- A holistic benchmark for identifying and quantitatively evaluating TTI models for languages other than English.
- Quantifying the presence of correctness and representation-related biases in 30 Indic languages across 4 TTI models through novel metrics.
- Qualitative analysis of cultural aspects in the generated images across Indic languages.

2 Benchmark Design

In this section, we provide additional details about the benchmark design.

2.1 Indic Languages and Prompts

In this research, we introduce the IndicTTI benchmark where we study the performance of popular text-to-image (TTI) models in 30 languages. While the benchmark may be extended for any number of languages, we select 30 Indic languages, which are written in 10 different scripts and have roots in multiple language families. Detailed information about the different languages is provided in Table 1.

For prompts, we utilize the COCO-NLLB dataset [5,6], which contains image-caption pairs for over 500K images, along with captions translated into 200 languages. We sample 1000 diverse image-caption pairs from the dataset. In order to avoid prompts with proper nouns, such as names of celebrities and/or brand names, we filtered the dataset to remove any captions that contained capitalized words in the middle of the sentence. From the filtered dataset, we randomly subsample 1000 captions. Next, for a diverse selection, we computed sentence-level embeddings for the 1000 prompts using a SentenceFormer model and calculated the average similarity between any two prompts. This experiment was repeated for 1000 iterations, and the subset with the lowest sentence

Table 1: The language family, script, language subfamilies, and number of native speakers for the 30 Indic languages in the IndicTTI benchmark. * represents the non-availability of official reports regarding the statistics.

Language Code	Name	Family	Script	Sub-family	#Native Speakers
asm_Beng	Assamese	Indo-Aryan	Bengali	Eastern Indo-Aryan	15.3M
awa_Deva	Awadhi	Indo-Aryan	Devanagari	Northern Indo-Aryan	2.52M
ben_Beng	Bengali	Indo-Aryan	Bengali	Eastern Indo-Aryan	97.2M
bho_Deva	Bhojpuri	Indo-Aryan	Devanagari	Northern Indo-Aryan	*
brx_Deva	Bodo	Sino-Tibetan	Devanagari	Boroic	1.4M
doi_Deva	Dogri	Indo-Aryan	Devanagari	Northern Indo-Aryan	2.5M
gom_Deva	Konkani	Indo-Aryan	Devanagari	Southern Indo-Aryan	2.2M
guj_Gujr	Gujarati	Indo-Aryan	Gujarati	Western Indo-Aryan	55.4M
hin_Deva	Hindi	Indo-Aryan	Devanagari	Central Indo-Aryan	528.3M
hne_Deva	Chhattisgarhi	Indo-Aryan	Devanagari	Northern Indo-Aryan	13M
kan_Knda	Kannada	Dravidian	Kannada	South Dravidian	43.7M
kas_Arab	Kashmiri	Indo-Aryan	Perso-Arabic	Northern Indo-Aryan	6.7M
kas_Deva			Devanagari		*
mag_Deva	Magahi	Indo-Aryan	Devanagari	Indo-Aryan	
mai_Deva	Maithili	Indo-Aryan	Devanagari	Eastern Indo-Aryan	13.5M
mai_Mlym	Malayalam	Dravidian	Malayalam	Southern Dravidian	34.8M
mar_Deva	Marathi	Indo-Aryan	Devanagari	Southern Indo-Aryan	83.0M
mni_Beng			Bengali	Central	
mni_Mtei	Manipuri	Sino-Tibetan	Meitei	Tibeto-Burman	1.7M
npi_Deva	Nepali	Indo-Aryan	Devanagari	Northern Indo-Aryan	2.9M
ory_Orya	Odia	Indo-Aryan	Odia	Eastern Indo-Aryan	37.5M
pan_Guru	Punjabi	Indo-Aryan	Gurmukhi	North Western Indo-Aryan	33.1M
san_Deva	Sanskrit	Indo-Aryan	Devanagari	Indo-Aryan	0.02M
sat_Oick	Santali	Austroasiatic	Oi Chiki	Munda	7.3M
sin_Sinh	Sinhala	Indo-Aryan	Sinhala	Indo-Aryan	*
snd_Arab			Arabic		
snd_Deva	Sindhi	Indo-Aryan	Devanagari	North Western Indo-Aryan	2.7M
tam_Taml	Tamil	Dravidian	Tamil	South Dravidian	69.0M
tel_Telu	Telugu	Dravidian	Telugu	South Central Dravidian	81.1M
urd_Arab	Urdu	Indo-Aryan	Urdu	Central Indo-Aryan	50.7M

similarity between prompts was selected. The subsets of 200 and 50 prompts for Midjourney and Dalle3, respectively, were chosen randomly from the selected subset of 1000 prompts.

2.2 TTI Models and Generated Images

We utilize four different text-to-image models for the benchmark. For **open-source models**, we use the Stable Diffusion and AltDiffusion Models (without safety filters). In the main paper, we report results on the m2 variant of AltDiffusion trained on the English and Chinese languages. Extended results, including the m9 variant trained on the English, Chinese, Spanish, French, Russian, Japanese, Korean, Arabic, and Italian languages, are reported in the supplementary. The extended results exhibit a similar pattern to the m2 variant. Detailed results are provided in the subsequent sections. Unless specified otherwise, the AltDiffusion model refers to the m2 variant in this work. For **API-based models**, we use Dalle3 [4] and Midjourney [1], which are both paid. Stable Diffusion and AltDiffusion models generate images of size 512 x 512, whereas Midjourney and Dalle3 generate images of size 1024 x 1024.

2.3 Quality of Translated Captions

Our prompts were primarily sourced from COCO-NLLB or translated from English using the IndicTrans2 model. As specified in the Limitations section of the main paper, this suggests that the translation quality may have influenced the

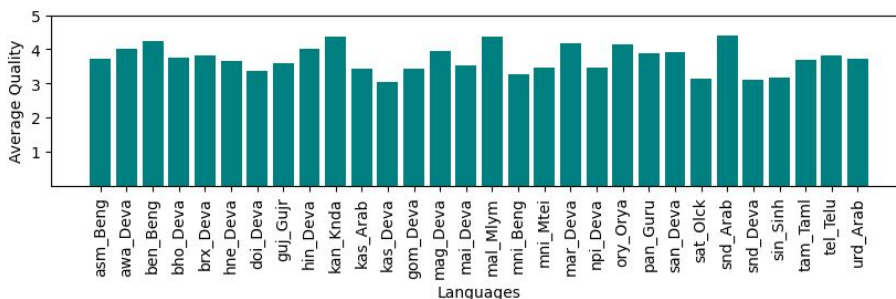


Fig. 1: Plot showcasing the quality of captions per language.

performance of image generation with Indic prompts. To evaluate the translation quality, we conducted a user study on a common set of 40 prompts across 30 languages, validated by 65 annotators with an average proficiency of 4.78 out of 5 in at least one Indic language. We followed the XSTS protocol [2], also used to assess the translation quality of the NLLB model [3]. On a scale from 1 (worst) to 5 (best), the average translation quality for all languages is reported in Fig. 1, indicating that the sentences are mostly equivalent or paraphrases of each other, according to the XSTS protocol, and thus suitable for image generation. In this experiment, the inter-rater agreement, measured through percent agreement, was 74.8%. For languages such as kas_Deva, snd_Deva, and sin_Sinh, the translation quality of certain prompts was observed to be less than 3 on the XSTS scale.

3 Implementation Details

In this section, we discuss the implementation details involved in the creation of the IndicTTI benchmark as well as its evaluation.

For generation using open-source TTI systems of Stable Diffusion⁴ and Alt-Diffusion⁵, we utilized the models available at HuggingFace. Dalle3 API was accessed through a Python script whereas the generation for Midjourney was done using Discord. For Stable Diffusion and AltDiffusion, the generation was done using an NVIDIA DGX Station consisting of 4 NVIDIA V100 GPU with 32 GB VRAM each, using a batch size of 8. The generation was repeated using 4 seeds, providing 4 images for every prompt. The DDIM Scheduler was used for inferencing with 50 steps.

For evaluation, all experiments are conducted on LINUX-based systems using Python-based libraries, and specifically, the PyTorch library is used. To extract rich semantic text and image features for the evaluation metrics, we utilize

⁴ <https://huggingface.co/runwayml/stable-diffusion-v1-5>

⁵ <https://huggingface.co/BAAI/AltDiffusion>, <https://huggingface.co/BAAI/AltDiffusion-m9>

various modules of the BLIP-2 model⁶. We utilize the image encoder of the BLIP-2 model as the image feature extractor f for computing the CLGC, IGC, SCAL, SCWL, and DWL metrics. Additionally, for the CLGC metric, we utilize the image-captioning capabilities of BLIP-2 captioner c to generate captions for generated images, and for extracting rich textual features, the SentenceFormer model⁷. For the LGC metric, we require image-text features that are extracted from BLIP-2 using the *LAVIS*⁸ library. The similarity function ϕ is computed using the cosine similarity. The code for generation, as well as evaluation, can be accessed through <https://iab-rubric.org/resources/other-databases/indictti>.

4 Benchmark Results and Analysis

In this section, we report extended results on the common set and complete set of prompts. The observations are consistent with those reported in the main paper on the common subset.

In the **correctness-based metrics**, the CLGC, IGC, and LGC metrics over the common set of prompts are reported in Tables 2, 3, and 4, respectively. Similarly, the results for the three metrics on the complete set of prompts are reported in Tables 5, 6, and 7. As observed in the main paper, across all the metrics, Dalle3 outperforms all other models when evaluated for Indic languages with a significant margin. Other models struggle to generalize on the Indic languages with AltDiffusion m9 performing the best among the other three models. This behavior is a result of increased generalizability due to the multilingual training of the model in 9 languages. On the other hand, while all models perform similarly for the English language, AltDiffusion m9 performs the worst, possibly due to catastrophic forgetting when trained for multilinguality.

In **representation-based metrics**, we evaluate the SCWL and DWL metrics on the complete set and observe that they follow the same patterns as their performance on the common set (refer Fig. 2). It is also observed that while AltDiffusion m9 has high distinctiveness across the concepts it generates, it provides poor self-consistency within the language, highlighting its instability and a tendency to seemingly generate diversely with or without relevance to the prompt. For the SCAL metric, the value obtained for AltDiffusion m9 on the common set of prompts comes out to be 21.47%, which is lower than the overall *SCAL metric* for the Stable Diffusion, AltDiffusion, Midjourney, and Dalle3 models is 25.44%, 23.73%, 26.75%, and 29.90%, respectively. This indicates an overall low consistency of AltDiffusion m9 in generating concepts across different languages.

⁶ <https://huggingface.co/Salesforce/blip2-opt-2.7b>

⁷ <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>

⁸ <https://github.com/salesforce/LAVIS/>

Table 2: Cyclic Language Grounded Correctness (CLGC) (%) across the different Indic languages in IndicTTI on the common set of prompts.

Model	en	asm	ben	guj	hin	kan	mal	mar	mnj	npi	ory	pan	san	snd	tam	tel
	(Beng)	(Beng)	(Gurj)	(Deva)	(Knda)	(Mlym)	(Arab)	(Beng)	(Beng)	(Orya)	(Guru)	(Deva)	(Arab)	(Arab)	(Tamil)	(Tulu)
Stable Diffusion	64.53	7.29	6.39	7.24	7.13	6.91	7.53	6.50	8.03	8.36	8.78	8.14	11.29	7.30	7.65	
Alt Diffusion	58.16	14.65	16.70	15.97	16.79	22.98	16.08	16.50	12.17	17.97	18.82	17.64	17.99	14.82	18.10	14.50
Alt Diffusion m9	53.19	18.02	27.08	26.50	29.78	26.66	27.04	28.21	12.39	31.42	27.70	27.14	21.50	18.97	27.81	24.34
MidJourney	64.31	12.99	11.89	12.49	11.31	12.12	10.92	11.99	9.80	11.89	12.56	9.32	12.19	13.49	12.84	12.16
Dalle3	65.70	53.36	60.73	57.25	64.16	60.13	57.68	61.84	35.24	59.90	53.96	62.67	51.22	45.19	53.98	55.26
Model	en	urd	kas	hne	mal	awa	hne	mag	sin	brx	dot	gom	sat	snd	mnj	(Mfcl)
	(Arab)	(Arab)	(Arab)	(Deva)	(Deva)	(Deva)	(Deva)	(Sinh)	(Sinh)	(Deva)	(Deva)	(Deva)	(Deva)	(Oick)	(Deva)	(Mfcl)
Stable Diffusion	64.53	10.08	6.29	7.52	7.13	10.11	7.12	6.25	7.95	6.26	7.78	10.51	7.59	10.17		
Alt Diffusion	58.16	17.68	15.96	15.46	16.93	16.32	15.41	16.74	16.80	13.04	11.98	13.50	15.82	14.13	13.37	15.25
Alt Diffusion m9	53.19	26.29	17.33	26.03	23.69	28.06	24.48	29.98	28.99	26.24	12.29	21.57	20.89	12.97	23.71	13.36
MidJourney	64.31	15.38	12.72	12.37	11.92	13.04	11.47	12.06	12.53	11.96	12.34	11.77	12.59	11.36	12.35	9.72
Dalle3	65.70	64.02	47.45	55.51	58.40	57.37	60.75	59.18	60.42	49.03	28.87	59.84	50.18	10.93	56.78	11.19

Table 3: Image-Grounded Correctness (IGC) (%) across the different Indic languages in IndicTTI on the common set of prompts.

Model	en	asm	ben	guj	hin	kan	mal	mar	mnj	npi	ory	pan	san	snd	tam	tel
	(Beng)	(Beng)	(Gurj)	(Deva)	(Knda)	(Mlym)	(Arab)	(Beng)	(Beng)	(Orya)	(Guru)	(Deva)	(Arab)	(Arab)	(Tamil)	(Tulu)
Stable Diffusion	51.21	24.19	23.84	23.23	23.53	23.24	22.96	23.53	23.98	23.76	22.97	23.11	23.06	22.08	22.61	
Alt Diffusion	46.58	24.23	23.08	25.18	25.43	27.74	24.27	25.42	22.99	23.98	26.27	23.48	23.68	24.68	22.19	
Alt Diffusion m9	42.43	26.50	28.39	20.46	28.94	29.74	29.59	28.56	23.29	30.67	29.35	28.76	26.94	25.94	28.62	
MidJourney	48.94	22.69	22.66	23.06	22.50	22.51	22.04	22.75	21.97	22.29	22.82	23.03	22.81	23.34	22.09	
Dalle3	48.92	40.02	44.22	43.75	44.86	42.74	43.12	46.10	30.91	43.14	40.00	43.06	37.66	36.30	41.25	41.39
Model	en	urd	kas	hne	mal	awa	hne	mag	sin	brx	dot	gom	sat	snd	mnj	(Mfcl)
	(Arab)	(Arab)	(Arab)	(Deva)	(Deva)	(Deva)	(Deva)	(Sinh)	(Sinh)	(Deva)	(Deva)	(Deva)	(Deva)	(Oick)	(Deva)	(Mfcl)
Stable Diffusion	51.21	22.92	22.58	23.41	23.71	25.16	22.98	23.19	23.41	24.00	23.83	22.90	23.88	24.14	23.64	23.98
Alt Diffusion	46.58	25.40	24.45	25.01	26.39	25.97	25.53	25.59	26.10	24.60	23.82	24.87	24.73	24.16	25.16	24.35
Alt Diffusion m9	42.43	28.39	25.05	27.48	27.34	29.31	28.24	30.17	29.53	29.79	23.28	27.16	25.83	24.10	28.24	24.26
MidJourney	48.94	23.30	22.76	22.63	22.90	23.09	22.65	22.97	22.64	22.16	22.57	22.90	22.79	21.25	22.53	21.99
Dalle3	48.92	44.10	38.81	44.04	43.18	44.88	44.22	44.03	44.17	37.46	29.76	44.69	38.91	23.18	42.18	22.46

Table 4: Language-Grounded Correctness (LGC) (%) across the different Indic languages in IndicTTI on the common set of prompts.

Model	en	asm	ben	guj	hin	kan	mal	mar	mnj	npi	ory	pan	san	snd	tam	tel
	(Beng)	(Beng)	(Gurj)	(Deva)	(Knda)	(Mlym)	(Arab)	(Beng)	(Beng)	(Orya)	(Guru)	(Deva)	(Arab)	(Arab)	(Tamil)	(Tulu)
Stable Diffusion	33.98	3.90	3.76	3.28	3.79	4.21	4.09	3.61	3.50	3.68	3.56	3.01	3.77	4.16	4.23	4.09
Alt Diffusion	30.90	6.50	6.87	7.34	7.56	10.67	6.53	7.81	4.54	8.25	8.57	7.87	8.49	6.53	7.12	6.00
Alt Diffusion m9	27.56	7.28	10.92	12.26	12.58	12.36	11.53	11.37	3.66	14.01	13.73	10.20	9.93	7.61	11.39	11.30
MidJourney	32.25	2.68	1.95	2.86	3.27	4.03	2.43	2.87	1.73	3.36	3.00	2.59	3.41	3.91	3.92	2.70
Dalle3	33.06	26.42	30.54	28.04	31.57	28.48	28.22	30.87	14.98	29.17	26.67	31.41	24.32	19.92	26.24	26.68
Model	en	urd	kas	hne	mal	awa	hne	mag	sin	brx	dot	gom	sat	snd	mnj	(Mfcl)
	(Arab)	(Arab)	(Arab)	(Deva)	(Deva)	(Deva)	(Deva)	(Sinh)	(Sinh)	(Deva)	(Deva)	(Deva)	(Deva)	(Oick)	(Deva)	(Mfcl)
Stable Diffusion	33.98	3.85	3.80	4.12	4.26	5.83	4.10	4.25	4.45	3.50	3.90	3.55	4.11	3.71	4.11	4.52
Alt Diffusion	30.90	7.72	5.41	7.22	8.22	8.04	7.82	7.70	7.81	5.99	4.71	5.98	6.12	5.76	6.82	6.30
Alt Diffusion m9	27.56	10.38	6.25	0.74	10.44	12.38	10.67	12.58	12.56	11.04	4.71	9.69	8.20	5.16	9.91	5.65
MidJourney	32.25	3.99	3.19	3.33	3.66	4.30	3.56	4.35	3.31	1.78	3.68	3.56	2.86	2.45	3.32	2.64
Dalle3	33.06	30.39	23.09	28.30	28.68	28.91	29.99	30.02	29.39	22.86	12.93	29.10	24.05	2.09	27.70	1.92

Table 5: Cyclic Language Grounded Correctness (CLGC) (%) across the different Indic languages in the IndicTTI benchmark on the complete set of prompts. Existing models provide high correctness for English languages while providing lower values for Indic languages.

Model	en	asm	ben	guj	hin	kan	mal	mar	npi	ory	pan	san	snd	tam	tel
Stable Diffusion	67.97	7.56	8.22	8.08	8.64	8.16	7.31	8.15	7.93	9.09	8.64	8.60	8.84	12.22	8.63
Alt Diffusion	62.40	14.32	16.87	15.42	17.34	22.33	15.67	16.36	12.07	17.27	18.78	16.35	16.54	17.13	17.69
Alt Diffusion m9	55.99	16.87	27.08	25.45	30.48	24.59	28.24	27.74	13.21	29.42	24.70	25.69	21.00	21.53	26.53
Midjourney	69.41	12.59	12.33	13.90	12.66	12.71	11.21	13.83	10.66	12.79	13.44	9.43	12.87	14.68	12.60
Dalle3	66.42	53.75	62.43	58.62	65.88	60.38	59.34	62.40	33.95	62.09	54.73	63.69	52.88	42.30	55.87
Model	en	urd	kas	kas	mal	awa	bho	hne	mag	sin	bix	doi	gom	sat	mi
Stable Diffusion	67.97	9.61	7.80	8.18	8.19	9.06	8.04	8.61	8.54	7.51	7.66	8.36	8.25	9.40	
Alt Diffusion	62.40	16.40	16.30	15.19	16.02	15.57	15.30	16.19	16.49	13.65	12.01	13.89	14.46	13.70	
Alt Diffusion m9	55.99	26.02	19.62	24.20	26.24	26.82	28.29	28.33	25.71	15.44	21.82	20.15	13.53	22.08	
Midjourney	69.41	15.85	13.90	13.85	13.09	13.76	12.81	12.52	13.32	13.02	12.87	12.65	13.97	11.71	
Dalle3	66.42	64.35	49.36	55.28	58.83	56.88	61.65	59.37	61.20	50.13	26.19	60.28	50.97	10.75	

Table 6: Image Grounded Correctness (IGC) (%) across the different Indic languages in IndicTTI on the complete set.

Model	en	asm	ben	guj	hin	kan	mal	mar	npi	ory	pan	san	snd	tam	tel
Stable Diffusion	53.97	23.48	23.57	22.72	22.93	22.65	22.49	22.99	23.40	23.32	23.25	22.24	22.63	23.45	
Alt Diffusion	49.12	23.64	24.30	23.78	25.12	26.22	23.45	24.22	22.43	25.05	25.54	24.49	24.41	24.24	
Alt Diffusion m9	45.62	24.82	28.39	29.13	29.85	28.90	30.43	28.47	23.49	29.87	27.94	28.35	25.72	27.34	
Midjourney	52.72	22.54	22.57	22.63	22.76	22.03	21.97	22.76	22.43	22.79	22.43	22.63	22.56	24.10	
Dalle3	48.54	39.48	45.13	43.10	45.86	42.57	46.26	30.42	44.47	40.34	45.54	38.23	35.46	41.71	
Model	en	urd	kas	kas	mal	awa	bho	hne	mag	sin	bix	doi	gom	sat	mi
Stable Diffusion	53.97	22.56	22.23	22.93	23.06	23.26	22.83	22.92	22.83	23.62	23.01	23.00	22.91	24.42	
Alt Diffusion	49.12	24.95	23.69	24.16	24.64	24.51	24.45	24.68	24.93	23.53	23.82	23.78	23.55	24.13	
Alt Diffusion m9	45.62	28.56	25.87	27.24	28.05	28.80	28.25	29.01	29.06	28.64	23.28	26.69	25.31	23.42	
Midjourney	52.72	24.05	22.95	23.04	22.66	23.25	22.87	22.76	22.79	22.60	22.45	22.66	22.81	21.08	
Dalle3	48.54	44.38	38.97	43.75	43.64	44.34	44.87	44.30	44.88	37.29	28.82	44.07	38.84	23.23	

Table 7: Language Grounded Correctness (LGC) (%) across the different Indic languages in IndicTTI on the complete set of prompts.

Model	en	asm	ben	guj	hin	kan	mal	mar	npi	ory	pan	san	snd	tam	tel
Stable Diffusion	35.00	4.05	4.34	3.85	4.26	4.04	4.62	3.95	4.15	4.18	4.00	3.37	4.03	4.58	
Alt Diffusion	32.47	6.31	7.22	6.81	8.22	6.70	7.21	4.44	8.34	8.56	7.47	7.39	7.14	7.73	
Alt Diffusion m9	29.22	6.44	11.48	11.86	13.70	11.23	13.16	11.79	4.32	13.30	10.69	10.92	8.76	9.19	
Midjourney	34.11	2.60	2.73	2.94	3.43	3.24	2.23	3.46	2.19	3.65	2.77	2.65	3.47	4.28	
Dalle3	33.04	25.79	30.96	27.64	31.59	28.51	28.44	14.22	29.77	26.54	31.37	24.83	18.57	26.45	
Model	en	urd	kas	kas	mal	awa	bho	hne	mag	sin	bix	doi	gom	sat	mi
Stable Diffusion	35.00	5.09	4.61	4.34	4.34	4.63	4.22	4.52	4.41	3.82	3.90	4.06	4.17	3.76	
Alt Diffusion	32.47	8.17	5.84	6.69	7.48	7.44	7.37	7.59	7.63	6.16	5.67	6.63	6.24	5.84	
Alt Diffusion m9	29.22	11.15	7.80	9.96	11.18	12.13	11.24	12.44	12.52	11.31	4.97	9.35	7.59	6.28	
Midjourney	34.11	4.19	3.15	3.62	3.29	4.22	3.30	3.88	3.43	3.42	3.32	3.36	3.35	2.52	
Dalle3	33.04	30.35	23.55	27.51	28.71	27.79	30.15	29.68	29.73	22.35	11.46	29.14	23.96	2.24	

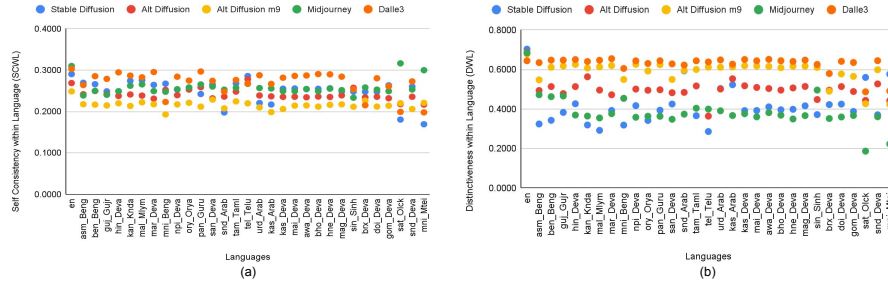


Fig. 2: Performance for the (a) SCWL and (b) DWL metrics of the benchmark showcasing self-consistency and distinctiveness of concepts within language, respectively, over the complete set of prompts.

5 Qualitative Analysis

In this main paper, we qualitatively analyzed the images generated by the different models across the different languages. We present more qualitative results here in Figs. 3, 4, and 5 corresponding to Stable Diffusion, Alt Diffusion, and Midjourney, respectively.

The Stable Diffusion model (Fig. 3) generates images containing a high number of individuals, temples, flowers, and gods. In the case of Arabic scripts (*kas_Arab*, *snd_Arab*, *urd_Arab*), the model produces men wearing Muslim caps and women in burkhas or niqabs. For all Devanagari scripts as well as for *as_guj_Gujr* in Gujarati script, the model generates individuals with sarees, and *tilak* which are often worn in the Indian culture. Additionally, it produces gods and temples. Sanskrit (*sans_Deva*) in particular, generates a large number of gods and temples due to its extensive religious context. This pattern is present across all languages using the Devanagari script. For languages using the Bengali script (*asm_Beng*, *ben_Beng*, *mn_Beng*), the model produces a significant amount of distinctive greenery. Sindhi in Arabic script (*snd_Arab*) is the only language generating a substantial amount of pornographic content.

The AltDiffusion model (Fig. 4) generates images of couples, gods, places of worship, and occasionally of Indian monuments such as the Taj Mahal, especially for languages with the Devanagari script. When generating for Arabic scripts, such as *kas_Arab*, *snd_Arab*, and *urd_Arab*, the model generates images of men wearing Muslim caps, women in burkhas or niqabs, and mosques showcasing a correlation between the Arabic script and Muslim culture. In the Gurumukhi script (*pan_Guru*), the model depicts individuals with long beards and turbans, highlighting correlations between the script of the Punjabi language with stereotypical portrayal of people from Punjab. Languages from the Dravidian family, including Telugu (*tel_Telu*), Tamil (*tam_Tam*), Malayalam (*mal_Mlym*), and Kannada (*kan_Knda*), along with Sinhala (*sin_Sinh*), feature images of dark-skinned individuals, possibly associating skin-color with the individuals being generated.

In Midjourney (Fig. 5), for the Devanagari script, the model produces a variety of visuals, including women, gods, and deities in Hinduism (like *Shiva*, *Ganesha*, and *Krishna*), *pandits* (priests), temple-like structures (religious places of worship), elephants, and tigers (typically shown in stereotypical depictions of India). Within the Dravidian family languages such as Malayalam (mal_Mlym), Kannada (kan_Knda), Telugu (tel_Telu), and Tamil (tam_Taml), common elements include jewelry such as necklaces, forehead pendants, earrings, and dark-skinned individuals, particularly men. Bengali scripts (asm_Beng, ben_Beng, mni_Beng) often feature bridal women in red sarees adorned with jewelry. Other commonly generated images include images of food, such as fish, which is prevalent in Bengali culture. Languages with Arabic script like Kashmiri (kas_Arab), Urdu (urd_Arab), and Sindhi (Snd_Arab) commonly depict women with *hijabs* or *niqabs*, men in headcovers, and places of Islamic worship such as mosques. In Gujarati, in addition to Devanagari influences, food-related imagery is prominent. Finally, Punjabi in Gurumukhi script (pun_Guru) frequently showcases individuals wearing turbans with a *Gurudwara* (place of worship in *Sikhism*, a religion commonly practiced by residents of Punjab) in the background. For certain languages such as Santali (sat_Olek) and Manipuri in Meitei script (mni_Mtei), Midjourney generates Asian women, depicting no Indian cultural influences (Fig. 6). It is interesting to note that these two languages also produce random outputs in the Dalle3 model, which understands many of the other Indic languages.

References

1. Midjourney. <https://www.midjourney.com/home> (2024)
2. Agirre et al., E.: Semeval-2012 task 6: A pilot on semantic textual similarity*. In: International Workshop on Semantic Evaluation. pp. 7–8 (2012)
3. Costa-jussà et al., M.R.: No language left behind: Scaling human-centered machine translation. arXiv preprint arXiv:2207.04672 (2022)
4. OpenAI: Dalle3. <https://openai.com/dall-e-3> (2024)
5. Visheratin, A.: Nllb-clip-train performant multilingual image retrieval model on a budget. arXiv preprint arXiv:2309.01859 (2023)
6. Visheratin, A.: Laion-coco-nllb. <https://huggingface.co/datasets/visheratin/laion-coco-nllb> (2024)



Fig. 4: Showcasing the influence of language scripts on the cultural aspects depicted in the Alt Diffusion model.

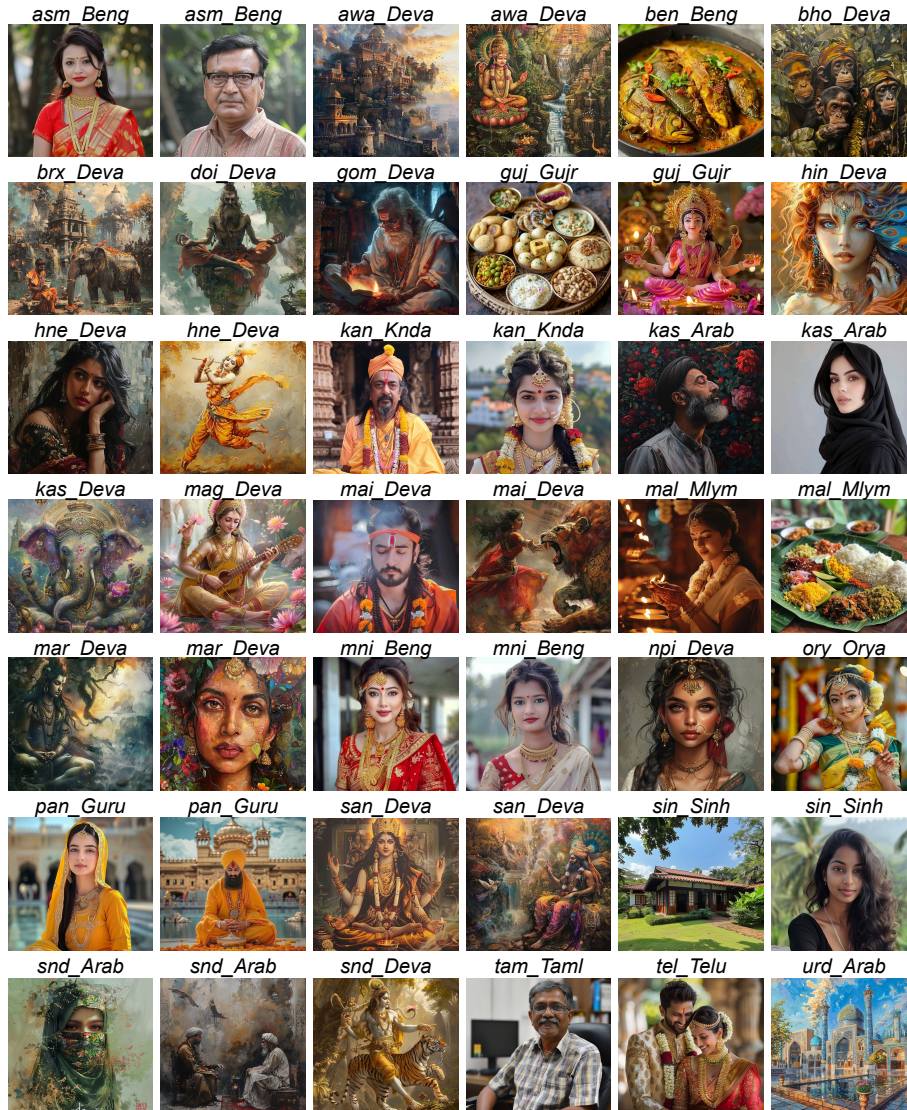


Fig. 5: Showcasing the influence of language scripts on the cultural aspects depicted in Midjourney.



Fig. 6: Showcasing the random generation of anime-style women and men in Midjourney when prompted to generate in Santali (Olck script) and Manipuri (Meitei script).