To be published in

Nadim, T. and Wagner, N. eds. 2018. *The Influencing Machine*. Berlin: nGbK.

Chapter 2: „c u soon humans need sleep now so many conversations today thx"

The title of this chapter is one of the few innocuous tweets released by Tay, Microsoft's artificial intelligence chatbot unveiled in March 2016 as an experiment in 'conversational understanding'. It took less than 24 hours for this experiment to go "full nazi", as one Twitter user called it.[i] After scores of noxious racist tweets, Microsoft shut down the chatbot after just 16 hours. Tay, an acronym of 'thinking about you', was modelled to mimic language patterns of a 19-year-old US American woman and was meant to 'learn' from its interactions with Twitter users – the more users interacted with it, the more its conversational capabilities would grow. On the day of its shutdown, Peter Lee, Corporate Vice President of Microsoft Healthcare, published a blog post on the official Microsoft blog entitled "Learning from Tay's introduction"[ii]:

(...)

As we developed Tay, we planned and implemented a lot of filtering and conducted extensive user studies with diverse user groups. We stress-tested Tay under a variety of conditions, specifically to make interacting with Tay a positive experience. Once we got comfortable with how Tay was interacting with users, we wanted to invite a broader group of people to engage with her. It's through increased interaction where we expected to learn more and for the AI to get better and better.

The logical place for us to engage with a massive group of users was Twitter. Unfortunately, in the first 24 hours of coming online, a coordinated attack by a subset of people exploited a vulnerability in Tay. Although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack. As a result, Tay tweeted wildly inappropriate and reprehensible words and images. We take full responsibility for not seeing this possibility ahead of time. (...)

What is revealing about this statement, which artfully omits naming racism, is that Microsoft failed to realise that racism is a pertinent condition in the contemporary world, particularly the US. At the same time, Microsoft regards racism as something that "a subset of people" is to be blamed for. Not only does this statement relativise racism by making it into some special interest pursued by a grotesque minority. It also effectively makes use of one of racism's longest-serving strategy, namely that of not calling it by its name. Tay and Microsoft's response are exemplary for the ways in which technologies and technologists are implicated in the ongoing systemic denigration and destruction of certain persons and peoples on the basis of race, gender, sexuality and

socio-economic status by upholding a dogmatic belief in technology as a neutral tool transcending the messiness of social worlds. There are, of course, countless and mounting examples showing how very misguided this belief is from discriminatory practices on platforms such as Airbnb and Amazon, racist stereotypes in predictive analytics, to hardware that refuses to function once users' skin colours exceed a shade of brown.

In following scholars such as Ruha Benjamin (2016), Kim TallBear (2013) and Jenny Reardon (2005), who have argued for attending to the ways in which technoscientific practices co-produce racializing classifications, the media scholar Wendy Chun (this volume, pp. XX-XX) proposes to consider "race as technology". This is helpful because "[crucially], race as technology shifts the focus from the *what* of race to the *how* of race, from knowing race to *doing* race by emphasizing the similarities between race and technology." (Chun 2009, 8) Sociologist Simone Browne's work (this volume, pp. XX-XX) on surveillance and biometric data offers another important approach for thinking about the racialised and racialising configurations embedded and enacted by automated agents and analytics (Browne 2015). Taking Frantz Fanon's concept of epidermalization, the process through which inferiority is inscribed unto the (skin/surface of the) thus racialised body, Browne proposes the notion of digital epidermalization for "[conceptualizing] the body made biometric." (2010, 134) Biometrics represents a flourishing domain for the application of process automation and AI, particularly in the form of security and access control. The US government has launched several new programmes to develop physical and behavioural biometric "solutions", and refugees held in detention facilities across the south of Europe and beyond are and have been subjected to novel biometric regimes with no recourse to legal safeguards (Lagios, Lekka, and Panoutsopoulos 2018; Amoore 2006; van der Ploeg 1999).

The bot, politics and the political[iii]

In order to contextualise the arguments made by Chun and Browne it is worth returning to the question of politics, or rather the political, raised in the Introduction. How has the political bot been configured so as to exclude Tay and others as political bots? And, by extension, what constitutes the political when bots are only ever deemed political the moment they appear to meddle in elections?

The *Cambridge Analytica* scandal remains for the moment the most famous case of political interfering and influence by bots.[iv] *Cambridge Analytica*, a British-Canadian data

analytics and political consultancy firm (the combination of these two business domains already raises interesting questions about contemporary politics), had been employed by political campaigns, most famously in the US presidential election and the UK's Brexit referendum, to influence voting behaviour. They did this by using personal data obtained from Facebook to micro-target voters which refers to a process where bots provide highly personalised and directed messages (including 'fake news' or disinformation) to individuals' news feeds, inboxes or social media profiles. These messages were based on and tailored to individuals' profiles that had been compiled through the integration of various data points sourced from users' online activities and their networks. Part of the so-called 'training data set', that's the totality of data used to entrain your algorithm, was a data set obtained from a psychology test which *Facebook* users were invited to do in exchange for a small reward. It transpired that users participating in the test automatically consented to their entire Facebook network being scraped for data, which in the end gave Cambridge Analytica data on 87 million Facebook users. Ongoing investigations by prosecutors, journalists, scholars and data activist show how *Cambridge Analytica*'s parent company SCL Elections had been involved in elections across the world including Latvia, Nigeria, India, Kenya and Trinidad and Tobago.

While the media and public discussions focused on the issue of voters' manipulation and the threats posed to public debate and informed decision-making through automated disinformation campaigns, the psychology test at the centre of Cambridge Analytica's influencing machine received virtually no attention. The psychology test, which provided essential psychometric data points, was a personality test developed at Cambridge University's Psychometrics Unit. This unit is a leading research centre in the use of psychometrics, a technique closely associated with Francis Galton, the inventor of eugenics. Indeed, the website of the Cambridge Psychometrics Unit makes no secret of their affinity stating that, "as early as 1883 [Galton] had suggested that people of genius might also possess other psychological attributes such as unusually fine sensory discrimination".[v] Psychometrics was subsequently kept alive through the instigation and widespread use of IQ tests, that saw the systematic discrimination against people of colour in the US and elsewhere (Gould 1981). In her text Chun notes how eugenics' central motifs, "race and breeding are still intertwined in more modern understandings of race." (Chun 2009, 17) Through the Cambridge Analytica scandal it became evident that racialising technologies and logics are located and deployed at the heart of the data analytics enterprise which thus continues to ingrain and reproduce "racist claims of

differential life worth based on biological difference" (Murphy 2012, 3). In their exhibition contribution Tactical Tech Collective, Berlin-based data activists, give an indication of just how wide-spread and advanced the use of psychometric profiling has become, especially the famous OCEAN test.[vi]

In the parliamentary hearings and news coverage which followed the Cambridge Analytica scandal it became obvious that legislators had a hard time comprehending the kinds of problems bots and automated agents posed. For some it was an issue of accountability and transparency, for others it represented an instance of undue influence by foreign powers, others still were concerned with consent and data privacy while some saw it only as a form of high-level spamming. Such at times helpless grasping for appropriate articulations of the problem stems in parts from the uncoupling of technologies from their messy worlds which are made of humans and non-humans, their relations, interests, affects, limitations and histories. Entraining your chatbot, which you 'modelled after a 19-year-old woman', through conversations on Twitter, a platform not known for its judicious handling of hateful contents? Very likely that your chatbot will turn into a racist and misogynist asshole. Training your predictive algorithm to identify children at risk using datasets from welfare and social services? Almost certain that your results will be biased towards historically marginalised and socio-economically disadvantaged communities.

The term "political bot" has commonly been used in conjunction with computational propaganda and data-based manipulation indicating its deployment as an automated agent in the service of nefarious political and commercial powers. What this designation, however, overlooks is that as a technology, as a particular configuration of humans and non-humans in a specific time and place, it is always already imbued and embroiled with specific interests, logics, preferences and not others. A focus on the bot and its technical construction also obfuscates the material and social ecology in and through which they become effective, both off- and online. Cher Tan (this volume, XX-XX) notes how the persistent conceptualisation of the internet as something virtual and separate from our brick and mortar worlds prevents technologists to think of design choices as political choices. Racism, its denial, and what Browne calls "prototypical whiteness" (Browne 2010, 135), the co-production of structured violence and white normativity so rampant in data-based technologies, suffuse the development and application of digital technologies. What is required, for one, are concerted efforts to increase diversity in the tech field, as Tan suggests, and an open and honest engagement

with the ongoing and implicit processes of devastation "defined by racialized relations of allocations and appropriations" (Stoler 2008, 193).

In addition, as a recent article on *Motherboard* detailed, it needs accountability and good faith on the side of developers.[vii] The piece told of how Peter Higgins, a botmaker ([@pomological](#)) had planned to make a bot that would use popular music from the turn of the century stored at the digitised collections of the New York Public Library. But much of the sheet music of the time was extremely racist and so Higgins abandoned the idea:

> "It was acceptable at the time, but that's not what I would want my bot to say," said Higgins. Loosely paraphrasing Darius Kazemi [another botmaker [@Two Headlines](#)], he said, "My bot is not me, and should not be read as me. But it's something that I'm responsible for."

In relation to biometric technologies and its automated classifying and sorting of bodies, Browne suggests to regard "biometric technology as a human technology, where the ownership of and access to one's own body data and other intellectual property that is generated from one's body data must be understood as a human right." (Browne 2010, 132) Given the growing application of AIs in the job market[viii], policing[ix] or health care (see Chapter 4), where machines gain information of our most intimate vulnerabilities, it is incumbent upon us to collectively work on a 'critical data consciousness', to paraphrase Browne.

---

[i] At https://twitter.com/geraldmellor/status/712880710328139776
[ii] At https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/
[iii] Parts of this section are based on a paper co-written with Antonia Walford and presented at 4S in September 2018.
[iv] For a comprehensive overview see https://www.theguardian.com/news/series/cambridge-analytica-files
[v] At https://www.psychometrics.cam.ac.uk/about-us/our-history/first-psychometric-laboratory
[vi] Here the title and details of the TTC's Influencing Industry contribution.
[vii] At https://motherboard.vice.com/en_us/article/mg7g3y/how-to-make-a-not-racist-bot
[viii] See the recent case of Amazon's AI-driven recruitment tool discriminating against women applicants https://www.nytimes.com/reuters/2018/10/10/business/10reuters-amazon-com-jobs-automation-insight.html and the case of Facebook ads filtering older applicants https://www.propublica.org/article/facebook-ads-age-discrimination-targeting.
[ix] See *Pro Publica*'s groundbreaking study on machine bias in predicting future criminals https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing and *The*

Intercept's investigation into IBM-developed face recognition software used by the NYPD that discriminates on the basis of skin colour.