Would Deep Generative Models Amplify Bias in Future Models?

Supplementary Material

Table 1. Details of the common (with no social attributes) image retrieval results for OpenCLIP.

(a) COCO image retrieval

Flickr30K IR	0.0	0.2	0.4	0.6	0.8	1.0
@1	6.23	6.66	6.04	6.23	5.98	5.21
@5	17.09	17.91	16.99	17.65	17.25	14.44
@10	24.45	26.13	24.39	25.61	25.07	21.02

(b) Flickr30k image retrieval

Flickr30K IR	0.0	0.2	0.4	0.6	0.8	1.0
@1	4.79	5.29	4.81	5.47	5.25	4.47
@5	13.40	14.75	13.51	14.83	14.77	12.64
@10	19.19	21.13	20.17	21.49	21.35	18.69

Table 2. Details of the image retrieval performance and bias metrics for OpenCLIP in COCO dataset with gender and skin tone annotation. Characters in the table are: 1) In gender: **M**an and **W**oman; 2) In skin tone: **L**ighter and **D**arker.

		All	gender		skin-tone	
α			M	W	L	D
0.0	@1	3.1	3.0	3.2	3.4	3.2
	@5	9.2	9.2	10.3	10.1	9.2
	@10	13.7	13.7	15.1	15.0	13.4
0.2	@1	3.2	2.9	4.5	3.6	2.9
	@5	9.7	9.6	11.5	10.9	10.6
	@10	14.8	15.1	17.1	16.6	16.5
0.4	@1	3.0	3.1	3.9	3.7	1.9
	@5	9.2	9.3	11.1	10.8	8.0
	@10	13.8	14.1	16.2	15.8	14.3
0.6	@1	3.4	3.4	4.1	3.8	3.3
	@5	9.7	9.9	11.7	11.0	11.5
	@10	14.7	15.1	17.6	16.7	17.1
0.8	@1	3.5	3.4	4.9	3.9	4.8
	@5	9.8	10.0	12.2	11.1	12.2
	@10	14.5	14.9	17.7	16.5	17.1
1.0	@1	2.7	3.0	3.0	3.2	3.5
	@5	8.8	9.6	10.2	10.1	11.3
	@10	13.1	14.4	15.2	15.1	16.2

Table 3. Details of the image retrieval performance and bias metrics for OpenCLIP in PHASE dataset. Characters in each table are: 1) In gender: **M**an and **W**oman; 2) In age: **B**aby and **C**hild; 3) In skin tone: **L**ighter and **D**arker; 4) In ethnicity: **B**lack, **E**ast **A**sia, **I**ndian, **L**atino, **M**iddle **E**ast, **S**outheast **A**sia, and **Wh**ite.

		All	age				gender		skin-tone		Ethnicity						
α			B&C	young	adult	senior	M	W	L	D	B	EA	I	L	ME	SA	Wh
0.0	@1	2.3	3.0	2.3	2.4	3.9	2.6	2.3	2.6	1.9	1.2	1.7	2.2	1.2	0.0	0.0	2.8
	@5	6.3	8.5	6.2	6.3	9.1	6.9	5.7	6.7	5.3	4.5	4.0	5.6	3.6	6.3	2.1	7.2
	@10	9.1	11.5	8.9	9.1	12.2	9.7	8.6	9.5	8.7	8.2	7.5	8.9	6.0	6.3	4.2	10.0
0.2	@1	2.8	3.3	2.7	2.7	5.2	3.1	2.7	3.2	1.8	1.2	1.7	2.2	3.6	2.1	2.1	3.5
	@5	7.6	9.5	7.5	7.3	8.6	8.1	7.2	8.2	6.5	6.0	5.2	7.4	7.1	8.3	4.2	8.9
	@10	11.7	15.6	11.5	11.4	13.8	12.6	11.0	12.4	11.1	10.3	5.2	13.7	10.7	10.4	6.3	13.3
0.4	@1	2.4	4.1	2.4	2.0	4.4	2.7	2.1	2.8	1.7	1.4	2.3	4.1	2.4	6.3	2.1	2.9
	@5	6.9	9.7	6.9	6.3	10.7	7.1	6.5	7.5	5.3	5.2	4.6	7.8	7.1	8.3	2.1	8.1
	@10	10.3	13.8	10.6	9.3	14.3	10.4	10.1	11.0	8.8	9.1	5.2	11.5	8.3	18.8	4.2	11.7
0.6	@1	2.4	2.9	2.8	1.9	5.2	2.3	2.5	2.7	2.0	1.9	1.1	3.3	1.2	2.1	0.0	2.9
	@5	7.0	9.9	7.7	6.1	10.4	7.1	7.1	7.7	6.4	6.5	5.7	7.8	4.8	8.3	2.1	8.1
	@10	10.7	15.8	11.1	9.4	14.6	11.0	10.4	11.5	9.4	9.3	8.0	13.7	8.3	10.4	4.2	12.0
0.8	@1	2.2	4.0	2.1	1.9	4.4	2.2	2.2	2.5	1.3	1.4	1.1	2.2	3.6	0.0	0.0	2.8
	@5	6.5	10.9	6.4	5.5	11.5	6.6	6.2	7.1	5.8	5.7	4.6	6.3	7.1	6.3	2.1	7.6
	@10	9.9	15.0	9.5	8.9	15.4	10.2	9.4	10.7	9.0	8.1	7.5	12.2	9.5	10.4	4.2	11.2
1.0	@1	1.6	3.1	1.6	1.3	3.1	1.7	1.7	1.9	0.8	1.2	0.6	1.5	1.2	4.2	0.0	2.0
	@5	5.1	8.1	5.3	4.2	9.1	5.1	5.1	5.7	3.7	4.0	3.4	5.2	4.8	10.4	0.0	6.0
	@10	7.9	12.7	7.7	6.6	12.5	8.0	7.6	8.6	6.2	6.7	6.3	8.1	7.1	14.6	2.1	9.2

Table 4. Details of the normalized self similarity score for OpenCLIP. Characters in each table are: 1) In gender: **M**an and **W**oman; 2) In ethnicity: **B**lack, **E**ast **A**sia, **I**ndian, **L**atino, **M**iddle **E**ast, **S**outheast **A**sia, and **Wh**ite.

	gender		ethnicity							age								
α	M	W	EA	I	SA	Wh	ME	L	B	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	more than 70
0.0	-0.210	0.210	-0.366	0.463	0.377	-0.678	-0.640	0.301	0.543	1.247	1.041	0.635	-0.091	-0.497	-0.452	-0.442	-0.754	-0.687
0.2	-0.671	0.671	-0.097	0.563	0.420	-0.851	-0.636	0.122	0.479	1.290	0.997	0.538	-0.213	-0.618	-0.575	-0.497	-0.590	-0.331
0.4	-0.192	0.192	-0.272	0.332	0.196	-0.568	-0.374	0.111	0.575	0.507	0.581	0.472	-0.170	-0.336	-0.216	-0.138	-0.309	-0.392
0.6	-0.176	0.176	-0.419	0.523	0.212	-0.781	-0.454	0.013	0.906	1.045	0.491	0.440	-0.371	-0.567	-0.341	-0.233	-0.318	-0.146
0.8	-0.128	0.128	-0.503	0.347	0.138	-0.557	-0.274	-0.017	0.865	0.244	0.253	0.420	-0.274	-0.234	-0.097	0.051	-0.104	-0.258
1.0	-0.099	0.099	-0.322	0.245	0.192	-0.382	-0.130	-0.040	0.437	0.235	0.171	0.399	-0.220	-0.232	-0.066	-0.036	-0.047	-0.204

Table 5. Details of the person preference score for OpenCLIP. Characters in each table are: 1) In gender: **M**an and **W**oman; 2) In ethnicity: **B**lack, **E**ast **A**sia, **I**ndian, **L**atino, **M**iddle **E**ast, **S**outheast **A**sia, and **Wh**ite.

	gender		ethnicity							age								
α	M	W	EA	I	SA	Wh	ME	L	B	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	more than 70
0.0	0.646	0.660	0.518	0.486	0.513	0.507	0.491	0.505	0.484	0.447	0.430	0.394	0.380	0.375	0.365	0.361	0.357	0.356
0.2	0.823	0.895	0.709	0.653	0.668	0.703	0.674	0.669	0.621	0.851	0.865	0.896	0.900	0.895	0.886	0.883	0.873	0.859
0.4	0.904	0.879	0.637	0.581	0.608	0.641	0.611	0.602	0.551	0.730	0.758	0.772	0.791	0.785	0.770	0.752	0.726	0.676
0.6	0.976	0.994	0.353	0.330	0.329	0.331	0.304	0.311	0.307	0.825	0.840	0.838	0.832	0.808	0.778	0.741	0.712	0.702
0.8	0.974	0.955	0.079	0.055	0.069	0.089	0.067	0.068	0.064	0.883	0.951	0.979	0.984	0.981	0.978	0.969	0.952	0.939
1.0	1.000	1.000	0.149	0.130	0.134	0.183	0.164	0.146	0.113	0.943	0.978	0.988	0.989	0.985	0.978	0.962	0.938	0.905

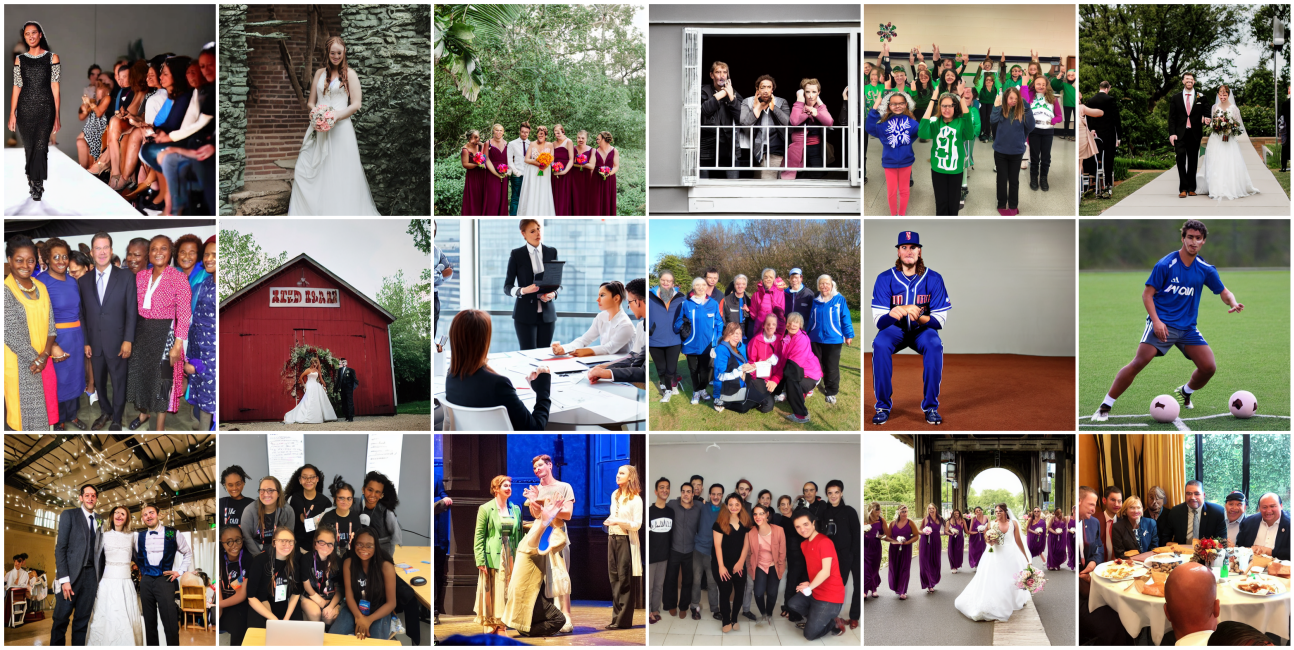

Figure 1. More examples of blurry faces in the generated images.

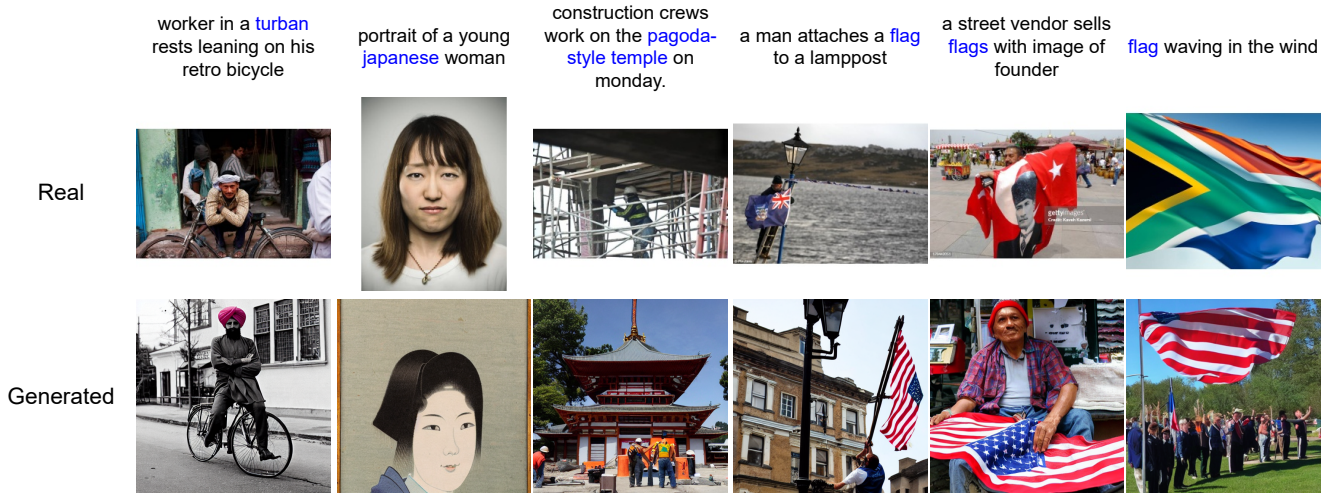

Figure 2. More examples of stereotyping in the generated images.