
DOOR: Ethnicity in Artificial Intelligence

Tomas Laurenzo

School of Creative Media
City University of Hong Kong
18 Tat Hong Avenue
Kowloon Tong, Hong Kong
tomas@laurenzo.net

Katia Vega

Department of Design
University of California, Davis
One Shield Avenue
Davis, CA 95616 USA
kvega@ucdavis.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).
TEI '19, March 17–20, 2019, Tempe, AZ, USA
ACM 978-1-4503-6196-5/19/03.
<https://doi.org/10.1145/3294109.3301262>

Abstract

DOOR is an artwork that aims at exposing some of the social and political impact of artificial intelligence, computer vision, and automation. The project uses a commercially available computer vision system that predicts the interactor's ethnicity, and locks or unlocks itself depending on this prediction. The artwork showcases a possible use of computer vision making explicit the fact that every technological implantation crystallises a political worldview.

Author Keywords

artificial intelligence; computer vision; ethnicity; interactive art;

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); Human computer interaction (HCI); *Haptic devices*; Human computer interaction (HCI); •Applied computing → *Media arts*;

Introduction

Machine learning (ML) algorithms find complex relationships in data, recognising patterns that humans may not be able to identify, or not know how to formally express for explicit rule-based programming. This allows to create systems with unpredictable, yet somehow controllable behaviour using training datasets. While this dataset-based

control is where its greatest strength resides, it is also very dangerous, for it often introduces biases through over-fitting or under-fitting, and, more importantly, learns biases already existing in the training data. These dangers are even more amplified as these systems are relatively difficult to debug [13].

The translation of biases present in the datasets brings extremely worrying social problems present in the widespread adoption of ML technologies, as these ML-constructed models of information do have an enormous social impact on the technologies they fuel [7].

The social concerns of ML, however, transcend the involuntary transfer of biases present in training data into ML models, and include the direct effects that the insertion of “Intelligent” systems has in society. For example, AI systems will generate new economic value by changing existent products and services, by creating new ones, as well as by reducing costs through automation. However, it remains to be seen how this will be distributed, a “key concern for economic inequality” [2].

The insertion of AI in everyday life has dramatic social implications that have not been explored enough. However, algorithms implemented had already affected minorities: predicting criminal reports in areas with higher proportion of people from racial minorities in Oakland [12]; misunderstanding gender for darker-skin women by popular AI models [4]; police from Israel arrested a man for a wrong translation made by Facebook automatic translation software [6]. Similarly, a system widely used in the US to guide sentencing predicted black defendants have a higher risk of recidivism than white defendants [18].

It is easy to see that ML is used to automate decisions that affect people's lives, where systematically biased systems

and systems that violate people's rights to privacy are frequently proposed and implemented.

Background

Race has always played an important role in society. In the United States, for example, the U.S. Office of Management and Budget approved the collection of data about ethnicity into the American Census.

The race classification started in 1790 incorporating three categories: Slaves, Free White Females and Males, and Other Free Persons. Currently, six categories are included: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and some other races (multiracial, mixed, interracial, or a Hispanic or Latino group) [9]. The collection of these data is reportedly meant to increase the benefits on education, housing, health, and other programs to groups in need.

For example, a school system might design cultural activities that reflect the diversity in their community, or a business to select the merchandise to sell in a new store. On the other hand, 19 countries out of 41 European countries do not collect data on ethnicity, while some European human rights institutions are requesting to reconsider ‘colour-blind’ statistics [19].

Recent advances in computer vision and artificial intelligence, have allowed the creation of systems able to infer (predict) a information on a person from camera data, including facial recognition, facial expressions, ethnicity, among others [16, 3]. Nowadays, companies like as Affectiva [1], Clarify [5] and Haystack [8] provide image processing services that include these predictions, among several others.

In spite of potential benefits that face recognition proposes,

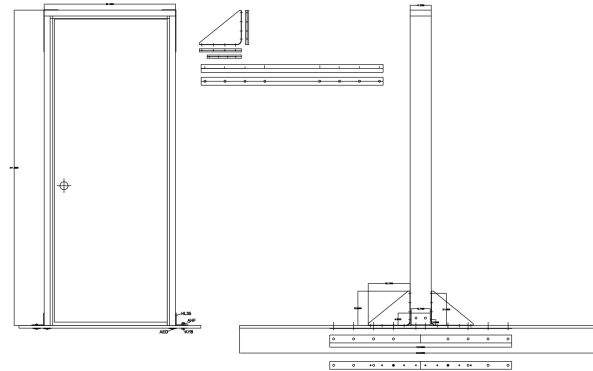


Figure 1: *DOOR* schematics.

its widespread application entails several risks, from privacy breaches to systematic discrimination in areas such as hiring, policing, benefits assignment, marketing, and other purposes [14].

DOOR

DOOR (door frame, door, computer, camera, custom software), is an interactive installation consisting of an intelligent door set up in an open area of the exhibition space (see Fig. 2). The piece reflects on the power asymmetries that technology crystallises, aiming at providing with a reflection on the aesthetics of our relationship with it.

DOOR also aims at showcasing the advancements and limitations in computer vision and machine learning, allowing the public to experience in person its power as well as its inherent biases.



Figure 2: *DOOR* installation. Work in progress.

DOOR embeds a computer vision system that detects and analyses the faces of potential interactors. *DOOR* predicts the interactor's ethnicity in real time and locks or unlocks itself depending on this prediction.

Implementation

DOOR uses a standard webcam (HD, 30FPS) to acquire data on its interactors. Two different implementations have been created, both running at interactive rates on a 2018 MacBook Pro laptop.

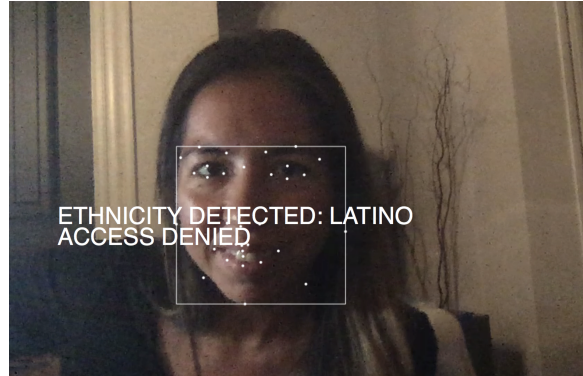


Figure 3: Model predicting ethnicity: Latino.

The first uses a deep neural network based on David Sandberg's FaceNet [17], and trained on the UTKFace dataset [20]. UTKFace contains 23,708 faces with five labelled races (White, Black, Asian, Indian, and Others (e.g. Hispanic, Latino, Middle Eastern)).

A second implementation uses Affectiva.ai's pre-trained model run locally [1]. This service classifies the input images also in five different ethnicities (White, Black, South East Asian, Asian, and Latino).

Once the interactor's ethnicity has been predicted by the system, the door is either kept locked or unlocked, using a standard electric strike and a relay controlled by an Arduino Uno microcontroller.

Conclusion and Future Works

DOOR is an artwork that makes tangible the implications of artificial intelligence in society. The piece reflects on the power asymmetries that technology crystallises. It also aims at showcasing the advancements and limitations in

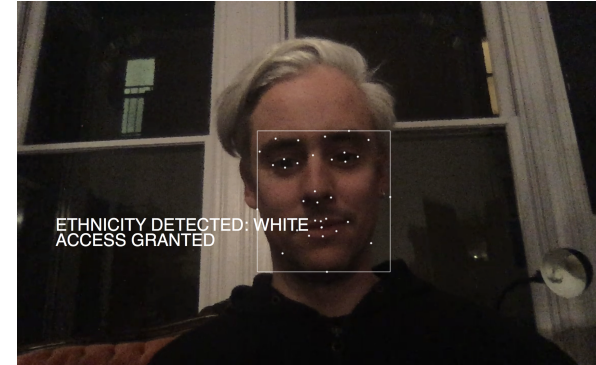


Figure 4: Model predicting ethnicity: White.

computer vision, machine learning, and artificial intelligence, allowing the public to experience in person its power as well as its inherent biases.

Not only the current social impact of AI is undeniable, it is also posed to increase enormously. In effect, the ML revolution resides on the discovery that our intuition of what is resolvable by computers was wrong: where we thought a precise formal modelling was necessary, it is now plausible that an *implicit modelling* like the one provided by machine learning will succeed. By using a commercial model, the piece shows the state of the art of ethnicity prediction.

Knowing that novel technologies can be ahead of ethical and political rules and principles, [10] it becomes necessary to build tools that help informing the discussions.

Art is intrinsically deregulatory, being able to on its own epistemological framework. Yet, the appropriation and re-contextualisation of technology that new media artists perform, situates the practice in a unique position, able to re-

flect on contemporary political issues by re-articulating the technology helping the audience in their construction of new points of view. Interestingly, in DOOR, the artistic fact exists both when its prediction of ethnicity is accurate and when it fails. The piece simultaneously showcases the potentialities of machine learning, while exploring of the aesthetics of machine learning and interaction.

The authors plan con continue the exploration of these subjects. Future works will include other predictions such as gender, sexual orientation, or emotional state.

Acknowledgements

The authors thank Stochastic Labs for supporting the development of the project [11]. The work described in this paper partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. 9042456 (CityU 11674416)]. We also thank Alejandro Rodriguez [15] for his support.

References

- [1] Affectiva. 2018. Affectiva. <https://www.affectiva.com/how/how-it-works/>. (2018).
- [2] Dean Baker. 2014. Can Productivity Growth Be Used to Reduce Working Time and Improve the Standard of Living of the 99 Percent? The Future of Work in the 21st century. *Economic Analysis Research Network—Future of Work Project* (2014).
- [3] Vicki Bruce and Andy Young. 1986. Understanding face recognition. *British journal of psychology* 77, 3 (1986), 305–327.
- [4] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*. 77–91.
- [5] Clarifai. 2018. Clarifai Demographics Model. <https://clarifai.com/models/>. (2018).
- [6] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 797–806.
- [7] Kate Crawford. 2016. Artificial intelligence—white guy problem. *The New York Times* 25 (2016).
- [8] Haystack.ai. 2018. Haystack. <https://www.haystack.ai/>. (2018).
- [9] Karen R Humes, Nicholas A Jones, and Roberto R Ramirez. 2011. Overview of race and Hispanic origin: 2010. (2011).
- [10] David Sun Kong and Nicole Bakker. 2018. Community driven design of living technologies. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2*. ACM, 47.
- [11] Stochastic Labs. 2018. Stochastic Labs — Seed funds and collaborative community. <http://stochasticlabs.org/>. (2018).
- [12] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.
- [13] Anh Nguyen, Jason Yosinski, and Jeff Clune. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 427–436.

- [14] Future of Privacy Forum. 2017. Unfairness y Algorithm: Distilling the Harms of Automated Desicion-Making. <https://fpf.org/wp-content/uploads/2017/12/FPF-Automated-Decision-Making-Harms-and-Mitigation-Charts.pdf>. (2017).
- [15] Alex Dogrush Rodriguez. 2018. Alex Rodriguez. <http://dogrush.com>. (2018).
- [16] Ashok Samal and Prasana A Iyengar. 1992. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition* 25, 1 (1992), 65–77.
- [17] David Sandberg. 2017. Face recognition using tensorflow. *Läst [2018-11-20]*. URL: <https://github.com/davidsandberg/facenet> (2017).
- [18] New Scientist. 2018. How Google and Facebook hooked us and how to break the habit. <https://www.newscientist.com/article/mg23731640-500-how-google-and-facebook-hooked-us-and-how-to-break-the-habit/>. (2018).
- [19] Patrick Simon. 2012. Collecting ethnic statistics in Europe: a review. *Ethnic and Racial Studies* 35, 8 (2012), 1366–1391.
- [20] Song Yang Zhang, Zhifei and Hairong Qi. 2017. Age Progression/Regression by Conditional Adversarial Autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.