Visual Artificial Intelligence's Matter of Taste
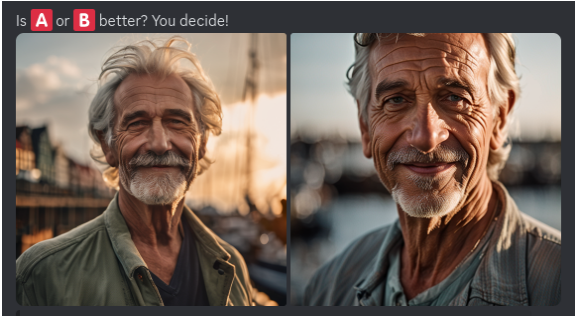Bruno Caldas Vianna, Uniarts Helsinki

"Taste classifies, and it classifies the classifier."(Bourdieu 2002, 41)

In times when machine learning visual generative methods dominate more and more the cultural landscape, ownership of cultural capital is obtained through statistical processes. The aesthetic quality of images must be evaluated in two crucial steps of the training of visual neural networks. First, by filtering out pictures that lack enough quality to be used in the models. The original five billion images collected by the Laion non-profit into their Laion-5B set are filtered down to 600 million for their Laion-Aesthetics collection, which was the one used to train the first versions of the Stable Diffusion model. (Schuhmann et al. 2022) Secondly, during the training process itself, test synthetic images are created and evaluated to guide the data towards better quality. But who dictates what is considered good quality? Whose taste gets imprinted in the data?

As it turns out, it is the taste of a very select group of people. Aesthetic evaluation data is obtained online through rating systems on public websites and chat forums. The users of these systems grade images, sometimes on a 0-9 scale, sometimes by comparing two pics and choosing the most appealing. The problem is, even though these websites are available to a general audience, most of their users happen to be people who already have an interest in synthetic images and know where to vote.

The *simulacra aesthetics bot* developer, John Pressman, acknowledges in its own repository that most votes come from WEIRD people - white, educated, industrialized, rich, from democratic countries. (Pressman [2022] 2023) Another solution is proposed by researchers from Tsinghua University, who used the services of a professional company to rate images, most of them with university degrees. (Xu et al. 2023)

These current rating systems perpetuate the taste of people with a very specific cultural capital, built from access to computers, the internet, and an interest in computational creativity. To achieve a more diverse representation of visual culture in neural generative systems, we must incorporate ratings from people from the Global South, people who do not use generative systems or even computers in general, and people from marginalized cultures. This realization marks the beginning of a project to build a more diverse aesthetic guidance.

Stable Foundation's rating bot, accessible through its Discord channel

References

Bourdieu, Pierre. 2002. *Distinction: A Social Critique of the Judgement of Taste*. 11. print. Cambridge, Mass: Harvard Univ. Press.

Pressman, John David. (2022) 2023. "Simulacra Aesthetic Captions Bot." Python. https://github.com/JD-P/simulacrabot.

Schuhmann, Christoph, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, et al. 2022. "LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models." arXiv. https://doi.org/10.48550/arXiv.2210.08402.

Xu, Jiazheng, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. "ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation." arXiv. http://arxiv.org/abs/2304.05977.