Selected Papers of #AoIR2021:
The 22nd Annual Conference of the
Association of Internet Researchers
Virtual Event / 13-16 Oct 2021

WHICH HUMAN FACES CAN AN AI GENERATE? LACK OF DIVERSITY IN THIS PERSON DOES NOT EXIST

Lucas Nunes Sequeira[1]
Group on Artificial Intelligence and Art (GAIA) and Molecular Sciences course, University of Sao Paulo

Bruno S. Moreschi[2]
Group on Artificial Intelligence and Art (GAIA) and Faculty of Architecture and Urbanism, University of Sao Paulo

Vinicius Ariel Arruda dos Santos[3]
Group on Artificial Intelligence and Art (GAIA) and Polytechnic School of the University of São Paulo

1. Introduction

In this abstract we show the results of an interdisciplinary research in which we audit fake human faces (deepfakes) generated by the website This Person Does Not Exist (TPDNE), and discuss how this system can help perpetuate normativities supported by a dependency on a limited database. Our analysis focuses on the "default generic face" we created by overlapping random samples of fake faces generated by TPDNE's algorithms. Independently of the group of fake human faces sampled, the same generic white face always appeared as a result (Fig. 1).

[1] Lucas N. Sequeira is an undergraduate student of Molecular Sciences at the University of São Paulo, researcher at Group on Artificial Intelligence and Art (GAIA / C4AI / Inova USP) and junior researcher at CPqD, in the area of Natural Language Processing. E-mail: lucasnseq@usp.br
[2] Bruno Moreschi is a visual artist, postdoctoral fellow at University of São Paulo's Faculty of Architecture and Urbanism, senior fellow at the Center for Arts, Design and Social Research (CAD+SR), member of the Histories of AI: A Genealogy of Power group (University of Cambridge) and co-coordinator of GAIA. E-mail: brunomoreschi@usp.br
[3] Vinicius Ariel Arruda dos Santos is an undergraduate student of the Polytechnic School of the University of São Paulo, researcher at Group on Artificial Intelligence and Art (GAIA / C4AI / Inova USP). E-mail: viniciusarielarruda@usp.br

Suggested Citation (APA): Sequeira, L., Moreschi, B., and Arruda dos Santos, V., (2021, October). Which Human Faces Can an AI Generate? Lack Of Diversity In This Person Does Not Exist. Paper presented at AoIR 2021: The 22nd Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from http://spir.aoir.org.

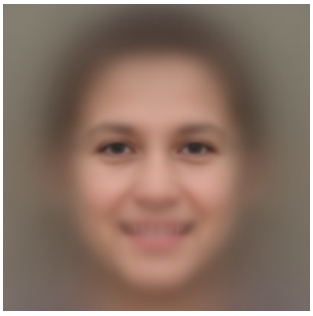

Fig. 1: “Default generic face” – overlapping 3000 random fake faces by TPDNE. Credit: Bernardo Fontes and Lucas N. Sequeira / GAIA

Faces algorithmically generated are part of the improvements in the CV field since 2014, with the development of a machine learning infrastructure called Generative Adversarial Network – GAN (Goodfellow et al., 2014). The increase of hardwares processing capacities made the intensive use of this infrastructure viable, recently allowing the construction of more realistic images of humans, animals and objects from the new GANs known as StyleGAN (Karras et al., 2019) and StyleGAN2 (Karras et al., 2020; Nie et al., 2020).

Created in 2019, TPDNE is the result of using this last specific GAN and it was only achievable thanks to the use of the database of real human faces from the Flickr-Faces-HQ database, with 70 thousand high definition images. Every time the website is refreshed, its AI renders a new (and fake) human face.

The rise of CV has prompted several debates such as the “algorithmic injustices” (Noble, 2018). In this research we are particularly inspired by specific analysis of datasets such as the auditing on ImageNet conducted by Prabhu and Birhane (2020), and also the social and cultural impacts that those systems reflects in society approaching cultural aspects (Mintz et al., 2019) or to certain phenotypic subgroups (Buolamwini and Gebru, 2018).

Fig. 2: Samples of fake human faces generated by TPDNE.

2. Methodology

We first built a database with 4100 fake human faces taken from TPDNE website via web scraping. Then, we analysed them through a Python language script created by us, and discussed behaviours identified in this StyleGan2. Our analyses are based on the use of specific images created from a subset of overlapping fake human faces available in our database. These resulting images, called here “cluster-images”, were made from the overlapping of *N* arbitrary images generated by the TPDNE's algorithms.

3. What does this face tell us?

3.1. A face that imposes itself independent of the selection in the dataset

To prove this statement, we first created six “cluster-images” by overlapping fake faces scraped from TPDNE (Fig. 3), each formed by different amounts of images. We observed that the resulting “cluster-images” are similar: a person with white skin, dark eyes and brown hair, characteristics that are preserved regardless of the number of different fake faces overlapped.

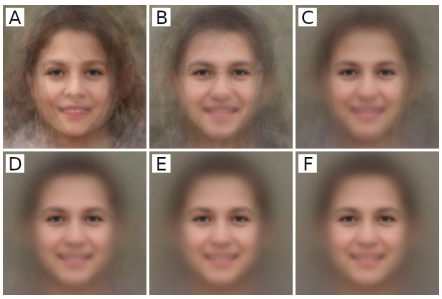

Fig. 3: Set of six images resulting from overlaps each with a number of different images. The “cluster-images” A, B, C, D, E and F were generated by, respectively, 10, 30, 100, 300, 1000 and 3000 images taken from our database randomly and without repetition.

To assess how these “cluster-images” are similar to each other, we created a set of 20 “cluster-images” for each value of *N*. In each set “cluster-images” with a fixed *N*, we calculated the mean deviation between the images in them (Fig. 4). This curve reveals that the degree of similarity increases exponentially for “cluster-images” that were generated with a larger number of data. From more than 1000 images, we can assume that the “cluster-images” generated are the “default generic face” of the database (Fig.1; E and F in Fig. 3). This experiment shows that, when we talk about the TPDNE's algorithms, a larger data volume does not mean an increase in the diversity of results, but the contrary.

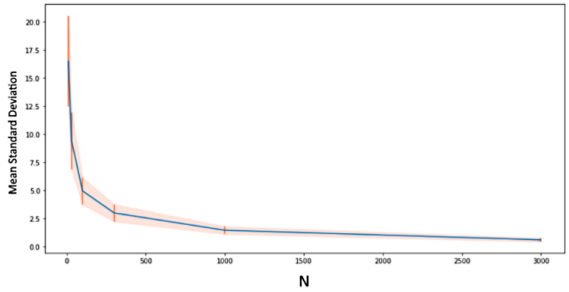

Fig. 4: The chart represents the mean standard deviation curve between set images assuming different values of *N*, each generated by 20 different “cluster-images”.

3.2. Faces of black people as unlikely

We also analyzed how the image database behaves in relation to skin tones diversity. For this, we selected all the faces of black people from a set of 4100 TPDNE results – there were only 54 faces, which represents 1.4% of the whole set, which per se shows this racialization and whiteness of the dataset. With these images we generated a “cluster-image” (A). We also selected 54 random images of white faces to create another “cluster-image” (B) for comparison (Fig. 5). Finally, we conclude that the “cluster-image” generated by white people is visually more similar to the “default generic face” than the “cluster-image” generated by faces of black people. This is empirical evidence that the creation of fake faces is not independent of characteristics such as whiteness imposed by processes that are already historically known.

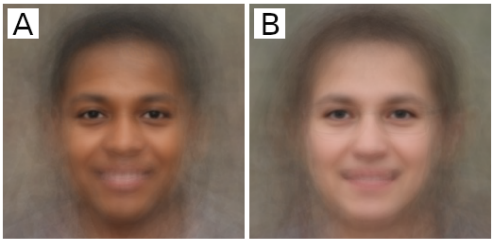

Fig. 5: “Cluster-images” of two overlaps: A (54 images of different fake faces of black people) and B (54 different fake faces of white people).

4. Why is this research important?

These results intrigue us because they are part of a first generation of realistic images created independently from the human eye, since they are synthesized solely by algorithms – a new chapter in the history of post-photography (Beiguelman, 2020). But also, particularly because the lack of diversity of TPDNE's generated faces is not a bug in this digital infrastructure, but a reinforcing standard dynamic that historically regulates bodies, territories and practices, from which the computer sciences are not removed. It is not by chance that Beiguelman (2020) also raises the question: are fake human faces the announcement of a new era of the eugenics of images?

Why, for example, is there a generic smile on “default generic face” and in all the fake human faces in TPDNE? What does it hide from us? Thousands of precarious crowdworks (as Turkers) organizing the Flickr-Faces-HQ database may be one of the answers (Karras et al., 2019). The fact that this StyleGAN was only possible thanks to the use of amateur personal images posted on Flickr (Smits and Wevers, 2021) and taken without the consent of its authors also tells us a lot about how data is made available to train Artificial Intelligence.

5. References

Beiguelman, G. (2020). *As verdades dos deepfakes*. Retrieved March 25, 2021, from https://book.affecting-technologies.org/as-verdades-dos-deepfakes/

Buolamwini, J.; Gebru, T. *Gender shades: Intersectional accuracy disparities in commercial gender classification*. In: Conference on fairness, accountability and transparency. PMLR, 2018. p. 77-91.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. (2014). *Generative Adversarial Networks*. Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

Karras, T., Laine, S., & Aila, T. (2019). *A style-based generator architecture for generative adversarial networks*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4401-4410)

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). *Analyzing and improving the image quality of stylegan*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8110-8119).

Mintz, A., Gobbo, B., Silva, T., & Pilipets, E. (2019, April 17). *Interrogating Vision APIs*. Retrieved March 18, 2021, from: https://www.researchgate.net/publication/332910402_Interrogating_Vision_APIs/

Nie, W., Karras, T., Garg, A., Debnath, S., Patney, A., Patel, A., & Anandkumar, A. (2020, November). *Semi-supervised stylegan for disentanglement learning*. In: International Conference on Machine Learning (pp. 7360-7369). PMLR.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Prabhu, V. U.; Birhane, A. (2020). *Large image datasets: A pyrrhic win for computer vision?*. arXiv preprint arXiv:2006.16923.

Smits, T.; Wevers, M. (2021) *The Agency of Computer Vision Models as Optical Instruments*. Visual Communication. https://doi.org/10.1177/1470357221992097.