To be published in

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

Chapter 4: "IBM Watson is the Donald Trump of the AI industry"[i]

In 2009 a 'cognitive computer' named Watson, an IBM product, shot to fame by winning *Jeopardy!*, a televised US American quiz show. In an article that followed this stunt IBM Watson's research team described their work as follows

> We believe advances in question-answering (QA) technology can help support professionals in critical and timely decision making in areas like compliance, health care, business integrity, business intelligence, knowledge discovery, enterprise knowledge management, security, and customer support (Ferrucci et al. 2010).

As a so-called 'deep learning expert system', Watson has since integrated many artificial intelligence (AI) techniques, including voice recognition and sentiment analysis, also known as opinion mining, which infers, extracts and quantifies affective states from biometrics, text analysis and natural language processing. The breadth of application mentioned by Ferrucci et al. is staggering and, indeed, Watson found application in, among others, healthcare, call centres, the film industry, cooking, law enforcement and fashion. It thus serves as a prominent example for investigating some of the ways in which AI is being integrated into existing domains of work and life more generally. At the same time, it provides a cautionary tale about the inflation of expectations that has accompanied, that might well be constitutive of, AI developments. The quote which provides the title for this chapter is indicative of some of the responses – from experts and users – to IBM's spectacular marketing campaign including celebrities and preposterous claims about being able to predict global weather, improve education and provide treatment suggestions for 80 per cent of global incidence of cancers. It did not take long for the hype to deflate, internal documents and interviews with doctors revealing that decisions were often inaccurate and that there were "serious questions about the process for building content and the underlying technology." (Ross and Swetliz 2018) Spectacular claims are not rare in AI development and they make it easy to see its failures and shortcomings. However, what these rhetorics omit are the very "serious questions" about the underlying logics and infrastructural conditions that enable something to claim artificial intelligence. This is particularly alarming when AI

rhetorics and AI-based systems combine to shore up dominant social and political biases such as in predictive policing.

We have seen in Chapter 2 how design decisions are always political in that they enact certain inclusions and exclusions. In this chapter we will explore the role of classification in the decision-making of expert systems and, by extension, raise critical questions about the current push for AI and machine learning, which is seen not only as the future for the tech industry but for knowledge production more generally.[ii] All the tech giants, Facebook, Apple, Microsoft, Google and Amazon (FAMGA) are investing heavily in various AI ventures, from Facebook's DensePose, which seeks to automatically recognise human body poses by mapping 2D human image data unto 3D body models, to Google's Duplex, which can make a call and book a hair salon appointment for you, and Amazon's AI-powered drone delivery service (currently in the prototype phase). Car companies are zealously pursuing the vision of the self-driving car, and goods manufacturers are equally ardently sticking chips into every conceivable object, no matter how banal. The Internet of Things now includes smart yoga pants and, as Kashmir Hill so impressively recounts (this volume, pp. xx–xx), smart tooth brushes. AI systems are already integrated in vital domains of our public and private life from travel and dating to employment, criminal justice, health and education (Wagner 2018; Hofstetter 2016; Schlieter 2015; Barocas et al. 2014).[iii]

What does this practically mean in relation to, for example, healthcare? IBM Watson Health has developed applications to interpret genetic testing results, automate care management workflow, integrate and analyse different health-related data, optimise business performance and, perhaps most astoundingly, provide doctors with second opinions on treatment options. This last product, designed specifically for application in cancer care, is deployed by 230 hospitals around the world (in the US, China, India, South Korea and Slovakia among others).[iv] At the core of AI is the idea that machines can make relevant, some might say *good*, decisions autonomously. One origin of AI is the work of Alan Turing, especially his famous question, published in 1950 in *Computing Machinery and Intelligence*, "Can machines think?". In their film *The Outlawed* (2018), showing as part of the exhibition, filmmakers Fabien Giraud and Raphaël Siboni focus on the figure of Alan Turing (played by actress Aurore Broutin) as he studies natural specimens on his travels to Greece while being forced to undergo invasive hormonal treatment for being gay, deemed deviant and criminalised at the time. Homosexuality had been classified as a mental illness by the World Health Organisation's ICD, the International Classification of Diseases that functions as the international standard

diagnostic tool for medical researchers and practitioners as well as related domains such as public health management and legal systems. Homosexuality was only removed from the ICD in 1990 but its classification had had devastating and long-lasting effects on people and their communities. "Classifications matter", as the sociologists Geoff Bowker and Leigh Star note in their brilliant book *Sorting Things Out* (1999), which also includes a lengthy discussion of the ICD. And they particularly matter for AI and machine learning.

How machines (don't) think

Machines do not think, they process data. We may not be able to accurately predict what an AI might turn out given the enormous amounts of data and advanced neural network algorithms that comprise it. But this does in no way equate with human thought processes. Machine learning refers to a technique by which computers are designed so that they can act without being explicitly programmed. It appears really very simple as Cassie Kozyrkov, Chief Decision Intelligence Engineer at Google, explains in a blog post for *hackernoon*:[v] Machine learning identifies patterns in data, uses these patterns to label things and runs an algorithm to cluster/fence these labels. Once this is accomplished it works as a model that can then be unleashed unto other data out there in order to return labels. And neural network algorithms are just "flexible, squiggly shapes" (instead of straight lines) for drawing boundaries around your labels. Kozyrkov's breezy tone is intended to disenchant the often impenetrable language used by AI developers and businesses and thus correct the tall rhetorics deployed by the AI business that, as scholar M.C. Elish and danah boyd have argued, "[manufactures] legitimacy" (2018, 2). This, they argue, "provides cover for nascent technologies to potentially create fundamentally unsound truth claims about the world, which has troubling implications for established forms of accountability." (ibid.) In fact, some members of the AI community themselves are concerned about the indiscriminate use of algorithms without a proper understanding of how they actually work ('black boxing') and about the dogmatic pursuit of optimisation, that is, ensuring your algorithm fences the right labels in the right boundaries (Hutson 2018; Schleim 2018).

One of the documented failures of IBM Watson's healthcare oncology treatment was due to the way in which the system was trained: it had been fed data from 'synthetic' health cases, meaning that hypothetical patients, chosen by doctors at a specific hospital, provided the baseline information from which the system developed its treatment recommendations. So the data with which Watson was entrained were partial and could not necessarily be made to bear on patients elsewhere. This is not to suggest that IBM's

mistake was to use too little or the wrong kind of data. Data are made, not found and as such they will always already be partial. As mentioned above, Watson does not think but instead processes information. In order to make sense of the vast amounts of data entraining algorithms, certain processes of classification need to happen. Certain data become classified as, for example, 'MRI scan data' or certain images as 'face' or as 'male' and such classifications are done relationally, that is, characteristics are defined in relation to other characteristics. One obvious difficulty with settling on a classificatory system to build and train your algorithm on is that the categories by which we classify are situated and historically contingent. It is not just that they are mutable but that the kinds of categories one chooses will be a reflection of one's history, position, environment and preferences. A good example to illustrate bias comes from a friend who recently visited the Association for the Advancement of AI (AAAI) conference, an important Palo Alto-based organisation promoting and supporting AI research and community. In one of the conference panels an AI engineer introduced an algorithm that could produce social networks identifying 'gang membership'. He was asked about why he chose 'gangs' and how he defines 'gangs', a complex and place-dependent social phenomena that comprises real estate speculation, police brutality, socio-economic inequality, urban planning and public health amongst other things (see, for example, Hagedorn 2006). The engineer said that he didn't know anything about gangs but just wanted to build an automated system for articulating networks. The administration in Chicago maintains an unconstitutional gang database that includes personal details on 33,000 young people, some of whom are underage, and 133,000 adults, mostly African-Americans and people of colour, without their knowledge or any mechanism to appeal (see @EraseetheDatabase). According to Amnesty the Metropolitan Police in London manages a similar database where gang membership is inferred on the basis of familial ties.[vi] Amnesty rightly describes this technology as racist. A third of all young people have never been charged with a crime. So when clueless engineers use the heavily biased classification of „gang membership"in order to train an algorithm for creating social network analyses, then they actively support racialising dynamics which further divide our societies.

Bowker and Star note that "a classification is not of itself an explanation." (2000, 319; this volume pp. XX-XX) What it does, however, is "tie the person into an infrastructure – into a set of work practices, beliefs, narratives, and organizational routines". And so "classification (…) has real effect." (ibid.) Once classified as 'non-white' (a problematic category already in its semantics that suggest 'white' as the norm), 'immigrant', 'suicidal'

or a 'gang member' a barrage of structural forces will bear on you in various moments of encountering infrastructures: border controls, loan agreements, job applications, hospital admissions, tenancy applications, university admissions, public transport systems. While the engineer might not have known anything about gangs, his algorithm – once built into AI systems – would indiscriminately classify some people as 'gang members' with dire consequences for their lives and the lives of their friends and families and the status of their neighbourhoods. They ways in which predictive algorithms are wrecking havoc in already disadvantaged communities and on vulnerable subjects (including teachers) are chillingly detailed in Virginia Eubank's *Automating Inequality* (2018), Safiya Umoja Noble's *Algorithm of Oppression* (2018) and Cathy O'Neil's *Weapons of Math Destruction* (2016).

But this is not to say that AI and its applications are inherently bad. Technological determinism has never been a helpful attitude. Instead, this is the time to get smart about smart homes and their promises, to build bots and robots for the flourishing of civic engagements, to take care of one's data and to reclaim high-tech for social justice.

---

[i] Oren Etzioni, CEO of the Allen Institute for AI, quoted in https://gizmodo.com/why-everyone-is-hating-on-watson-including-the-people-w-1797510888

[ii] The US government for example is seeking to replace research with AI. See https://www.heise.de/tp/features/Pentagon-will-wissenschaftliche-Forschung-durch-KI-Systeme-ersetzen-4145928.html

[iii] See https://www.pymnts.com/news/artificial-intelligence/2018/cognitive-spending-machine-learning-innovation-automation/amp/

[iv] https://gizmodo.com/ibm-watson-reportedly-recommended-cancer-treatments-tha-1827868882

[v] At https://hackernoon.com/machine-learning-is-the-emperor-wearing-clothes-59933d12a3cc

[vi] https://www.amnesty.org.uk/press-releases/met-police-using-racially-discriminatory-gangs-matrix-database