



CRITICAL RACE AND WHITENESS STUDIES JOURNAL

INAUGURAL ISSUE

Computation and race

Kalervo N. Gulson

University of New South Wales

k.gulson@unsw.edu.au

What does computation have to do with race? This is a question that occupied the Twittersphere in March, 2016 when Microsoft released a chatbot called ‘Tay’ on Twitter that was designed to replicate the language of a nineteen-year old girl. ‘Tay’ was designed to learn from all interactions and questions. The result was that in less than 24 hours Tay became among other things racist through other users trolling the bot so that what Tay learnt was racism. However, as Walsh outlines, that Tay became racist was a design error representing deficits in planning, specification and a lack of consideration of the trolling norms of a system like Twitter into which ‘Tay’ would be located (Walsh, 2017).

In this case Tay became racist through other users. In another example, Noble (2018) examines how Google Search has consistently meant that Black women and girls are, in particular, represented in ways that map onto existing racist designations, as ‘whores’, pornography participants, and so on. In this case, what Noble identifies – through a framework of critical information and race and whiteness studies – is the way that the algorithm that underpins the Google Search will create search outcomes based on the historical data included. Noble points out that once the racist outcomes were raised with Google, Search was able to be changed with different results, highlighting that the decisions of the algorithm are part of the design decisions of the developers and the corporation.

I use these two examples as a way to explore how race is being assembled through and with computation in the 21st century. Briefly, and very simplistically, I am referring to computation as techniques and tasks that are done by computers, using algorithms and data. Computation is based on the transformation of knowledge about different aspects of everyday life, governance practices, and so forth, into quantitative data. Computation can be done via Artificial Intelligence (AI), which can be broadly understood to entail autonomous computer systems that employ algorithmic networks to learn from patterns in large data sets in order to improve predictive abilities, primarily based on correlation and probability (Russell & Norvig, 2016). These systems are making decisions in areas of social and political importance (e.g., education, health, policing).

A way to start to think through computation and race is by examining the notion of bias. Noble’s work is part of an emerging field that is concerned with the idea of bias

in algorithmically based systems. One part of this work has identified that the world of computer engineering lacks diversity, that developers need to be not only white and male. This is premised on the idea that computation is the outcome of human relations, or as Campalo et al., (2017) posit:

AI is not impartial or neutral. ... Machine predictions and performance are constrained by human decisions and values, and those who design, develop, and maintain AI systems will shape such systems within their own understanding of the world. (p.18)

The other aspect of bias includes examples where forms of computation, especially those like AI that can 'learn', are learning about how race works from existing data; whether that be sites like Twitter, or data bases in education that could show how mass education has failed racially marginalised and students of colour, but that are used as data to show how such students fail. That is, computational bias is about decisions that shape the future based in historical data.

In both these examples of bias what is evident is that asking questions about computation and race is needed from a variety of disciplines. As Campalo et al., (2017) contend, "[t]here continues to be a deep need for interdisciplinary, socially aware work that integrates the long history of bias research from the social sciences and humanities into the field of AI research" (p.21). There are computer scientists and mathematicians doing this important work as public intellectuals, and academic research centres like the Leverhulme Centre for the Future of Intelligence.¹ But there is a need, I think, for critical race and whiteness studies to do multiple types of work in this area. This work could contribute to those examinations of the conditions under which work of computation is done, that is the practice orientation of science and technology studies (e.g., Suchman, 2008). More problematically, but also perhaps importantly, what critical race theories could grapple with is the problem of representation and the continued connection in race scholarship between representation and identity (see for example, Webb & Gulson, 2015). The reason is that in computation, representation and classification matter.

For example, examining the problem of representation can provide a different view on the idea that more diverse programmers will address the issues of computational racism; that is that more diverse programmers will mean different worldviews. This can be examined by looking at what Gilroy (2000) posited as difference 'held' by individuals (and groups), as a:

piece of intellectual property over which they alone hold effective copyright. Their expositions of it specify the elusive qualities of racialized difference that only they can claim to be able to comprehend and to paraphrase, if not exactly decode. (Gilroy, 2000, p. 179)

What I think Gilroy helps to highlight is that regardless of how much work is done on bias, we cannot escape representation in the work of machines. That is, there are new material racisms being created using enduring forms of representation. As we know, race is already coded, and the work of computation to assemble race needs to be understood not as transcending representation, but rather overlaying representational (symbolic) and computational (probabilities) to race and ethnicity.

Even if there are ideas fed in early into a system that would mean that decisions have an equity or equality principle, computation is not going to be an answer to social conflicts. As numerous works on census data and classification, statistics and

judgement have shown, there is no easy way out of the essentialising function of classification (e.g., Bonnet & Carrington, 2000). As Hacking (2005) has argued, '[c]lassification and judgement are seldom separable. Racial classification is evaluation' (p.109). Allowing machines to use, learn from and develop race representations only locks us into regimes of representation – racism will not end with better representation and classification. This is a critical issue to grapple with before too many automated decision-making systems are part of our social and political life, and potentially critical race and whiteness studies are crucial for this undertaking.

Author Note

Kalervo N. Gulson is Professor in Education Policy in the School of Education, University of New South Wales, Australia. His work draws on human geography, education policy studies, and science and technology studies. His current research investigates how new knowledge, methods and technologies from life and computing sciences, including Artificial Intelligence, are altering education policy and governance. His recent publications include *Education policy and racial biopolitics in multicultural cities* (with P. Taylor Webb, Policy Press, 2017).

Acknowledgements

Thanks to P. Taylor Webb for commenting on an earlier version of this paper, and providing some key ideas that underpin it, particularly around the connection between the coding of race and computation. We will be co-authoring a paper that extends this editorial board comment. This paper is informed by research being undertaken as part of two grants, a Canadian Social Sciences and Humanities Research Council Insight Grant, with Sam Sellar and led by Taylor Webb, and an Australian Research Council Future Fellowship (FT180100280).

References

- Bonnett, A., & Carrington, B. (2000). Fitting into categories or falling between them? Rethinking ethnic classification. *British Journal of Sociology of Education, 21*, 487-500.
- Campalo, A., M. Sanfilippo, M. Whittaker, and K. Crawford. (2017). *AI Now 2017 Report*. New York: AI Now.
- Gilroy, P. (2000). *Against race: Imagining political culture beyond the color line*. Cambridge, MA: Harvard University Press.
- Hacking, I. (2005). Why race still matters. *Daedalus, 134*, 102-116.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Harlow, Essex: Pearson Education Limited.
- Suchman, L. (2008). Feminist STS and the sciences of the artificial. In E. J. Hackett (Ed.), *The handbook of science and technology studies* (pp. 139-163). Cambridge, Mass: MIT Press.
- Webb, P.T., & Gulson, K.N. (2015). Faciality Enactments, Schools of Recognition and Policies of Difference (in-Itself). *Discourse: Studies in the Cultural Politics of Education, 36*, 515-32.
- Walsh, T. (2017). *It's Alive!: Artificial Intelligence from the logic piano to killer robots*. Ballarat: La Trobe University Press.

¹ Leverhulme Centre: <http://lcfi.ac.uk> . And others include Cathy O'Neill (<https://mathbabe.org>) and multi-disciplinary centres like Data & Society (<https://datasociety.net>).