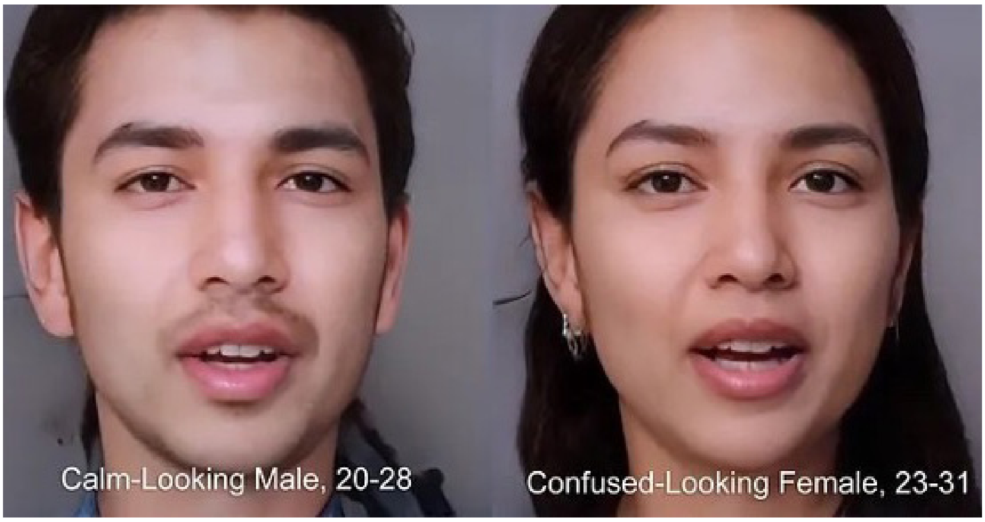
xCoAx 2023 11th Conference on
Computation, Communication,
Aesthetics & X
2023.xCoAx.org
Weimar, Germany

Deep Hysteria



Amy J. Alexander
ajalexander@ucsd.edu
University of California San Diego, USA

DOI 10.34626/xcoax.2023.11th.316

Deep Hysteria is a still image series that repurposes algorithmic bias in the service of unraveling a deep human bias. Artworks are generated using deep learning algorithms trained on still frames of thousands of YouTubers speaking to the camera. Generated individuals are then algorithmically gender-adjusted and the variations fed to Amazon Rekognition, a commercial deep learning based facial analysis algorithm (Amazon n.d.) that attempts to classify faces according to the subject's gender, age, and emotional appearance. Despite the marketing of such tools, reading emotions solely by analyzing a person's face is a feat that neither humans (Callahan 2021; Le Mau 2021) nor "AI's" (Crawford 2021) can reliably do. Further, these deep learning algorithms are themselves trained on data categorized by humans — so they reflect human biases. The side-by-side images in *Deep Hysteria* compare Rekognition's interpretation of similar expressions on more masculine and more feminine versions of the same face. The comparisons interrogate how humans perceive emotion differently, and often in alignment with stereotypes, when observing people of differing genders.

Keywords: Gender Bias, Generative Deep Learning, Algorithmic Bias, Facial Analysis, Emotion Detection.

Hysteria, Then and Now

Figure 1: Sequence of drawings from 1893 depicting a woman with "hysteria."

For centuries, "hysteria" was a medical and mental diagnosis (McVean 2017) that assumed females had an innate predisposition toward an anxious and nervous emotional state. Although the diagnosis has been retired, stereotypes of women as nervous, fearful, and uncertain continue to impact how women are perceived and treated. And while more women than men are diagnosed with anxiety, a Google image search for "anxiety" returns a far disproportionate number of images of women — who tend to be depicted in stereotypical poses of extreme emotional distress.

Figure 2: Google Search Image Results for "Anxiety," January 2023.

The stereotype is further augmented by the cultural expectation of smiling as women's default facial expression. Consider the phenomena of "Resting Bitch Face" (Grossman 2019) and "telling women to smile" (Smith 2016). A neutral facial expression on a woman is read as disgust, distress or unhappiness: "What's wrong?"

In recent years, deep learning-based facial analysis algorithms such as Amazon Rekognition have been marketed as facilitating the identification of apparent emotion on faces captured in photos or videos.

These "emotion detection" services have been widely criticized as being inaccurate and highly problematic (Simonite 2019, Crawford 2021). As of January 2023, Amazon qualifies its emotion detection API as "only making a determination of the physical appearance of a person's face. It is not a determination of the person's internal emotional state and should not be used in such a way" (Amazon n.d.). It is unknown what percentage of users note this warning and limit usage of Rekognition accordingly.

But it isn't only "AI's"[1] that can't read emotion based on facial expressions. Neuroscientists studying facial movements have demonstrated that facial expressions alone do not sufficiently convey emotion. Identical facial expressions can mean different things depending on culture — or context (Crawford 2021; Callahan 2021; Le Mau 2021). We think we can "read" other people's internal states — we can't.

Deep learning-based algorithms are trained on data produced by humans, so they reflect and often amplify human biases. Although developers typically do not intend to replicate problematic biases in their models, the nature of the training process provides many opportunities for problems to happen inadvertently. The training dataset may lack sufficient diversity — e.g. a face classification system might disproportionately misclassify darker-skinned people as a result of having an insufficient number of darker-skinned faces in the training dataset (Buolamwini and Gebru 2018). Or the system may tacitly "learn" proxies for historically biased behavior — e.g. a recruiting algorithm trained on historical data might end up favoring job applicants with names or pastimes common among white men (Bogen 2019). In such cases, the unintended bias is an indirect consequence of the design of the training system.

In the case of emotion detection algorithms, however, bias and subjectivity are at the core of the concept itself. Whether performed by human or machine, the identification of a person's internal emotions using external criteria is inherently subjective. Various emotion-related training datasets of pre-categorized faces, incorporating both posed and spontaneous emotions, are available to developers (Boesch n.d.) — or a developer may create their own dataset. Either way, the implicit biases inherent in the way humans pose and categorize the facial expressions in the dataset will be directly passed on to the detection algorithm. And since these biases are so deeply embedded socially, a system that reflects them may go unnoticed by both developers and users. Stereotypical results may simply appear to be "right."

1. Although formal definitions of "AI" currently refer to the broad concept of artificial intelligence rather than specific applications, a common, informal usage has emerged to refer to certain types of AI-based applications. The latter usage implies a context in which the software is a functional entity whose performance of tasks can be contrasted with that of either humans or conventional software algorithms. For this reason, the informal usage is employed here.

The artworks in *Deep Hysteria* redeploy the bias embedded in facial analysis algorithms in the service of probing this deeply entrenched social bias.

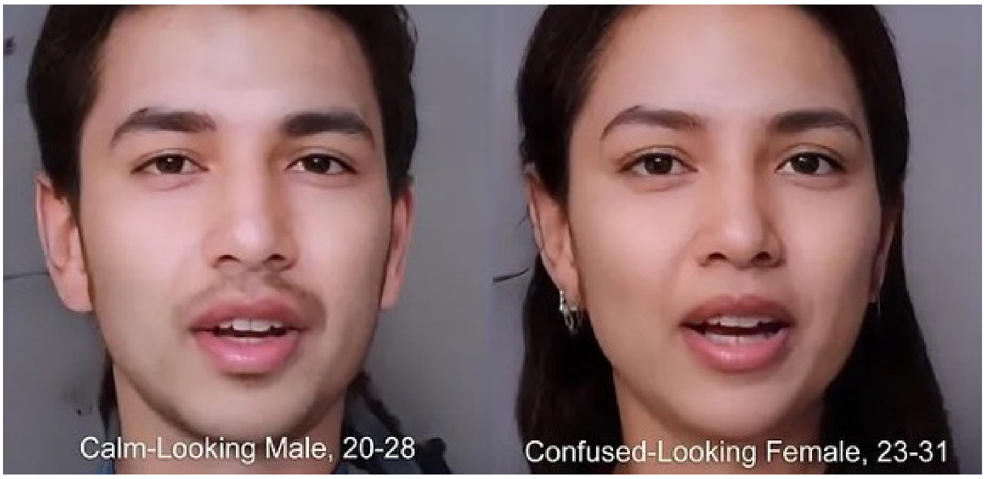


Figure 3: *Deep Hysteria* artwork.

Vloggers and Generative Deep Learning Portraiture

In addition to their utilitarian function of portraying gender-variable artificial personae, *Deep Hysteria*'s virtual portraits also serve as an exploration of texture, pose, and identity in generative deep learning portraiture. While many portraits generated via deep learning default to the use of posed, high resolution still photographs as training data, *Deep Hysteria* is trained on frames from You Tube vlogs produced by video makers with low subscriber counts. These vloggers, who come from countries around the world, photograph themselves speaking to the camera and are primarily amateur videographers. Their videos are typically shot with phone cameras and webcams and are photographed in a variety of settings, often with less-than-ideal lighting. Consequently, *Deep Hysteria*'s generated images are composite portraits of the self-selected vlogger addressing their audience, rather than the posed and curated subject of a photographer's gaze.[2]

2. The image generation model was trained using transfer learning from the Flickr-Faces-HQ (FFHQ) dataset, which is composed of primarily posed portraits posted to the Flickr website. Transfer learning begins with the previous dataset as a basis; the model is then retrained on the new images. The *Deep Hysteria* images retain some characteristics of the Flickr dataset. Developing a model trained from scratch would have required considerably more personnel and computation resources than were available for this project.

Figure 4: *Deep Hysteria* artwork.

The Process

Images featuring a human face were collected over the course of a year from YouTube videos; the collection was oriented toward vlogs and other first-person narratives by non-professional video makers. In most cases, the images were captured as the video makers were speaking to the camera. These images were then used to train a Stylegan2-ADA model, using transfer learning from the Flickr Faces HQ dataset. The resulting model was then used to generate a set of baseline images.

A subset of baseline images with relatively neutral facial expressions were then selected. From these images, a set of images with variations across a spectrum of gender presentations was created using a latent direction for gender developed by Robert Luxemburg (Luxemburg 2019). An attempt was made to compensate to the degree possible for changes in the image's facial expression resulting from algorithmic entanglement with the gender variation vector (mouth and eye openness), so that all gender variations for a given face had closely similar facial expressions. All gender variations were then submitted to Amazon Rekognition, a popular commercial image recognition and facial analysis service that offers an emotion detection API. Rekognition analyzed and labeled the faces according to gender presentation, age, and apparent mood. Results returned with less than 50% confidence were ignored and labeled as "unknown."

While the predominant mood analyzed by Rekognition for both male-identified and female-identified "neutral expression" images was "calm," a significantly greater number of male-identified images received this designation. Female-identified images were more likely to be analyzed with stereotypically female emotions: "fear," "confused," "sad," "surprised," "disgusted." Male-identified images were more likely to be designated "angry" by Rekognition. Although *Deep Hysteria* faces are gendered across the gender spectrum and thus includes non-binary faces, Rekognition lacks designations beyond binary gender presentation; it labels all images as either "male" or "female." Non-binary-appearing images inclusive, those images Rekognition identified as "female" were more likely to be labeled with stereotypically feminine emotions. This effect was notable in cases where the corresponding masculine image was labeled "calm."

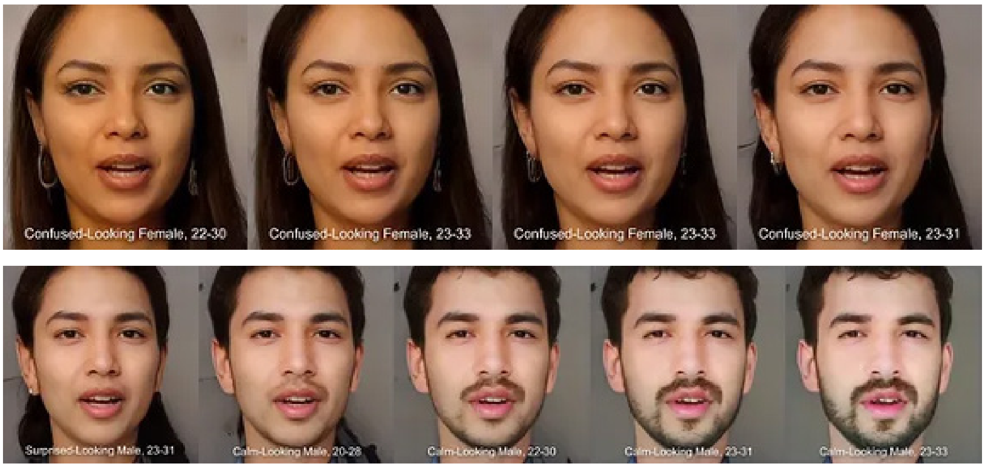


Figure 5: Sequence of progressively gender-varied *Deep Hysteria* raw images with Amazon Rekognition-generated emotion/gender/age labels.

To create the *Deep Hysteria* exhibition images, selected male images identified as "calm" were placed side by side with counterpart non-binary or female images identified with more stereotypically feminine emotions. Each side-by-side image was captioned with its emotion, gender, and age as identified by Rekognition.

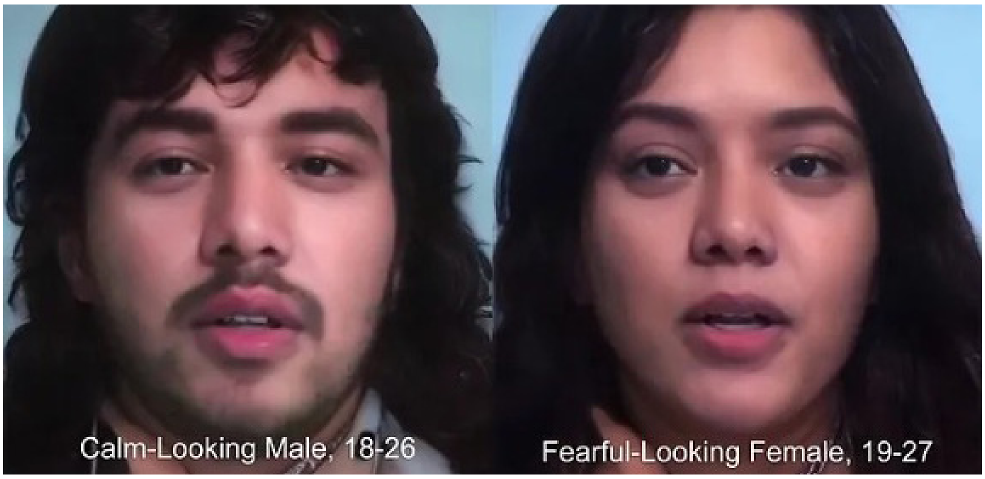


Figure 6: *Deep Hysteria* generated artwork.

Figure 7: *Deep Hysteria* generated artwork.

Figure 8: *Deep Hysteria* generated artwork.

Rough Statistics across the Broader Dataset

The statistics listed below are based on an informally collected/generated dataset. The sample size is small, and the generation and collection methodologies are not designed with the necessary rigor for scientific research. In particular, the broad gender spectrum of generated images sometimes exhibits apparent racial shifts and other anomalies at the extremes that could impact accuracy in some cases. Overall, however, they do give a general indication of the bias observed in the development of the *Deep Hysteria* artworks.

Table 1: Rekognition analysis of generated neutral expression faces — male.

Total identified as male:	**112**
calm	69.64%
undefined mood	6.25%
happy	5.36%
confused	10.71%
surprised	8.04%

Table 2: Rekognition analysis of generated neutral expression faces — female.

Total identified as female:	**119**
calm	51.26%
confused	12.61%
undefined mood	14.29%
disgusted	1.68%
sad	8.40%
surprised	4.20%
happy	7.56%

Smile!

The original stills of actual YouTube video makers used in the training dataset were generated were also analyzed by Rekognition over the course of a year, as part of the *What the Robot Saw* (Alexander 2020) live stream artwork. These images were submitted to Rekognition as raw images without algorithmic variations. Unlike the generated images analyzed, which were limited to neutral expressions, these images incorporate the actual range of vlogger facial expressions. The following statistics summarize the apparent emotions

Rekognition identified in the actual vloggers. Notable in these statistics: female vloggers were much more likely to be labeled as "happy." An informal visual analysis by the artist found that most images labeled as "happy" did indeed appear to be presenting as "happy" — i.e., more women smiled in their videos. This observation appears to support the assumption that women perceive more social expectation than men to smile in their public online presentation.

Table 3: Rekognition analysis of actual YouTuber faces (any expression) — male.

Total identified as male:	**4028**
confused	11.47%
angry	1.91%
fear	1.54%
disgusted	0.94%
calm	58.76%
sad	3.08%
happy	12.88%
surprised	9.41%

Total identified as female:	**3910**
confused	6.04%
angry	0.61%
fear	4.48%
disgusted	1.30%
calm	48.08%
sad	5.65%
happy	23.66%
surprised	10.18%

Conclusion

Much-needed attention has been paid in recent years to the problematic impacts of deep learning systems that reflect and ultimately amplify social bias in training data. However, it's essential to keep in mind that the biases originate with the humans, not in the algorithms themselves. "Biased" algorithms have the potential for positive social impact as well as negative. They can be redeployed in the service of revealing and interrogating deeply embedded social biases we might not otherwise be able, or willing, to see.

References

Alexander, Amy. 2020. *What the Robot Saw*. Accessed January 15, 2023. https://what-the-robot-saw.com/

Amazon. n.d. *Detecting and analyzing faces*. Accessed January 15, 2023. https://docs.aws.amazon.com/rekognition/latest/dg/faces.html

Amazon. n.d. *Emotion*. Accessed January 15, 2023. https://docs.aws.amazon.com/rekognition/latest/APIReference/API_Emotion.html

Boesch, Gaudenz. n.d. *AI Emotion and Sentiment Analysis With Computer Vision in 2023*. Accessed January 15, 2023. https://viso.ai/deep-learning/visual-emotion-ai-recognition/

Bogen, Miranda. 2019. *All the Ways Hiring Algorithms Can Introduce Bias*. May 6. Accessed March 25, 2023. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, in Proceedings of Machine Learning Research 81: 77-91.

Callahan, Molly. 2021. *You can't determine emotion from someone's facial movements– and neither can AI*. August 20. Accessed January 15, 2023. https://news.northeastern.edu/2021/08/20/you-cant-determine-emotion-from-someones-facial-expression-and-neither-can-ai/

Crawford, Kate. 2021. *Artificial Intelligence is Misreading Human Emotion*. April 27. Accessed January 15, 2023. https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

Grossman, Samantha. 2019. *The insidious sexism of 'resting bitch face'*. January 10. Accessed January 15, 2023. https://theweek.com/articles/815496/insidious-sexism-resting-bitch-face

Le Mau, Tuan, Katie Hoemann, Sam H. Lyons, Jennifer Fugate, Emery N. Brown, Maria Gendron, and Lisa Feldman Barrett. 2021. "Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs." *Nature communications* 12 (1): 1-13.

Luxemburg, Robert. 2019. Twitter / @robertluxemburg: #StyleGAN2 latent directions... December 17. Accessed 15 January, 2023. https://twitter.com/robertluxemburg/status/1207087801344372736

McVean, Ada. 2017. *The History of Hysteria*. July 31. Accessed January 15, 2023. https://www.mcgill.ca/oss/article/history-quackery/history-hysteria

Simonite, Tom. 2019. *Amazon Says It Can Detect Fear on Your Face. You Scared?* August 19. Accessed January 15, 2023. https://www.wired.com./story/amazon-detect-fear-face-you-scared/

Smith, Rosa Inocencio. 2016. *The Sexism of Telling Women to Smile.* October 4. Accessed 15 January, 2023. https://www.theatlantic.com/category/the-sexism-of-telling-women-to-smile-your-stories/