Stepping Inside the *Classification Cube*: An Intimate Interaction with an AI System

Avital Meshi and Angus G. Forbes

ABSTRACT

The *Classification Cube* art installation invites participants to become familiar with a machine-learning classification system. Inside a private space within the gallery, participants' bodies are subjected to a classification process that detects their faces and estimates their age, gender, emotion and actions. Participants are also able to see how their own classification compares with how the installation classifies a series of animated figures. Rapidly changing results encourage participants to actively perform their behavior to the system and alter the way it "sees" them. The entanglement with the system raises awareness regarding the effectiveness of machine interpretation.

Avital Meshi
Artist
Digital Arts and New Media
University of California, Santa Cruz
1156 High St
Santa Cruz, CA 95064, U.S.A.
ameshi@ucsc.edu

Angus G. Forbes
Researcher
Department of Computational Media
University of California, Santa Cruz
1156 High St
Santa Cruz, CA 95064, U.S.A.
angus@ucsc.edu

See www.mitpressjournals.org/toc/leon/53/4 for supplemental files associated with this issue.

Machine learning (ML) algorithms are becoming increasingly prevalent in our environment. Embedded in products and services we use on a daily basis, they rely on our personal information, searching for patterns in it and producing corresponding outcomes in return [1]. Mostly we are unaware of this process, and we do not know how these systems "see" us. Yet the outcomes may seriously impact our lives [2–4]. Recent studies demonstrate that some ML algorithms reflect social disparities and biases encoded into the algorithms. Moreover, data sets used for the training of these algorithms often fail to include underrepresented groups and historically marginalized communities [5–7]. Therefore, it feels crucial that individuals familiarize themselves with these systems and see how their own bodies are being detected and classified.

The *Classification Cube* art installation invites viewers to interact with an ML classification system. Inside a seemingly private space within the context of an art gallery, viewers are confronted with two screens. One presents a prerecorded video of a diverse group of animated figures representing realistic yet unusual combinations of body features and movements. On the other screen, viewers see a video feed of their own bodies given by a live webcam that is situated inside the space. Images on both screens are subjected to analysis by an ML classification system whose outcomes rapidly update and are displayed on the screens. This analysis includes a face detection process and estimations of age, gender, emotion and action of the subjected bodies. The immediate capture and analysis of viewers' bodies, along with the comparison to other bodies inside the space, turns the installation into a platform for exploration. Viewers become aware that their appearance serves as the system's input and that corresponding classifications are its output. The comparison between one's own body and the bodies of others suggests how effective or ineffective these classifications may be. The rapid shifting from one classification to another invites viewers to perform different movements inside the space, and to observe how these movements modulate the outcome of their classification: Will the system classify me as younger if I stand straight? Am I classified as "happy" if I smile? Viewers can also mimic the behaviors of the animated figures and investigate whether the system provides a classification similar to or different from the classifications of the animated figures.

Viewers are informed before entering the space that the system does *not* collect or share any information. This approach differs from art projects involving data collection that explicitly share personal data [8]. Given that their visibility through the lens of the system is for their eyes only, participants feel more comfortable engaging with the system in a performative manner.

Classification Cube reveals information that is usually kept obscure. We become aware of the coupling between the information and our own bodies. This understanding opens a broader discussion regarding the way we are seen through the lens of this technology. Assuming these algorithms will eventually dominate the entire public domain, our behaviors will be constantly detected, classified and stored, and our records will be determined accordingly [9]. Hence, figuring out how these systems see us is essential. Furthermore, if we understand that our appearance serves as the system's input, we can potentially learn that we have the

Fig. 1. Participants see an ML analysis of their own body. They can compare their classification with that of other bodies and examine their ability to alter their classification by engaging in a performative behavior. (Photo © David Pace. Artwork © Avital Meshi.)

power to modulate it. Engaging in a performative behavior that changes the outcomes of our classifications suggests an opportunity to manifest ourselves as we wish to be seen. In this sense, *Classification Cube* drafts a futuristic practice space in which we can shape and accomplish a performance that may lead to a radical identity transformation outside of the installation and in the public domain. Figure 1 shows a participant interacting with the *Classification Cube* installation.

Related Work

Inside *Classification Cube*, viewers examine the coupling between the physical body and the corresponding information produced. A similar invitation is demonstrated in *Zoom Pavilion* (2015), in which surveillance cameras use facial recognition algorithms to detect participants and record their spatial relationship within the exhibition space. The outcomes, projected on the walls, show zooming sequences that amplify the images of some participants and disorient the entire image to focus on one person or another. Viewers become subject to surveillance systems, which are usually used for detection and control [10]. In the artwork *Sight Machine* (2018), musicians playing on stage are subjected to the analysis of an algorithmic system. The analysis, projected on a screen above them, reveals their information to the audience. This includes detection of their faces, movements of their bodies and details regarding their gender, age and emotion. Occasionally an image showing the face detection of members of the audience itself reminds viewers that they are subject to that same process [11]. The artwork *Machine Readable Hito* (2017) presents a series of portraits of the artist Hito Steyerl along with captions output from an ML classification stating her age, gender and emotion. Steyerl intentionally performs a range of facial expressions that change the outcomes of her classification [12]. The artwork *Faking a Smile Is Easier Than Explaining Why I'm Sad* (2019) invites the viewer to carry out a similar performance while observing a dancing avatar as a real-time ML classification system detects the viewer's facial expressions. The avatar changes its dance style according to the participant's detected emotion: A happy viewer will make the avatar belly dance, a surprised viewer will make it dance ballet and so on. This work, while revealing subject relations between the body and the machine, also examines our ability to control the system and change its outcomes by performing different behaviors [13].

Other artworks address the oppressive nature of ML interpretation by suggesting ways to avoid detection. *How Not To Be Seen. A Fucking Didactic Educational .mov File* (2013) provides "lessons" in invisibility, offering solutions such as wearing an invisibility cloak or being a female over 50 [14]. In *URME, Personal Surveillance Identity Prosthetic* (2013), participants are offered the opportunity to free themselves from any threat of surveillance by wearing a 3D printed mask of the artist's own face [15]. Similarly, *Face Weaponizing Suite* (2011–2014) displays masks that cannot be detected as human faces [16]. A different approach is examined in *CV Dazzle* (2010–2017), which uses fashionable makeup designs as camouflage [17]. While it is unclear whether these solutions can lead to actual avoidance of detection, it is important to note that in some places wearing masks in the public domain is considered illegal. Also, the attempt to avoid detection may be considered as a suspicious behavior in itself, and recent reports share incidents of people being arrested only for trying to avoid detection [18]. Moreover, detection of disguised faces has already become a technological challenge, studied by several research groups who are trying to overcome any kind of disguising element that might interfere with face detection [19,20].

The work of these artists attempts to avoid oppressive visibility and manifests critique against the use of ML-based systems for surveillance and control. Scholar Shoshana Amielle-Magnet claims that when biometric systems fail, they reveal a violent, racist, homophobic, classist and sexist structure that we cannot run away from and that places us inside "cages of information" [21]. The artwork *Face Cages* (2013–2016) features performances of four queer artists wearing metal masks inspired by biometric diagrams. Being extremely painful to wear, the masks emphasize the violent imposition of these diagrams on human faces. The project also highlights how ML systems often fail to accurately recognize many people, especially those from marginalized communities [22].

Conceptual Overview

If ML systems do, in a sense, lock us inside cages of information, how do we set ourselves free? *Classification Cube* shifts the perspective by following a conceptual framework of the "New Aesthetic," which proposes that humans can communicate with machines—we wave at them and then see how they wave back at us. In interrogating machine behavior, we reveal ways in which these systems can be platforms for identity transformations [23]. In *Manifesto for a Theory of the New Aesthetic*, Curt Cloninger advises us to resist letting these systems dazzle our minds. Instead, "we should spend some time figuring out how these systems flow and function so we can more effectively modulate them (or sabotage them)" [24]. Inspired by these ideas, *Classification Cube* offers a space in which we can interact with the system to see how it responds to our appearance, posture and behavior. It enables us to learn how the system interprets us, and it gives us insight into how we might influence this interpretation (Fig. 2).

Implementation Details

Classification Cube is a cubical construction (10 × 10 × 10 ft) wrapped with an opaque, spandex fabric and lit with LED light strips. Its narrow entrance is designed to encourage viewers to step in, but also suggests the private nature of this environment (Fig. 3). Inside the space are two side-by-side vertical monitors. One shows a prerecorded looping video of animated figures. The other monitor is connected to a webcam inside the cube and presents a live video feed of the interior space.

The algorithms used in the *Classification Cube* installation are off-the-shelf, pretrained models derived from computer vision algorithms developed and made freely available by the open source community. The rationale behind using these particular algorithms is that they are embedded in a wide range of commercial products and services, and through their use we enable participants to overtly experience this classification process. The algorithms include a face detector, which displays a green bounding box around any automatically detected faces [25], and age and gender detection models that estimate an age from 0 to 100 and classify gender according to the binary "female" or "male" gender labels [26,27]. We also included an emotion recognition model that detects and classifies facial expressions as one of seven primary emotions (angry, sad, disgusted, scared, happy, surprised or neutral) [28]. Lastly, an action recognition model recognizes bodily actions and classifies them according to a list of 400 distinct human action classes, such as jumping, dancing and kicking, among many others [29,30]. Estimation of age, gender, emotion and action are all clearly displayed using four lines of red text, showing the estimated labels and the confidence level for each estimation. All of the models in this system were assembled into a single codebase that provides the analysis outcomes while processing the webcam's live feed of user motion in real time; the codebase also produces the analyzed video files for the animated figures (as shown in Fig. 4).

The animated figures shown inside the cube are created using Adobe Fuze CC and the Mixamo online service. In order to create a diverse representation, all body elements, clothes, textures and animations are randomly chosen in the assembly

Fig. 2. A performative engagement with the system inside *Classification Cube* allows viewers to better understand how the system "sees" them. Once a viewer realizes that the appearance of their body serves as one of the system's inputs, they understand that they can influence its results. (Photo and Artwork © Avital Meshi)

process of each figure. This randomization is meant to challenge the ML classification system with diverse and unusual-looking bodies. This diversity is also presented to viewers so they can compare the classification of their own bodies with that of many different bodies (Fig. 1). Unfortunately, this diversity cannot be claimed to be comprehensive, as the character creation platform excludes certain body types, such as those with functional diversity or nonbinary gender types. This is also true for the classification system itself, which offers a limited set of labels and ignores gender complexities, emotional diversity and a wide range of possible human behaviors and actions.

Audience Reception

Classification Cube was shown at a group exhibition at the University of California, Santa Cruz, in spring 2019. The installation received over 2,000 visitors over a two-and-a-half-week period, and we hosted critique sessions for students enrolled in art, art history, media studies, computational media and computer science courses. Because we wanted to ensure a safe space in which users felt free to interact with the classification system without being surveilled, documented or analyzed, we did not conduct a formal evaluation of user behavior within the installation. Nevertheless, the piece generated much discussion about sociotechnical systems and data ethics, and participants were eager to share thoughts about their experience inside it. For example, one viewer noted that her gender classification was always "wrong." She said, "I tried manifesting typical feminine gestures but the system kept classifying me as 'male.' I should have applied makeup to convince it that I'm a female!" Similar responses were common among female viewers who claimed that the system misgendered them repeatedly. In contrast, gender classification among male viewers seemed to be working as expected, especially for those with facial hair. A few viewers laughed about the confidence level of the male classification, noting that without facial hair, the confidence level of the male classification was about 60–70 percent, but with facial hair it was above 90 percent. They joked that having a beard makes you more "manly." One viewer, who identifies as nonbinary,

Fig. 3. The cube is designed as an immersive space that invites viewers to step into a different environment where they can become intimately engaged with an ML system. The white, futuristic, glowing space removes viewers out of any other context and allows them to focus on their own private experience. (Photo © David Pace. Artwork © Avital Meshi.)

commented on the fact that the system offers only a reductionist view of gender and therefore does not have the capacity to classify them accurately. Regarding age classification, viewers reported that the system classified their age as either younger than they are (mostly said with a smile) or older than they are (mostly said with a frown). A viewer who was eating a slice of pizza inside the space said that the system classified her as "playing the harmonica." Lastly, another viewer, who was rather happy, reported that the emotion classifier repeatedly labeled her as "angry." She said, "I turned angry to see that I was classified as 'angry.'"

We solicited feedback from artists, art historians and game studies faculty during the creation of the installation; we developed the installation through an iterative design process that was informed by critical discussion about the role of art in the information age. Scholarly critique noted that the innovation of this project is manifested through the invitation to form a direct playful interaction between the viewer and the classification system. The simple and immediate nature of this interaction turns the installation into an accessible, open-ended platform for exploration that is not available to us in the public domain. The white space of the cube corresponds with the traditional layout of the art gallery. It places viewers inside a "nonspace," an artificial, clean simulacrum where they can devote their attention to examining the ML aesthetics. The design of the space was also compared with that of virtual caves, photo booths and theater stages.

Scholarly critique also noted that the side-by-side comparison between the viewer's own body and those of the animated figures was one of the most powerful elements of the installation, explicitly showing viewers how their bodies are prone to be classified (and misclassified). This notion emphasizes the awareness that our entanglement with such systems is not just a personal one; rather, it is relevant to the diverse range of our society. Lastly, ideas of identity transformation were discussed in relation to other technologies that may transform identity, such as biotechnology and virtual reality. Interactive data artworks such as *Classification Cube* can play a valuable role in contemporary society, giving participants the ability to perform new identities and to test how they will be interpreted. This aspect of *Classification Cube* encourages

Fig. 4. A diverse group of animated figures shown inside the cube allows viewers to examine classifications of bodies other than their own and to understand the effectiveness of the system. Viewers can also mimic the behavior of these figures while performing their own behaviors. (Animation © Avital Meshi)

participants to think about current and future classification technologies, and the way our interactions with them may shape our culture.

Conclusion

ML technology is becoming more and more apparent in our environment. While our visibility through the lens of this technology can be an oppressive one that subjects us to biases and discrimination, this technology may also provide an opportunity for an empowering visibility. *Classification Cube* is a space in which we can become familiar with the system and play and interact with it. As we begin to understand how it "sees" us and others, we have the opportunity to influence it by performing different behaviors in front of it. Unlike similar systems in the public domain, this system does *not* collect or share any information. It is designed as a private, neutral space that presents a safe environment in which one can practice and shape one's performance in response to the algorithms. Assuming a future of constant surveillance and ubiquitous interpretation by machine learning systems, *Classification Cube* serves as a futuristic mirror, reflecting how we are detected and classified in an increasingly digitized public domain.

Acknowledgments

The authors wish to thank Ofer Meshi for his invaluable help in designing and integrating the ML systems used in this artwork, and Edward A. Shanken, Katherine Isbister, Marianne Weems and Micha Cárdenas for their advice and support during the development of this project.

References and Notes

1. Sumit Das et al., "Applications of Artificial Intelligence in Machine Learning: Review and Prospect," *International Journal of Computer Applications* **115**, No. 9 (2015).
2. Julia Angwin et al., "Machine Bias," *ProPublica*, 23 May 2016.
3. Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* **89** (2014) p. 1.
4. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, Broadway Books, 2016).
5. Tolga Bolukbasi et al., "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings," *Advances in Neural Information Processing Systems (Proceedings of NeurIPS 2016)*, 2016, pp. 4349–4357.
6. Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," presented at the Conference on Fairness, Accountability, and Transparency (New York: 2018).
7. Aylin Caliskan, Joanna J. Bryson and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," *Science* **356**, No. 6334, 183–186 (2017).
8. George Legrady and Angus G. Forbes, "Data in Context: Conceptualizing Site-Specific Visualization Projects," *Leonardo* **50**, No. 2, 200–204 (2017).
9. O'Neil [4].
10. Rafael Lozano-Hemmer and Krzysztof Wodiczko, *Zoom Pavilion*, interactive installation, 2015.
11. Trevor Paglen, *Sight Machine*, multimedia performance, 2018.
12. Trevor Paglen, *Machine Readable Hito*, photographic portraits with metadata, 2017.
13. Ruth Patir, *Faking a Smile Is Easier Than Explaining Why I'm Sad*, mixed media installation, 2019.
14. Hito Steyerl, *How Not to be Seen: A Fucking Didactic Educational .MOV File*, short film, 2013.
15. Leo Selvaggio, *URME Personal Surveillance Identity Prosthetic*, 3D printed mask of the artist's face, 2013.
16. Zack Blas, *Face Weaponizing Suite*, collective workshop-generated 3D masks, 2011–2014.
17. Adam Harvey, *CV Dazzle*, facial-recognition-blocking fashion pieces, 2010–2017.
18. Lizzie Dearden, "Police Stop People from Covering Their Face from Facial Recognition Camera Then Fine Man £90 after He Protested," *Independent*, 31 January 2019.
19. Vineet Kushwaha et al., "Disguised Faces in the Wild," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1–9.
20. Amarjot Singh et al., "Disguised Face Identification (DFI) with Facial Keypoints Using Spatial Fusion Convolutional Network," *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1648–1655.
21. Shoshana Magnet, *When Biometrics Fail: Gender, Race, and the Technology of Identity* (Duke Univ. Press, 2011).
22. Zack Blas, *Face Cages*, installation and performance, 2013–2016.
23. James Bridle, "Waving at the Machines," keynote given at Web Directions South conference, 2011. Video recording and transcript available at www.webdirections.org/resources/james-bridle-waving-at-the-machines (accessed 9 April 2020).
24. Curt Cloninger, "Manifesto for a Theory of the 'New Aesthetic,'" *Mute* **3**, No. 4, 3 (2012).
25. Yixuan H.Y. Hu, Tensorflow Face Detector, software tool. www.github.com/yeephycho/tensorflow-face-detection (accessed 9 April 2020).
26. Rasmus Rothe, Radu Timofte and Luc Van Gool, "Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks," *International Journal of Computer Vision* **126**, No. 2–4, 144–157 (2018).
27. Yusuke Uchida, Age and Gender Estimation, software tool. www.github.com/yu4u/age-gender-estimation (accessed 9 April 2020).
28. Omar Ayman, Emotion-Recognition, software tool. www.github.com/omar178/Emotion-recognition (accessed 9 April 2020).
29. Joao Carreira and Andrew Zisserman, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.
30. Joao Carreira and Andrew Zisserman, I3D Models Trained on Kinetics. www.github.com/deepmind/kinetics-i3d (accessed 9 April 2020).