Facial Recognition is the Plutonium of AI

It's dangerous, racializing, and has few legitimate uses; facial recognition needs regulation and control on par with nuclear waste.

By Luke Stark

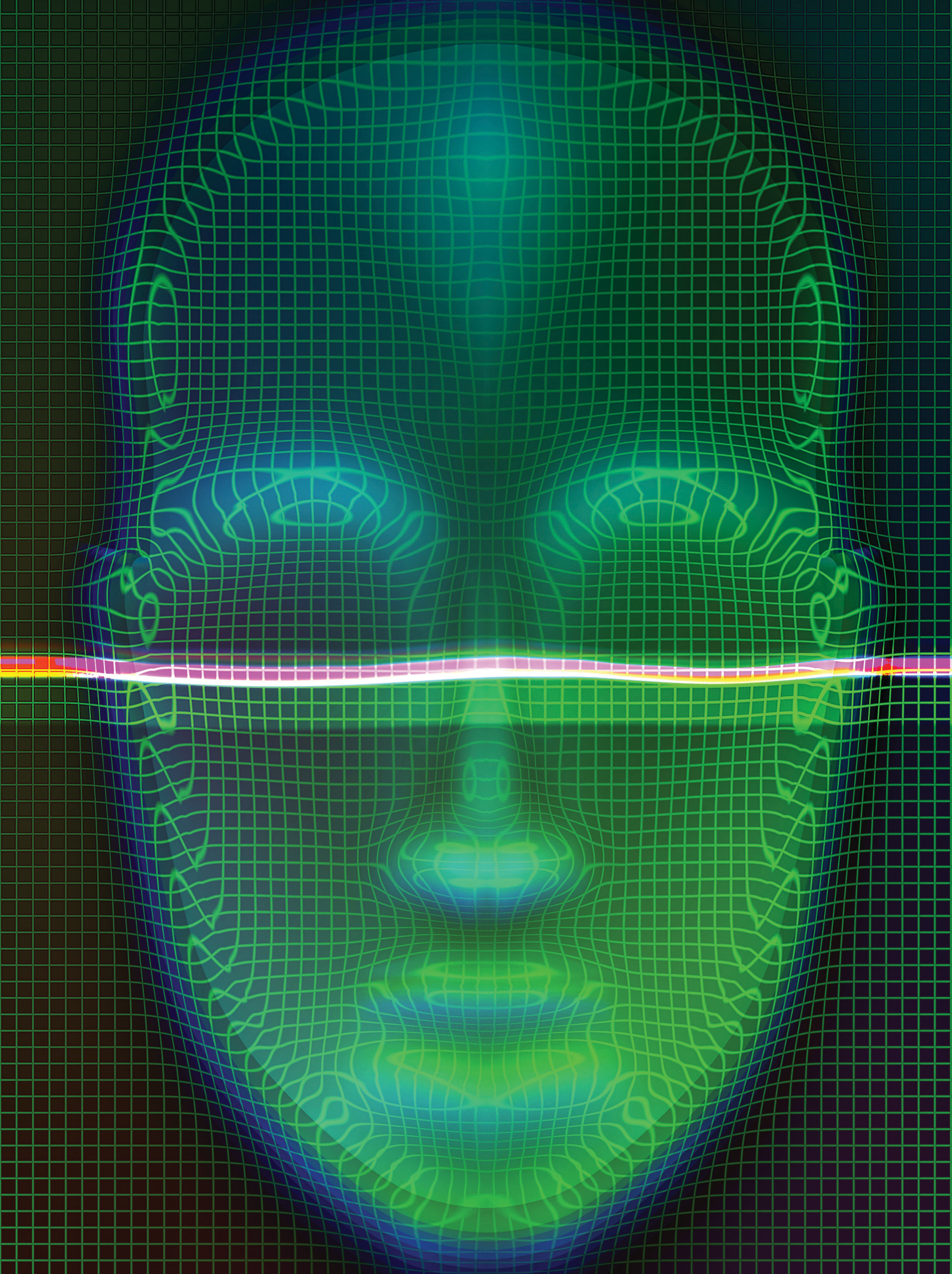
DOI: 10.1145/3313129

When, in 1941, Glenn T. Seaborg and his colleagues at the University of California Berkeley isolated—and subsequently named—plutonium, the radioactive element 93, Seaborg reportedly suggested the periodic symbol Pu for the discovery. According to Seaborg, it "sounded like the words a child would exclaim, 'Pee-yoo!' when smelling something bad" [1]. Plutonium, industrially produced for the American atomic bombs dropped on the Japanese cities of Hiroshima and Nagasaki in August 1945, was ill favored even by its discoverers.

Today, plutonium has very few non-military uses (its application in nuclear weapons being, of course, a moral abomination in itself). Plutonium is produced as a byproduct of uranium-based nuclear power, and is the chief component of nuclear waste; in miniscule amounts, is also used as a power source in specialized scientific instruments, such as aboard space probes. Plutonium has only highly specialized and tightly controlled uses, and poses such a high risk of toxicity if allowed to proliferate that it is controlled by international regimes, and not produced at all if possible.

Plutonium, in other words, is an apt material metaphor for digital facial

Image by Neosiam / Shutterstock.com



recognition technologies: Something to be recognized as anathema to the health of human society, and heavily restricted as a result.

Readers might object that the analogy between plutonium and facial recognition technologies is not just alarmist, but nonsensical. Yet in forthcoming work, the University of Washington's Anna Lauren Hoffmann and I argue the metaphors we use to make sense of digital systems can reveal important similarities between a new technology or practice, and other, older technological problems [2]. By analogizing facial recognition to plutonium, I want to add two broad points to an increasingly lively debate about the risks of facial recognition technologies. First, facial recognition technologies, by virtue of the way they work at a technical level, have insurmountable flaws connected to the way they schematize human faces. These flaws both create and reinforce discredited categorizations around gender and race, with socially toxic effects. The second is, in light of these core flaws, the risks of these technologies vastly outweigh the benefits, in a way that's reminiscent of hazardous nuclear technologies. That is why the metaphor of plutonium is apt. Facial recognition, simply by being designed and built, is intrinsically socially toxic, regardless of the intentions of its makers; it needs controls so strict that it should be banned for almost all practical purposes.

There have been a number of recent warnings about the dangers of facial recognition. Last August, Woodrow Hartzog and Evan Selinger cautioned that facial recognition technology "is the most uniquely dangerous surveillance mechanism ever invented" [3]. Arguing for a total ban, the authors gave several reasons why facial recognition technologies are uniquely dangerous: images of human faces are hard to hide or change; there is an existing store of databases, such as of drivers licenses, matching faces to names; video surveillance mechanisms are cheap and already widespread; and most crucially, faces, unlike other biometric indicators, are central to our personal identities and social lives. We can't escape or permanently hide our faces; as a result, our freedom to exist outside constant surveillance, Hartzog and Selinger write, is threatened by this "menace disguised as a gift."

Facial recognition's racializing effects are so potentially toxic to our lives as social beings that its widespread use doesn't outweigh the risks.

Hartzog and Selinger are right in their concerns, but the problems with facial recognition technology go even further. In a recent article, I laid out some of the reasons why facial recognition is troublesome at the conceptual and technical levels, even if all of the social use cases laid out by Hartzog and Salinger could be satisfactorily solved [4]. That article, which explores facial recognition through the lens of digital animation grounded in systems like Apple's FaceID, is indebted to a number of brilliant scholars of technology and race, including Simone Browne, Wendy Hui Kyong Chun, Lisa Nakamura, Sianne Ngai, and Safiya Umoja Noble. These thinkers (all women of color) are at the forefront of the critical interrogation of technologies, like facial recognition, both in the role these systems play in shaping our social lives, as well as how pre-existing forms of bias, racial animus, and asymmetries of power are built into novel digital tech.

The fundamental problem with facial recognition technologies is they attach numerical values to the human face at all. As Browne [5] and other scholars have observed, facial recognition technologies and other systems for visually classifying human bodies through data are inevitably and always means by which "race," as a constructed category, is defined and made visible. Reducing humans into sets of legible, manipulable signs has been a hallmark of racializing scientific and administrative techniques going back several hundred years. The systems used by facial recognition technologies to code human faces perform an essentializing visual schematization.

These systems thus enact a process Browne terms, "digital epidermalization," or "the imposition of race on the body" through the classification and schematization of human facial features [5]. The imposition of racial categories onto human bodies is of course scientifically unsound. As a recent op-ed, authored by more than 60 academics observed, "a robust body of scholarship recognizes the existence of geographically based genetic variation in our species, but shows that such variation is not consistent with biological definitions of race," and, moreover,

Image by Street VJ / Shutterstock.com

that such variation does not "map precisely onto ever changing socially defined racial groups" [6]. Race, in other words, is a set of categories dreamed up by humans and perpetuated by human activity—including through digital systems of classification.

Marta Maria Maldonado describes this process of racialization as, "the production, reproduction of and contest over racial meanings and the social structures in which such meanings become embedded" [7]. "Racial meanings," she observes, "involve essentializing on the basis of biology or culture." Essentialization, or abstracting away context to focus on particular elements, is central to racial animus: Not all essentialization is racist, but all racism involves some form of essentialization.

Like genetic variation, physiological facial variation is not dispositive of racial categories, either biological or sociological. Yet it is precisely because facial recognition technologies do not "see" in the human sense that they are so dangerous. Facial recognition involves identifying, extracting, and selecting contrasting patterns in an image, and then classifying and comparing them to a previously compiled database of other patterns. Facial recognition technologies assign numerical values to schematic representations of the face, and make comparisons between those values. At a technical level, it is not possible to separate the work of associating schematically mapped parts of the face with real humans with quantitative comparison, ordering, and ranking.

Critical race scholars, from W.E.B. DuBois and Franz Fanon in the early 20th century to Browne, Achille Mbembe, Eduardo Bonilla-Silva, and Kimberlé Williams Crenshaw today, have articulated the connections between systems of racial oppression and quantification [8]. In the case of facial recognition, the schematization of human facial features is driven by a conceptual logic that these theorists and others, such as the French philosopher Michel Foucault, have identified as fundamentally racist because it is concerned with using statistical methods to arbitrarily divide human populations.

This process of biopolitical management is grounded in finding numerical reasons for construing some groups as subordinate, and then reifying that subordination by wielding the "charisma of numbers" to claim subordination is a "natural" fact. As such, racism's function, as Foucault describes it, is "a way of introducing a break into the domain of life [...] of fragmenting the field of the biological that power controls" [9]. Race and racism are "the preconditions that make killing acceptable" in societies focused on making discriminations based on technical norms and standards—the justification in turning authority's custodianship of life and living into that of death and dying [10].

There are strong arguments for an outright ban on facial recognition systems, but there have also been increasing calls for regulation from the tech sector itself.

Recent work by Joy Buolamwini and Timnit Gebru at MIT [11] documents the existing bias in facial recognition training sets; the difficulty many commercial facial recognition systems have in recognizing darker female faces illustrates one aspect of digital epidermalization's privileging of whiteness [12]. As Buolamwini observes, "monitoring phenotypic and demographic accuracy of these systems as well as their use is necessary to protect citizen rights" [12]. Facial recognition technologies were neither invented for nor exist in a social vacuum. They are, in Browne's words, "designed and operated by real people to sort real people." As an example of these systems' discriminatory effects, both Browne and Buolamwini note how, in Browne's words, "particular biometric systems privilege[e] whiteness, or lightness, in the ways in which certain bodies are measured for enrollment." By introducing a variety of classifying logics that either reify existing racial categories or produce new ones, the automated pattern-generating logics of facial recognition systems both reproduce systemic inequality and exacerbate it.

Yet even if facial recognition systems were ever able to map each and every human face with technical perfection, the core conceptual mechanism of facial recognition would remain, in my view, irredeemably discriminatory. If human societies were not racist, facial recognition technologies would incline them toward racism; as human societies are often racist, facial recognition exacerbates that animus. This

Image by Street VJ / Shutterstock.com

claim is a diagnostic one about systemic problems, not a polemic against the designers and makers of these technologies. The analogy to plutonium is apt: As a radioactive element, plutonium's biological toxicity comes from its structure, just as facial recognition's social toxicity comes from the very parameters of what its algorithms do.

Like the refining of plutonium, the basic research programs developing facial recognition technologies was funded by, but formally separated from, the military [13]. And like many scientists involved in the Manhattan Project in the 1940s, computer scientists are also sounding the alarm regarding the technologies they and their colleagues have made. Yet what is the harm-to-benefit ratio around facial recognition technologies? Hartzog and Selinger observe, "when technologies become so dangerous, and the harm-to-benefit ratio becomes so imbalanced, categorical bans are worth considering." I have laid out the harms inherent in these systems, of which racial categorizing is one of many. Where, if any, are the benefits?

Proponents of facial recognition point to several arenas in which they claim these technologies will bring benefits; these include public safety, consumer convenience, and the general verification of individual identity online. Yet given the ways facial recognition systems embed racializing and racist logics into its structure, the potential harm for these systems' use in public safety and law enforcement contexts should be obvious; it is the equivalent of deploying a tactical nuclear weapon to demolish an ordinary office building. Reports prepared by the Electronic Frontier Foundation (EFF)'s Jennifer Lynch [14], and by Clare Garvie, Alvaro Bedoya, and Jonathan Frankle from Georgetown's Center on Privacy and Technology [15] exhaustively document the dangers to civil liberties and potentials for racial discrimination posed by facial recognition, as does AI Now's recent 2018 report [16].

Racial discrimination by facial recognition is not only a problem in the United States, as recent reporting on Chinese detention of members of the minority Muslim Uyghurs in western China make clear. But in the case of the United States, understanding facial recognition in its security context as a powerful means toward systemic oppression against black people highlights just how toxic its use within the broader edifice of securitization is, and how much that edifice's other techniques and procedures are structurally accomplice in racist violences—again, regardless of the individual feelings and beliefs of the professionals involved in its design and deployment.

Likewise, using facial recognition for more general forms of confirming identity online raises similar core questions regarding trading off its enormous risks for relatively meager gains. Recently, Sebastian Benthall, responding to a question I posed on Twitter of "When *shouldn't* you build a machine learning system?" suggested "one should not build an ML [machine learning] system for making a class of decisions if there is already a better system for making that decision that does not use ML" [17]. This response highlights how little is to be gained by the widespread deployment of facial recognition to confirm identity, and how much there is to lose. Why introduce an invasive technology with a wide range of ill effects, when other mechanisms will do as well? In effect, this move is the equivalent of using plutonium to heat not distant space probes, but residential homes: other options do the job just as well, and the risk of horrendous consequences is vastly reduced.

Perhaps the most widely cited use case for facial recognition is as a tool of consumer convenience and playfulness. Yet here, racial logics rear their ugly head too. Masking apps are one particularly egregious, and obvi-

The fundamental problem with facial recognition technologies is they attach numerical values to the human face at all.

Why introduce an invasive technology with a wide range of ill effects, when other mechanisms will do as well?

ous, avenue for racial discrimination via digital epidermalization. In 2016, Snapchat came under fire for a "Bob Marley" mask filter described by many commentators as "digital blackface." In 2017, the FaceApp app deployed a "Hot" filter than lightened a photograph's skin tone and applied smoothing to make a subject's facial features appear "white." Apparently immune to the widespread critiques of racism, the company later released a set of filters explicitly labeled as racial: "Asian, Black, Caucasian and Indian."

Likewise, digital animations like Samsung's memoji and Apple's animoji improve facial recognition technologies' ability to recognize all human faces, making this logic of racializing privilege even more pervasive and perverse. These systems serve as facial privacy loss leaders, getting users accustomed to cute, seemingly harmless applications of facial recognition tech. Precisely because they enlist multiple different technologies of classification within the mechanisms through which our digital social and emotional lives take place, these systems are racialized surveillance disguised as animation. In the consumer context, there is no reason to allow a technology with such toxic effects.

There are strong arguments for an outright ban on facial recognition systems, but there have also been increasing calls for regulation from the tech sector itself. In July of 2018, Brad Smith, President and General Counsel of Microsoft, called for both vigorous regulation of and heightened corporate social responsibility toward facial recognition systems [18]. [For full disclosure, I am an employee of Microsoft Research, though the views expressed in this article are my own and do not represent those of Microsoft Research or Microsoft more broadly.] Smith observed the utility of regulation in areas such as automobiles, air safety, food, and pharmaceutical products—one could easily have added hazardous waste to the list. Smith's position shows how even companies invested in some applications of facial recognition like Microsoft recognize some of the dangers these technologies pose.

Recognizing facial recognition as plutonium-like in its hazardous effects only underscores the need to build on calls for regulation like Smith's, paying close attention to how the government regulates a hazardous substance like plutonium. Smith notes one potential limited use case for facial recognition: As an accessibility tool for the visually impaired. Under a strong regulatory scheme, devices enabling this kind of functionality, like other digital accessibility devices and clinical health apps might be regulated by the Food and Drug Administration. Just as the use of a substance like plutonium for specialized medical or security applications is highly constrained and closely monitored, facial recognition technologies could be subject to similar constraints. Plutonium serves as a useful metaphor for facial recognition because it signals some technologies are so dangerous if broadly accessible that they should be banned for almost all practical purposes.

Facial recognition's racializing effects are so potentially toxic to our lives as social beings that its widespread use doesn't outweigh the risks. "The future of human flourishing depends upon facial recognition technology being banned before the systems become too entrenched in our lives," Hartzog and Selinger write. "Otherwise, people won't know what it's like to be in public without being automatically identified, profiled, and potentially exploited." To avoid the social toxicity and racial discrimination it will bring, facial recognition technologies need to be understood for what they are: nuclear-level threats to be handled with extraordinary care.

References

[1] Clark, D. L. and Hobart, D. E. Reflections on the legacy of a legend: Glenn T. Seaborg 1912–1999. *Los Alamos Science* 26, (2000), 5661.

[2] Stark, L and Hoffmann, A. L. Data is the new what? Popular metaphors & professional ethics in emerging data cultures. In *Journal of Cultural Analytics* Special Issue: Data Cultures, Culture as Data, (2019).

[3] Hartzog, W. Facial recognition is the perfect tool for oppression. *Medium*. (August 2, 2018); https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66

[4] Stark, L. Facial recognition, emotion and race in animated social media. *First Monday*. (September 3, 2018); https://firstmonday.org/ojs/index.php/fm/article/view/9406/7572

[5] Browne, S. *Dark Matters: on the Surveillance of Blackness*. Duke University Press, Durham, 2015.

[6] Maldonado, M. M. It is their nature to do menial labour: The racialization of 'Latino/a workers' by agricultural employers. *Ethnic and Racial Studies* 32, 6 (2009), 1017–36.

[7] How not to talk about races and genetics. *BuzzFeed News*. (March 30, 2018); https://www.buzzfeednews.com/article/bfopinion/race-genetics-david-reich

[8] Crenshaw, K. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43, 6 (1991), 1241-1299.

[9] Foucault, M. 17 March 1976. In *Security, Territory, Population*. Picador, New York, 2009, 239–63.

[10] Mbembe, A. *On the Postcolony*. University of California Press, Berkeley, CA, 2001.

[11] Buolamwini, J., and Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research* (PMLR'18). PMLR, New York, 2018, 1–15.

[12] Buolamwini, J. Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers. MIT. Master's thesis. 2017; https://www.media.mit.edu/publications/full-gender-shades-thesis-17/.

[13] Keyes, O., Stevens, N., and Wernimont, J. The government is using the most vulnerable people to test facial recognition software. Slate.com. (March 17, 2019); https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html

[14] Lynch, J. Face off: Law enforcement use of face recognition technology. *EFF*. (February 12, 2018); https://www.eff.org/wp/law-enforcement-use-face-recognition

[15] Garvie, C., Bedoya, A. and Frankle, J. The perpetual line-Up: Unregulated police face recognition in America. *Perpetual Line-up*. (October 18, 2018); https://www.perpetuallineup.org/

[16] AI Now. AI Now Report 2018; https://ainowinstitute.org/AI_Now_2018_Report.pdf

[17] Benthall, S. When shouldn't you build a machine learning system. *Digifesto*. (December 22, 2018); https://digifesto.com/2018/12/22/when-shouldnt-you-build-a-machine-learning-system/

[18] Smith, B. Facial recognition technology: The need for public regulation and corporate responsibility. *Microsoft On the Issues*. (July 13, 2018); https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/

Biography

Luke Stark is a postdoctoral researcher in the Fairness, Accountability, Transparency and Ethics (FATE) Group at Microsoft Research Montreal, and an Affiliate of the Berkman Klein Center for Internet & Society at Harvard University. His work explores the history, ethics and social impacts of computational media and AI. Stark holds a Ph.D. from the Department of Media, Culture, and Communication at New York University, and an Honours B.A. and M.A. from the University of Toronto.