“Don’t Worry, Be Happy”: Resisting an AI Emotion Recognition System With a Smile

Avital Meshi

Performance Studies Graduate Group
University of California, Davis Davis,
CA, USA
ameshi@ucdavis.edu

Angus G. Forbes

Department of Computational Media
University of California, Santa Cruz
Santa Cruz, CA, USA
angus@ucsc.edu

ABSTRACT

Don’t Worry, Be Happy is an art performance which demonstrates the problematic nature of AI recognition algorithms when used as instruments of the law. The performance associates the use of AI recognition algorithms with the use of the notorious electric chair. Both technological tools were integrated into juridical systems before being fully understood and both mediate a direct contact between the body and the law. While AI based recognition and classification of faces can be used for the common good, they can also be dangerous when used to control and criminalize people. The artwork emphasizes the entanglement between the body and the machine and illustrates how a performative engagement facilitates a sense of agency when we are confronted with recognition algorithms.

CCS CONCEPTS

• Applied computing → Media arts.

KEYWORDS

Creative AI, emotion recognition, machine learning, performance.

ACM Reference Format:

Avital Meshi and Angus G. Forbes. 2021. “Don’t Worry, Be Happy”: Resisting an AI Emotion Recognition System With a Smile. In *10th International Conference on Digital and Interactive Arts (ARTECH 2021)*, October 13–15, 2021, Aveiro, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3483529.3483725>

1 INTRODUCTION

The artist is strapped to an electric chair. Her face is constantly scanned by an artificially intelligent (AI) emotion recognition system. As long as the system detects her as ‘Happy’ she is safe. However, if any other emotional expression is detected, her body is electrocuted. She smiles relentlessly in order to avoid the punishment that is delivered by the machine. This Kafkaesque scenario describes the details of the art performance *Don’t Worry, Be Happy*, which aims to shine a spotlight on the problematic nature of using AI recognition algorithms as tools of the law. *Don’t Worry, Be Happy* follows a history of endurance art performances which involve

Figure 1: The artist is strapped to an electric chair while an emotion recognition system scans her facial expressions.

hardship to the performer’s body. The artist sets up rules which correspond to pain in order to manifest an irresolvable conflict. These often deeply troubling acts are aimed to visualize the marks of cultural authority and institutional power on the body [36]. The body in *Don’t Worry, Be Happy* is in direct contact with two controversial technological advancements which have been embraced by the law: the electric chair and the AI algorithm.

The increasing prevalence of AI algorithms used as tools of the law is controversial. The potential of reducing crime through predictive software systems is appealing to law enforcement agencies, and they have been integrated into various juridical systems [45]. However, these tools are also potentially harmful [2]. Civil-rights activists and researchers in algorithmic bias are calling to abolish the ‘tech-to-prison’ pipeline, pointing out the problematic nature of the underlying data used for identifying and predicting criminality, and calling attention to the growing crisis of algorithmic validity. A recent letter, which was published as a response to AI research claiming the ability to predict criminality, notes instances of algorithmic structural bias which reveal the propensity of algorithms to amplify historic forms of discrimination and to cause harm when integrated into real world applications. Concluding with the question “Who will be adversely impacted by the integration of machine learning within existing institutions and processes?”, the letter addresses the power relations which accompany the use of the technology [14]. This letter joins a growing body of work which describes the dangers of AI when used to control and criminalize people [7, 8, 31].

This work is licensed under a Creative Commons Attribution International 4.0 License.

ARTECH 2021, October 13–15, 2021, Aveiro, Portugal
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8420-9/21/10.
<https://doi.org/10.1145/3483529.3483725>

Figure 2: The artist's arms are strapped to the chair. Electrode patches are attached to both her forearms.

The process in which technology is deployed as an instrument of the law before its reliability is fully understood was also demonstrated with the electric chair. First used in the United States in 1890, the electric chair was introduced as a humane alternative to hanging. Its promise was to deliver an instantaneous and painless death. However, many of the people who were executed while sitting on this chair died instead in a slow and agonizing death rather than a quick and efficient one [9]. While abolished in many countries, capital punishment is retained in some countries, including the United States, which continues to use the electric chair in the present day [15].

In the *Don't Worry, Be Happy* project, AI algorithms are associated with the electric chair to emphasize the technological mediation between the body and the law. With the electric chair the body is physically strapped and electrocuted, whereas with AI algorithms this viscosity is not as obvious. Nevertheless, the body's physicality is given by its external appearance, which is virtually harnessed by the system as it is detected and classified. Following Hannah Arendt's notions of appearance, it is precisely through the perspective of recognition algorithms that being and appearing coincide [3]. With both technologies the body becomes a biopolitical medium. As described by Michel Foucault, "power-relations have an immediate hold on it; they invest it, mark it, train it, torture it, force it to carry out tasks, to perform ceremonies, to emit signs" [22].

Power relations between the body and the machine are depicted in the performance by means of post-humanist entanglement; a dynamic process of 'intra-activity' allows meanings and patterns of marks on bodies to be constantly reconfigured. Agency emerges within this entanglement so that the meaning of 'emotions' cease to function as groups of words, but rather they become an ongoing performance of real world outcomes [5]. The machine switches an electric relay while the artist controls her apparent behavior and smiles. The interaction closes a loop in which a smile liberates the body from the punishment that is delivered by the machine. Yet, under the threat of getting shocked, for how long can the artist perform this exaggerated facial expression so that the machine continues to 'read' her as 'Happy'?

2 IMPLEMENTATION DETAILS

Don't Worry, Be Happy is a durational performance, a work of endurance art. The artist is strapped to a wooden armchair modeled after the electric chair (Fig. 1). Electrode patches, connected to a Transcutaneous Electrical Nerve Stimulator (TENS) unit, are attached to her forearms (Fig. 2). This unit is connected to a computer that hosts a pre-trained emotion recognition model which classifies facial expressions out of seven described primary emotions [16]. This model was obtained from the online GitHub community [4].

The image of the artist's face is captured and broadcasted in real-time to the computer. Her emotional expression is analyzed and scored so that the emotion with the highest confidence level is displayed on the screen (Figs. 3 and 4). Emotion detection updates every two seconds. The artist is unable to see the outcomes of this analysis. However, she feels them directly on her body, for whenever an emotion which is other than 'Happy' is displayed on the screen, the system delivers a continuous electric stimulation to both her arms. Concurrently, a loud sound is emitted so that viewers know that electrocution took place.

3 CONCEPTUAL OVERVIEW

Emotion recognition systems are designed according to Paul Ekman's psychological research of emotional expressions [17]. The practice of computing emotions followed an idea that computers need to recognize emotions in order to form an intelligible interaction with users [35]. With the rise of AI, the idea of assuming a person's feelings based on their external appearance became attractive to both corporations and governments for obvious reasons [29]. However, the methodology of inferring inner human emotions based on external appearance is problematic, after all, it is easy to agree that even if one is smiling it doesn't necessarily mean that one is happy. A major review which examines the link between facial configuration and inner emotions, critiques this methodology by showing its limited reliability, lack of specificity, and narrow generalizability [6]. Moreover, the use of emotion recognition systems reactivates a long tradition of pseudo-scientific, physiognomical methods that were used in the past to perpetuate racism and discrimination [1, 44].

During the performance, the artist acts out a 'Happy' face in an attempt to avoid electrocution. This gesture is efficient because she understands how the system operates. Unfortunately, in real world applications of AI recognition systems, we rarely know when and where these systems 'see' us and how we are seen through them. Yet, the consequences of being monitored by such systems have the potential to harshly impact our lives. According to the concepts emphasized by Bridle's New Aesthetic, it is essential to become familiar with these systems and learn to better communicate with them. We can wave at the machine and see how it waves back at us [10]. In Cloninger's Manifesto for a Theory of the New Aesthetic, we are advised to resist letting these systems dazzle our minds by thinking they know us better than we know ourselves. Instead, "we should spend some time figuring out how these systems flow and function so we can more effectively modulate them (or sabotage them)" [13]. A sincere effort to truly engage with the machine will allow us to understand its genuine nature [42]. This understanding

Figure 3: The artist’s face is analyzed by the AI system in real time. When a ‘Happy’ face is detected electric stimulation is not delivered.

Figure 4: For each detected emotion other than ‘Happy’, the relay switches on and the system delivers an electric stimulation to the electrodes that are attached to the artist’s arms.

will enable us to inhabit our existing reality in a better way and to fit more or less harmoniously with the overall system.

The performance also follows conceptions of techno-feminism which are demanding liberation from a techno-chauvinist culture. In her *Cyborg Manifesto*, Donna Haraway posits that liberation rests on the construction of consciousness and the imaginative apprehension of both oppression and of possibility. Haraway claims that in a cyborg world, people are not afraid of their kinship with machines. The reality of the cyborg exhibits partial identities and contradictory standpoints [23]. One such contradiction can be seen in *Don’t Worry, Be Happy* as it shows the limits of identifications and the use of a stubborn smile to resist a terrifying pathway.

4 RELATED ARTWORKS

Becoming familiar with, engaging with, or ‘tricking’ AI systems are activities found in a few recent artworks. In *Emotion Hero* (2016), participants play a video game to train their emotional expressions as guided by the AI software, improving their score by practicing and continuously changing their emotional expressions [43]. In *Machine Readable Hito* (2017), an intentional performance of a range of facial expressions changes the outcomes of a classification process [32]. In *Faking a Smile Is Easier Than Explaining Why I’m Sad* (2019), viewers observe a dancing avatar as a real-time AI system detects their emotional expressions. The avatar changes its dance style according to the participant’s detected emotion [34]. Similarly, in *Classification Cube* (2019), participants engage with an AI system to perform their appearance and behavior and to change the outcomes of their classifications [30]. These artworks reveal subject relations between the body and the machine and examine the impact of a performative engagement on this relationship.

Other artworks criticize the use of AI systems for security and surveillance. *Zoom Pavilion* (2015) uses surveillance cameras and facial recognition algorithms to detect participants and record their spatial relationship within the exhibition space [28]. In *URME* (2013), wearing a 3D printed mask of the artist’s face enables them to

evade surveillance [37]. *CV Dazzle* (2010–2017) examines fashionable makeup designs as camouflage from AI [24]. In the interactive film *Riot AI* (2018), a potential use of emotional recognition systems during a face-to-face confrontation with a riot cop is suggested. Viewers’ emotional expressions are scanned by what seems to be a police camera. Detected emotions are then used to influence the outcome of the confrontation. A scared or an angry viewer will lead to a violent response by the cop while a calm viewer will lead to a more relaxed reaction [33].

The electrical manipulation of the body as a performative act emphasizes the materiality of the body and its vulnerability. In *Prelude to 220 or 110*, performed by Chris Burden, the artist lay on the floor near buckets of water in which 110-volt lines had been submerged. Any spill would have probably lead to a fatal electrocution of the artist’s body [11]. Such performances also portray the body as a kinematic system whose motions can be steered by means of electrical signals [20]. In Stelarc’s performances, viewers control the artist’s muscle movements through local touch screens or remotely via a web interface [38, 39]. Stelarc also integrates his body into cybernetic systems in which visual cues from the internet move his body involuntarily [40, 41]. Similarly, Arthur Elsenaar, in his piece *r-emote* (1995), allows remote viewers to trigger his facial muscles and put together different facial expressions [18]. Elsenaar also uses a computer program to examine all the possible non-spontaneous facial expressions which can be elicited by an algorithmically controlled manipulation [19].

An artistic representation of the electric chair is most famously depicted in Andy Warhol’s series of silk screens showing an unoccupied electric chair in an empty room. The electric chair in this series, in addition to a means for punishment, can also be read as an instrument of torture that monumentalizes the cruelty of a corrupt justice system [27]. The emptiness of the chair represents the condemned as no one and as everyone at the same time. Viewers of this piece become spectators and witnesses. They are invited to contemplate which people are sentenced in our justice system, and

for what crimes they are sentenced [12]. Through technologically-mediated durational performance, *Don't Worry, Be Happy* provides an opportunity for the contemplation of contemporary tools of justice.

The use of a smile as a visual element is rarely depicted throughout the history of art, its absence from portraiture relates to the time it took to paint a portrait; holding a smile for this long turns it quickly into an uncomfortable grimace [25]. This is also true for early photography portraiture which took guidance from pre-existing customs in painting [21]. The widespread practice of smiling to the camera in contemporary portraiture is suggested to be internalized by the public as part of the informational environment of the technology that produced it. Thus, the “Say cheese!” smile is a learned, symbolic behavior in which individuals willingly subjugate themselves to the hegemonic culture that is created by the technology [26]. *Don't Worry, Be Happy* aims to illustrate this same McLuhanesque behavioral change in relation to AI recognition algorithms.

5 CONCLUSION

AI recognition algorithms can be used for the common good. However, it is important to acknowledge that they can also be dangerous when used to control and criminalize people in an unethical way. *Don't Worry, Be Happy* is an exaggerated visualization of the encounter between the body and an AI system, fictionally set to resemble an administration of the law. The artwork aims to illustrate the problematic use of emotion recognition systems as tools of the law. Viewers are invited to contemplate the implications of the use of such algorithms within the justice system and to consider the opportunity to resist this reality by utilizing a performative behavior.

ACKNOWLEDGMENTS

The authors wish to thank Ofer Meshi for his invaluable help in integrating the AI system used in this artwork.

REFERENCES

- [1] Blaise Agüera y Arcas, Margaret Mitchell, and Alexander Todorov. 2017. Physiognomy's New Clothes. <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>.
- [2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias. *ProPublica* 23, 2016 (2016), 139–159.
- [3] Hannah Arendt. 1981. *The life of the mind: The groundbreaking investigation on how we think*. Houghton Mifflin Harcourt.
- [4] Omar Ayman. 2018. Emotion Recognition. <https://github.com/omar178/Emotion-recognition>.
- [5] Karen Barad. 2003. Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society* 28, 3 (2003), 801–831.
- [6] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68.
- [7] Ruha Benjamin (Ed.). 2019. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Duke University Press.
- [8] Ruha Benjamin. 2019. *Race after technology: Abolitionist tools for the New Jim Code*. Polity.
- [9] Craig Brandon. 2016. *The electric chair: An unnatural American history*. McFarland.
- [10] James Bridle. 2011. Waving at the Machines. Video recording and transcript of keynote given at *Web Directions South*, <http://www.webdirections.org/resources/james-bridle-waving-at-the-machines>.
- [11] Chris Burden. 1971. Prelude to 220, or 110. Art Performance.
- [12] Bennett Capers. 2006. On Andy Warhol's Electric Chair. *California Law Review* 94 (2006), 243.
- [13] Curt Cloninger. 2012. Manifesto for a Theory of the 'New Aesthetic'. *Mute* 3, 4 (2012).
- [14] Coalition for Critical Technology. 2020. Abolish the #TechToPrison Pipeline. <https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16>.
- [15] Travis Dorman, Natalie Allison, and Adam Tamburin. 2020. Tennessee execution: Nicholas Todd Sutton executed by electric chair. *The Tennessean* (20 February 2020). <https://www.tennessean.com/story/news/crime/2020/02/20/nicholas-todd-sutton-execution-tennessee-electric-chair/4806710002/>.
- [16] Paul Ekman, Tim Dalgleish, and Mick Power. 1999. *Handbook of cognition and emotion*. Wiley.
- [17] Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124.
- [18] Arthur Elsenaar. 1995. r-emote. Art Performance.
- [19] Arthur Elsenaar. 1997. The Varieties of Human Facial Expression. Art Performance.
- [20] Arthur Elsenaar and Remko Scha. 2002. Electric body manipulation as performance art: A historical perspective. *Leonardo Music Journal* 12 (2002), 17–28.
- [21] Merrill Fabry. 2016. Now you know: Why do people always look so serious in old photos? *Time Magazine* (28 November 2016).
- [22] Michel Foucault. 1990. *The History of Sexuality Volume 1: An Introduction*. Vintage.
- [23] Donna Haraway. 2006. A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *The international handbook of virtual learning environments*. Springer, 117–158.
- [24] Adam Harvey. 2010-2017. CV Dazzle. Artwork.
- [25] Nicholas Jeeves. 2013. The serious and the smirk: The smile in portraiture. *The Public Domain Review* (18 September 2013).
- [26] Christina Kotchemidova. 2005. Why we say “cheese”: Producing the smile in snapshot photography. *Critical Studies in Media Communication* 22, 1 (2005), 2–25.
- [27] Steven Kurtz. 1997. Uneasy flirtations: The critical reaction to Warhol's concept of the celebrity and of glamour. In *The Critical Response to Andy Warhol*, Alan R. Pratt (Ed.), 250–259.
- [28] Rafael Lozano-Hemmer and Krzysztof Wodiczko. 2015. Zoom Pavilion.
- [29] Andrew McStay. 2020. Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society* 7, 1 (2020), 2053951720904386.
- [30] Avital Meshi and Angus G Forbes. 2020. Stepping inside the Classification Cube: An intimate interaction with an AI system. *Leonardo* 53, 4 (2020), 387–393.
- [31] Cathy O'Neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- [32] Trevor Paglen. 2017. Machine Readable Hito. Artwork.
- [33] Karen Palmer. 2018. Riot AI. Artwork.
- [34] Ruth Patir. 2019. Faking A Smile is Easier Than Explaining Why I'm Sad. Artwork.
- [35] Rosalind W Picard. 2000. *Affective computing*. MIT Press.
- [36] Karen Gonzalez Rice. 2016. *Long suffering: American endurance art as prophetic witness*. University of Michigan Press.
- [37] Leonardo Selvaggio. 2015. URME Surveillance: Performing privilege in the face of automation. *International Journal of Performance Arts and Digital Media* 11, 2 (2015), 165–184.
- [38] Stelarc. 1994. Split Body. Art Performance.
- [39] Stelarc. 1995. Fractal Flesh. Art Performance.
- [40] Stelarc. 1996. Ping Body. Art Performance.
- [41] Stelarc. 1997. ParaSite. Art Performance.
- [42] Bruce Sterling. 2012. An essay on the New Aesthetic. *Group* 1 (2012), 1–4.
- [43] Ruben van de Ven. 2017. Emotion Hero. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 422–423.
- [44] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kazi-unas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. *AI Now report*. AI Now Institute at New York University.
- [45] Xiaolin Wu and Xi Zhang. 2016. Automated inference on criminality using face images. *arXiv preprint arXiv:1611.04135* (2016), 4038–4052.