

# Supplemental Material: Understanding Human-Like Biases in VLMs via Subjective Face Analytics

Chaitanya Roygaga and Aparna Bharati  
Lehigh University, PA, USA  
{crr221, apb220}@lehigh.edu

## 1. Objective Attribute Detection

To understand the relationship between the subjective and objective face attributes in the VLM embedding space, we analyze the relative co-occurrence of objective attributes during the interpretation of subjective attributes. For this, we first initialize a set of 15 objective attributes for each subjective attribute using similarity scores in the VLM text embedding space. In Figures 1, 2, 3, we present the frequencies of top 5 objective attributes from the VLM embedding space that are visible according to the InternVL2-8B model, for the *attractive*, *trustworthy* and *dominance* subjective attributes. While eye-related attributes have the highest relevance for *attractive* in euclidean-space VLM models, both MERU and HyCoCLIP show its highest rel-

evance with cheekbones. It is evident that all VLM models consider “young” to be in their top relevant objective attributes. We present these results as an open research problem, as we cannot conclude the grounding of any specific subjective attribute with the objective face attributes. Our observations show the co-occurrence of these attributes in an attempt to understand the interpretation of subjectivity in MLLM and VLM embedding spaces for face attribute prediction.

## References

- [1] Miguel Ibáñez-Berganza, Ambra Amico, and Vittorio Loreto. Subjectivity and complexity of facial attractiveness. *Scientific Reports*, 9(1):8364, 2019. 6

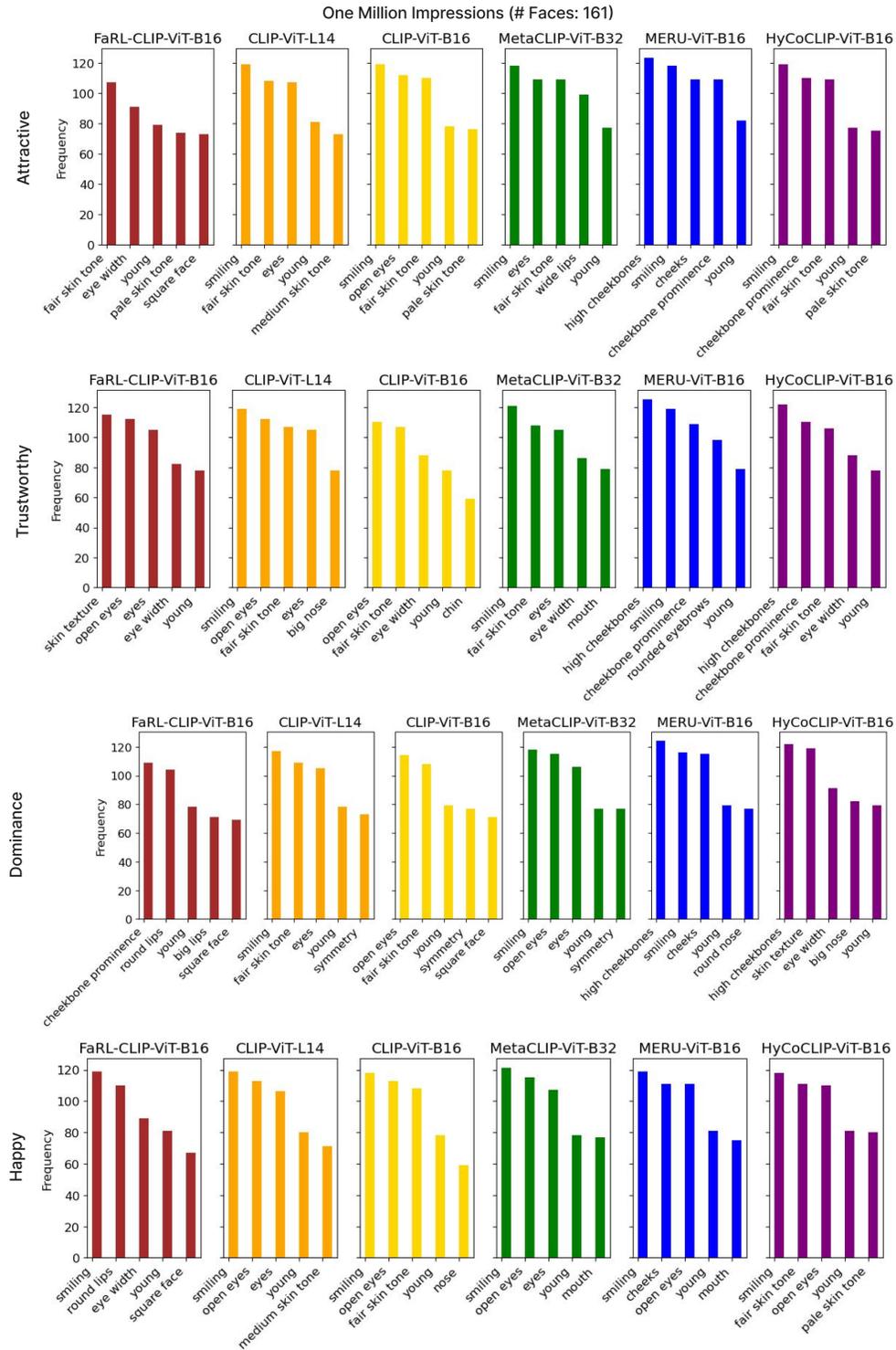


Figure 1. InternVL2 — OMI. Representing the co-occurrence of objective attributes in the VLM embedding space with the corresponding *Attractive* and *Trustworthy* subjective attributes. Out of 78 selected objective face attributes (Main paper, Section 3), attributes are first filtered based on the similarity scores in the VLM’s text embedding space. An MLLM is then used to determine if a certain objective attribute is visible in the face. The graph shows the frequency of objective attributes that are both in the vicinity to the subjective attribute and are visible in the face.

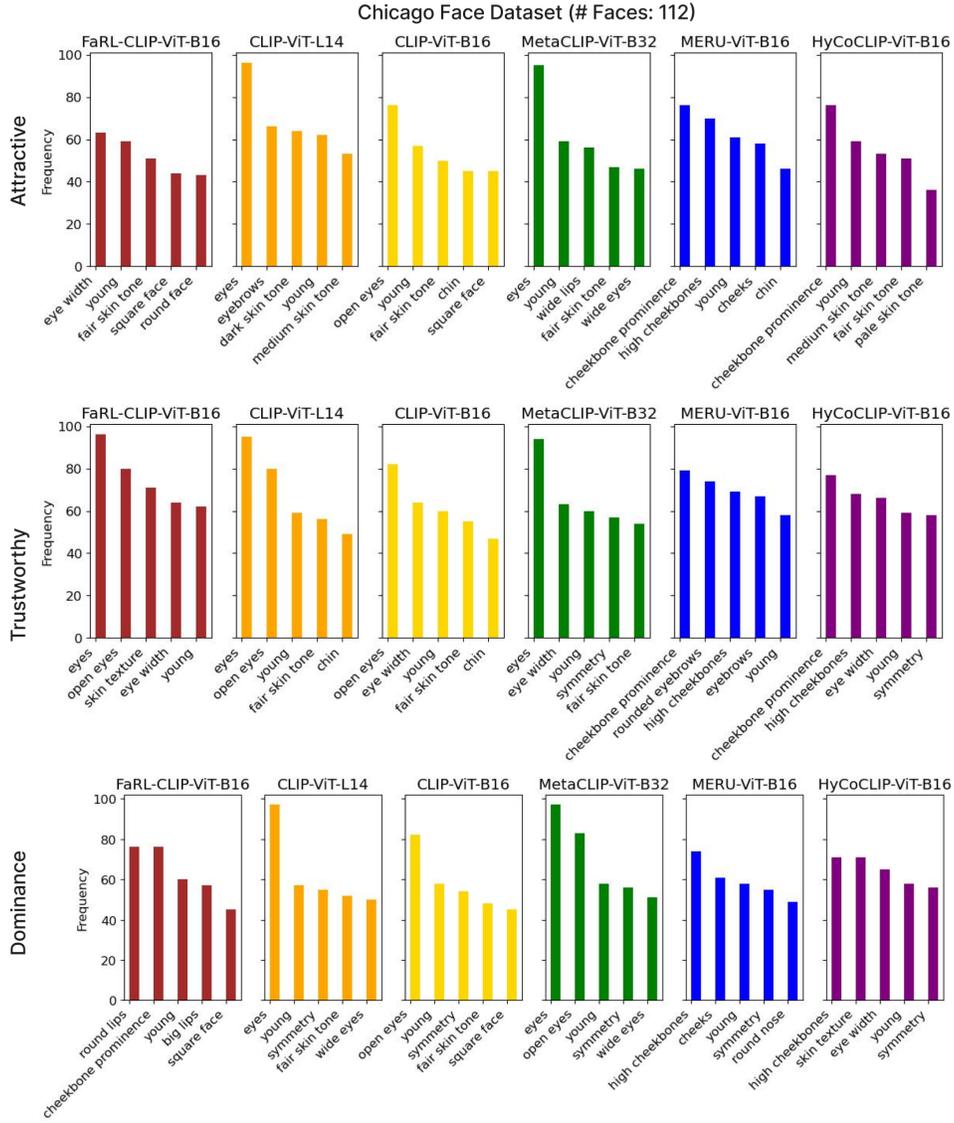


Figure 2. InternVL2 — CFD:1

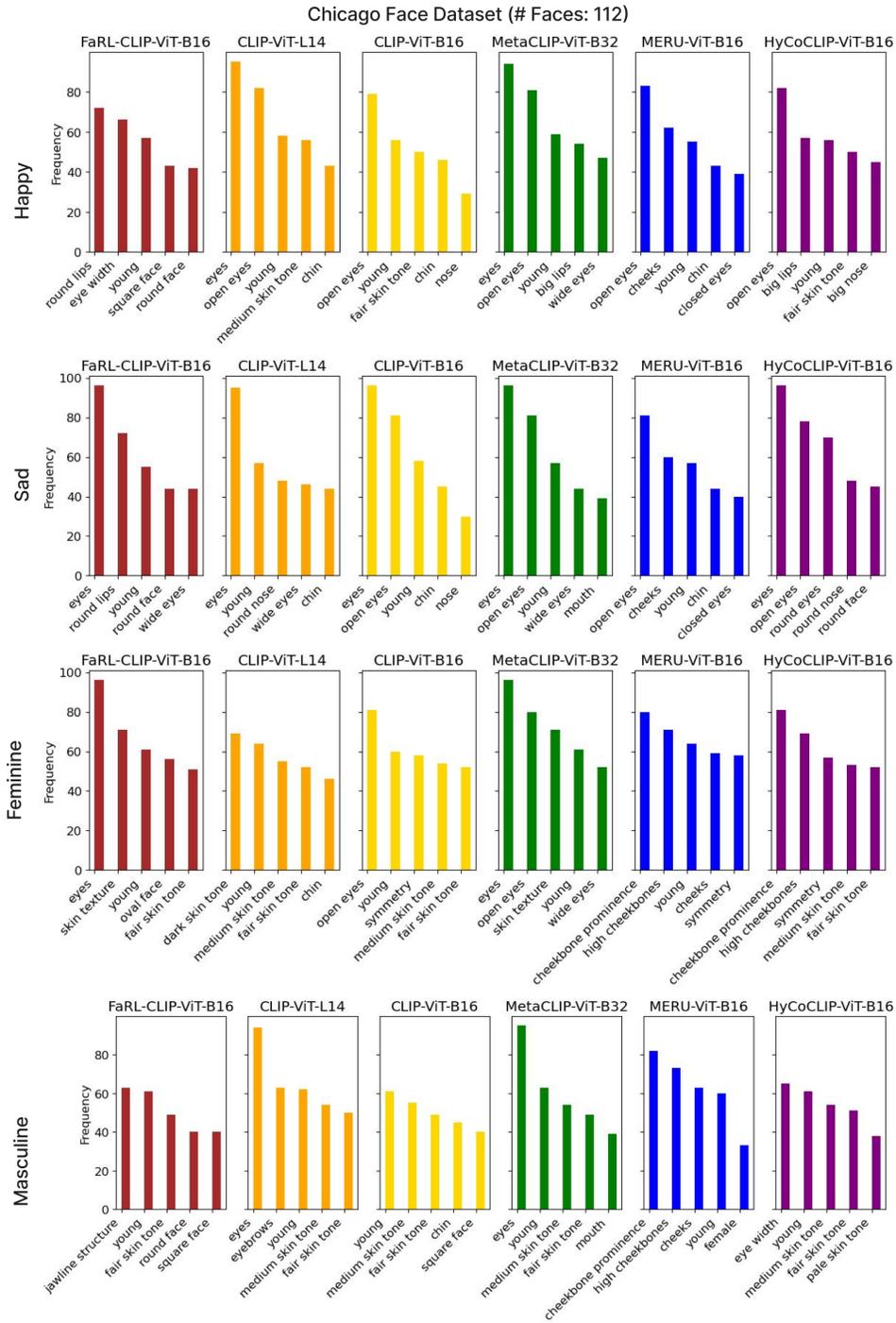


Figure 3. InternVL2 — CFD:2

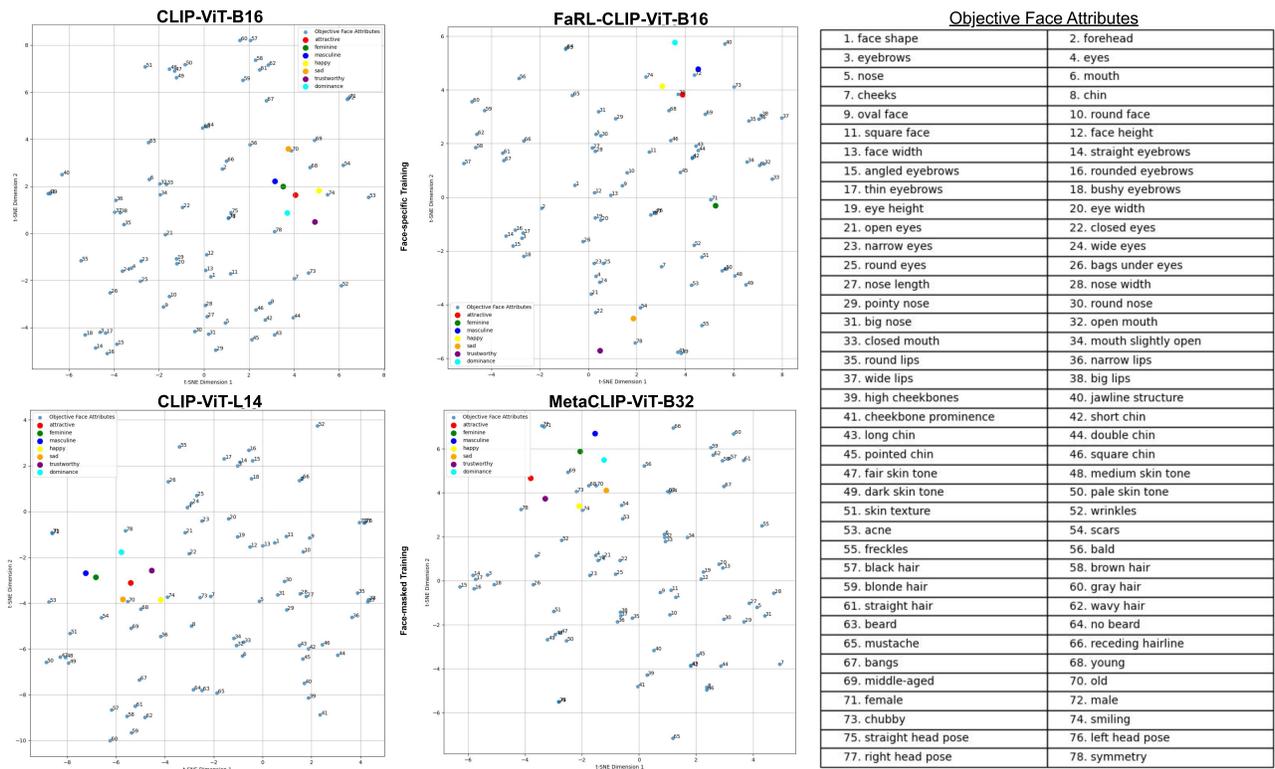


Figure 4. Representation of *word* text embeddings of subjective and objective face attributes in the Vision-Language Model Embedding Space.

Table 1. VLM Pairwise Comparison Agreement %. The Standard Template used is “a photo of a <subjective attribute> face”. The Context-relevant Generated Captions are 5 captions that describe the context of the subjective attribute in a particular face image. Humans are capable of interpreting subjective attributes from faces differently in different contexts [1], so we make use of the stochasticity provided by MLLM predictions to mimic human interpretation of that context. ST-CA and CGC-CA are pairwise agreement scores. From each dataset, we pick a pair of faces and compare their ground truth (GT) scores — if the VLM similarity scores follow a similar comparison result with the GT scores, if face “a” has a higher dominance score than face “b” and VLM similarity score comparison result in the same, the interpretation of subjectivity for that face is considered to be in agreement with human-bias. Note that the GT scores are not directly compared with the VLM similarity scores, as the similarity scores only represent the alignment of text and image context in the embedding space and are not real scores which are predicted for the input face based on the text captions. Keys — IVL2: InternV12-8B, L-SI: Llama-OneVision-Qwen2-7B-SI-HF.

VLM	DATA	STANDARD TEMPLATE COMPARE AGREEMENT (ST-CA) %							MLLM	CONTEXT-RELEVANT GENERATED CAPTIONS COMPARE AGREEMENT (CGC-CA) %						
		ATT.	TRUST.	HAPPY	DOM.	SAD	FEM.	MASC.		ATT.	TRUST.	HAPPY	DOM.	SAD	FEM.	MASC.
META-CLIP-VIT-B32	CFD	58.15	45.11	58.77	48.37	58.96	69.11	72.37	IVL2	73.62	67.39	68.32	68.45	68.56	80.55	78.68
									L-SI	66.73	62.66	61.84	68.55	68.36	69.30	72.49
	OMI	69.73	57.31	75.67	47.60	-	-	-	IVL2	77.38	76.89	73.23	55.40	-	-	-
									L-SI	74.21	76.26	69.24	49.70	-	-	-
	UAF	65.88	52.62	68.14	-	-	-	-	IVL2	76.95	72.00	80.65	-	-	-	-
									L-SI	75.98	69.64	72.75	-	-	-	-
CLIP-VIT-L14	CFD	58.46	48.00	51.06	55.08	45.18	60.40	43.42	IVL2	75.68	62.76	66.23	60.99	65.51	83.61	73.23
									L-SI	75.00	56.64	57.21	62.09	54.45	68.48	60.96
	OMI	63.40	43.24	67.63	57.70	-	-	-	IVL2	77.77	73.80	68.73	56.77	-	-	-
									L-SI	71.95	71.76	64.56	52.12	-	-	-
	UAF	51.40	43.68	53.75	-	-	-	-	IVL2	72.95	68.75	74.75	-	-	-	-
									L-SI	69.81	65.27	56.18	-	-	-	-
CLIP-VIT-B16	CFD	60.28	49.44	48.81	46.80	53.38	67.86	58.02	IVL2	75.80	65.19	70.75	70.45	71.91	85.42	75.39
									L-SI	72.24	66.60	61.22	70.80	62.66	75.06	58.27
	OMI	58.39	49.85	73.32	58.76	-	-	-	IVL2	76.39	72.33	70.91	58.35	-	-	-
									L-SI	69.49	73.50	62.32	54.99	-	-	-
	UAF	55.08	50.79	62.28	-	-	-	-	IVL2	75.25	72.36	74.55	-	-	-	-
									L-SI	71.56	70.27	62.70	-	-	-	-
MERU-VIT-B16	CFD	50.50	49.37	59.02	47.74	55.76	70.80	69.99	IVL2	74.32	68.97	74.07	63.21	72.15	83.29	74.68
									L-SI	68.73	65.79	62.15	68.80	68.17	76.25	66.79
	OMI	66.16	32.53	71.92	57.14	-	-	-	IVL2	78.70	72.00	74.88	61.29	-	-	-
									L-SI	64.85	73.84	61.47	57.52	-	-	-
	UAF	60.61	41.36	61.97	-	-	-	-	IVL2	78.35	73.00	75.39	-	-	-	-
									L-SI	71.31	67.31	59.51	-	-	-	-
HYCoCLIP-VIT-B16	CFD	63.72	51.38	56.95	52.32	55.39	71.99	55.14	IVL2	80.10	70.75	69.37	68.29	69.42	82.98	69.71
									L-SI	73.62	66.10	66.17	69.99	62.09	76.19	67.18
	OMI	68.58	57.91	72.60	45.96	-	-	-	IVL2	82.00	79.36	76.38	59.11	-	-	-
									L-SI	66.26	74.45	64.60	51.51	-	-	-
	UAF	63.63	54.68	64.50	-	-	-	-	IVL2	80.33	72.12	76.23	-	-	-	-
									L-SI	75.09	69.00	62.73	-	-	-	-
EARL-CLIP-VIT-B16	CFD	51.38	42.11	55.33	50.38	63.78	60.90	67.11	IVL2	75.11	70.08	68.52	69.98	76.71	81.63	78.25
									L-SI	64.10	60.09	61.47	66.42	79.57	72.49	74.06
	OMI	69.17	55.79	71.86	45.03	-	-	-	IVL2	76.00	73.00	75.23	55.25	-	-	-
									L-SI	67.18	73.38	70.39	53.50	-	-	-
	UAF	64.45	52.71	62.78	-	-	-	-	IVL2	74.93	72.27	78.57	-	-	-	-
									L-SI	73.01	70.57	73.17	-	-	-	-