Mary Mulvihill Award

Science: Whose Truth? Whose Facts?

Science: Whose Algorithms? Whose Data?

Abeba Birhane and Siobhán Grayson

Doctoral Students in Cognitive Science and Computer Science

School of Computer Science - Insight Centre for Data Analytics

University College Dublin

March 2018

[1] The Wheel of Science

The capacity for self-correction is the source of science's immense strength, but the public is unnerved by the fact that scientific wisdom isn't immutable.

David Barash - Evolutionary Biologist - Paradigms Lost

Truth is often seen as something non-negotiable, something that exists independent and regardless of an individual's views and interests. Many view science as the conveyor of facts and the key to establishing truths. The scientist sets the hypothesis, chooses the methods, carries out the experiment and analysis and voilà, arrives at an objective fact. All of which wouldn't be possible without one of the most important scientific resources that we have, data. Today, thanks to the phenomenal leaps that have been made in advancing technology, mining "raw" data is more accessible than ever. This in turn is assumed to allow scientists everywhere to derive facts and get at the truth not just about the natural world but also about humanity from societal and individual scales. Within data science, the common view is that "raw" data simply exist out there, to be collected, analysed and consequently, to provide us with the truth. In fact, the blessings of computational advances extend beyond experimentation such that we're now able to execute actionable insights in real time, granting the power of the "purest field" [1], mathematics, to play a more pivotal role in our lives via algorithms.

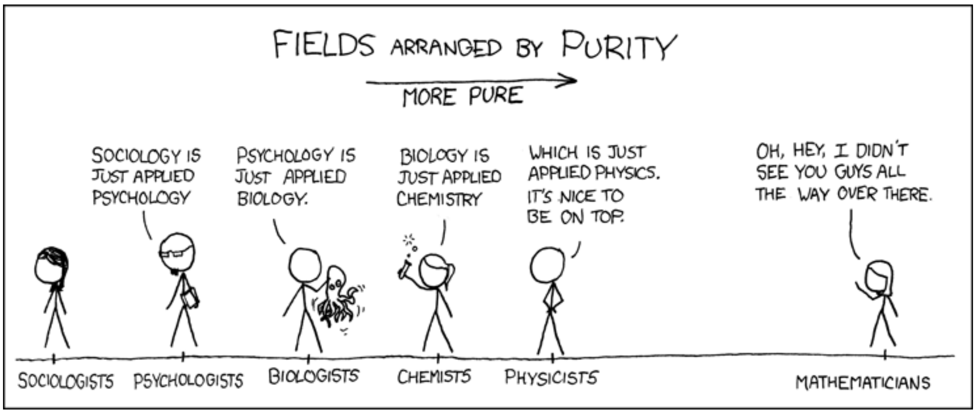

Figure 1.1: *Purity* by xkcd - A webcomic of romance, sarcasm, math, and language [1].

Such simplistic and naive misconceptions of science and truth are quite common [2, 3, 4]. However, it is important to note that science is a human endeavour that operates under certain, often unexamined, theoretical commitments [5]. Truths and facts that we derive from our scientific practices are situated within certain historical, cultural and societal contexts. The type of questions we ask (and those we ignore), the methods and tools we use, the datasets we choose to generate, even the kind of language we use to describe our findings, all impact what types of truths we bring to light, and what truths we may unwittingly bury [6]. Similarly, upon closer inspection, dichotomies that are emblematic to science, such as objective and subjective, are problematic and difficult, if not impossible, to clearly draw, which makes the very idea of "neutral" science unrealistic. What we consider truth and whose truth we unquestionably adopt and amplify as "the truth", then are bounded by these as well as other more complex factors.

Science needs to be viewed as an open and dynamic endeavour that continually goes through a process of reiteration and not a static, fixed and finalized body of knowledge [6, 7]. Viewing science as an open-ended process in turn implies that our truth and facts are not set in stone but also in need of continual revision and reiteration. This might be disconcerting for some. But aren't we better placed acknowledging the complex nature of reality and truths instead of clinging on to mere illusions? After all, if science provided fixed and final answers, then we would not need to revise and improve it. We can just stop enquiring about certain concerns when we arrive at some certain answers and the wheel of science may cease to spin.

Before we go any further, let's look at the possible assumptions that we might be adopting with the very idea of describing the current era as post-truth politics, for example. "Post-truth" in a sense, implies that it has all been truth until recently and undermines the fact that truth and falsehood have always existed alongside each other in dynamical tension as two sides of a coin. People have always been dishonest and manipulative especially when they are politically motivated. Newspapers, radio and television have been used (and continue to be used) as mediums to spread misinformation [8, 9]. The difference, of course, between newspapers and television spreading misinformation versus misinformation spread now via algorithms [10, 11] is that, while we have accountable and responsible agents in the former, accountability and responsibility for misinformation evaporate when it comes to algorithms [12].

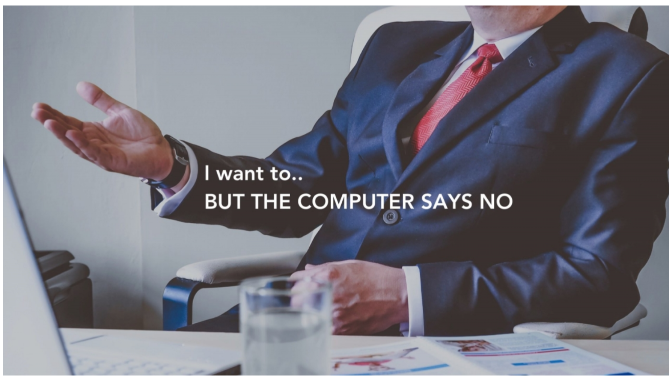

Figure 1.2: I want to.. But the computer says no. Algorithms need to be fair, transparent, and held accountable [13].

Perhaps the instant spread of information on social media platforms and advanced technologies such as machine learning might have given us the illusion that we live in a post-truth era and there might be some truth to that. There is a case to be made for the argument that the biggest casualty to machine learning algorithms will not be jobs but the final and complete eradication of trust in anything we see (Fig. 1.3 [14]) [15, 16] or hear [17, 18, 19] as such algorithms become ever so advanced, making it impossible to differentiate between reality and output generated by an algorithm [20, 21].

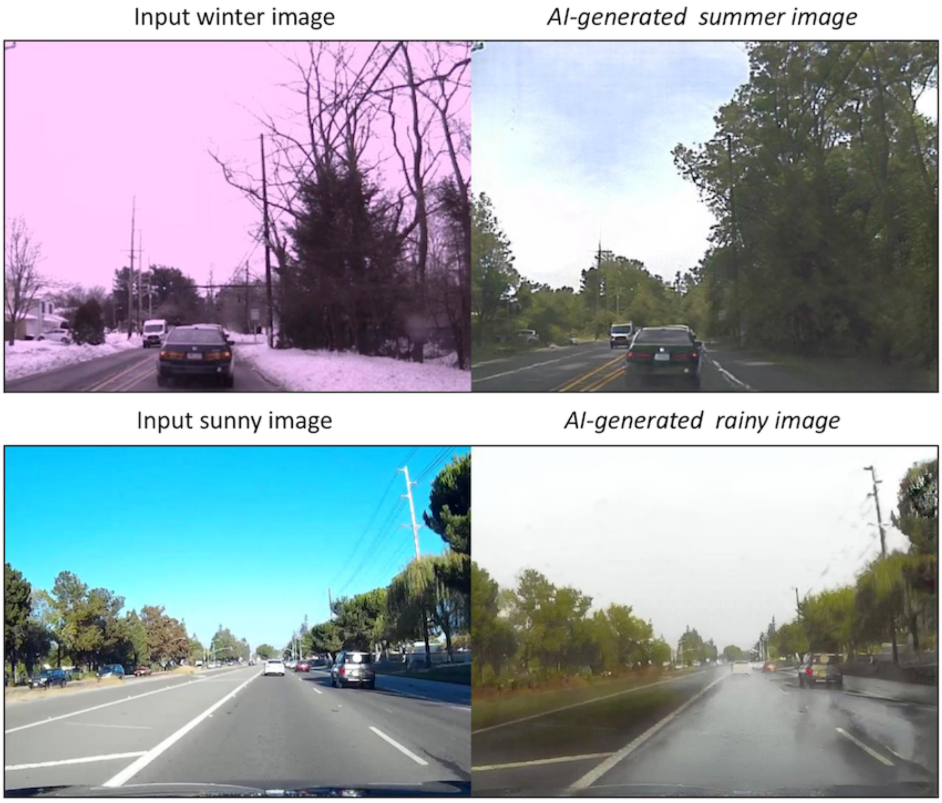

Figure 1.3: An example of Liu et al. [14] work on using unsupervised learning and generative modeling to give machines more "imaginative" capabilities. The winter and sunny scenes on the left are the inputs to the algorithm and the imagined corresponding summer and rainy scenes are the outputs on the right.

Nonetheless, technology advancements serve as a double-edged sword. On the one hand, it has become easy to spread "falsehoods" (for the lack of a better word) instantly and at a massive scale [11]. On the other hand, however, social media platforms have decentralized the source of information and where power resides when it comes to who tells truths and what truths are told. Individuals no longer need to rely on institutions and journalists in order to get information. Individual people share their stories and tell their own truths through these platforms. Injustices have become relatively easy to expose to the world. These platforms are also the driving force behind social and political activism [22]. The #BlackLivesMatter [23] and the recent #MeToo movement are exemplary in this regard.

[2] Science in the Algorithmic Age

Algorithms are opinions embedded in formal code.

Cathy O'Neil - Data Scientist - Weapons of Math Destruction

Up until recently, big data was practically unheard of outside the field of physics and algorithms were seen as just sets of rules to govern machines. Data was a precious commodity, often requiring a substantial amount of time, effort, and expertise to acquire. Today we're inundated with a wealth of data due to the increased use of digital technologies within our lives. So much so, that a new field of science is beginning to bloom known as Data Science whose main tools of enquiry are algorithms written to collect data and perform statistics at scale. Often the data to be studied is our online selves, meaning that many experiments are now being conducted digitally where the identity, methodologies, and motivations of the scientist may be at times opaque [24, 25]. Furthermore, despite the expanding role of algorithms from mere machine instructions to data curators and analysts, our perception of them as neutral has changed little since their invention.

From scientific journals to major newspaper headlines, excessive belief in the power, objectivity, and neutrality of algorithms and big data is widespread [26]. The situation is further convoluted by the fact that many of us are now the subjects of these experiments where consent is rarely explicitly sought if even offered. All our online interactions and behaviours, the virtual side of our lives, are generally viewed as appropriate proxies to study in lieu of the time, effort, finances, and ethic approvals that would be required for our offline worlds. Labelled digital data is regarded as a ground truth and how or why such data arrived at its existence is seldom investigated or considered during analysis [27]. This lack of digital literacy becomes a problem when it leads into thinking that correlation always implies causation and that enormous data sets and predictive models reflect objective truth [28, 29].

This perception of data and of algorithms as neutral and reflective of objective truth allows biases to go unchecked [26]. The idea that math is pure and far removed from the ambiguities of natural language plays a significant role in supporting this fallacy. When in reality, "Algorithms are opinions embedded in formal code" as data scientist, Cathy ONeil, explains in *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* [29]. From tagging black people as gorillas (see Fig. 2.1)

[30], to bringing down minority's credit scores [31], to spreading misinformation and suppressing some voices, algorithms wield societal prejudices (see Fig. 2.2) [32], historical discriminations and unfairness [28]. Since most such algorithms, used in the social sphere, use historical data as their input, it is not surprising that their output is biased. Decisions delivered from these algorithms may not have mattered so much if they simply were recommending what books to buy next based on our previous purchase. However, the stakes are higher if they are diagnosing illness, or holding sway over a person's job [33] or prison sentence [34, 35].

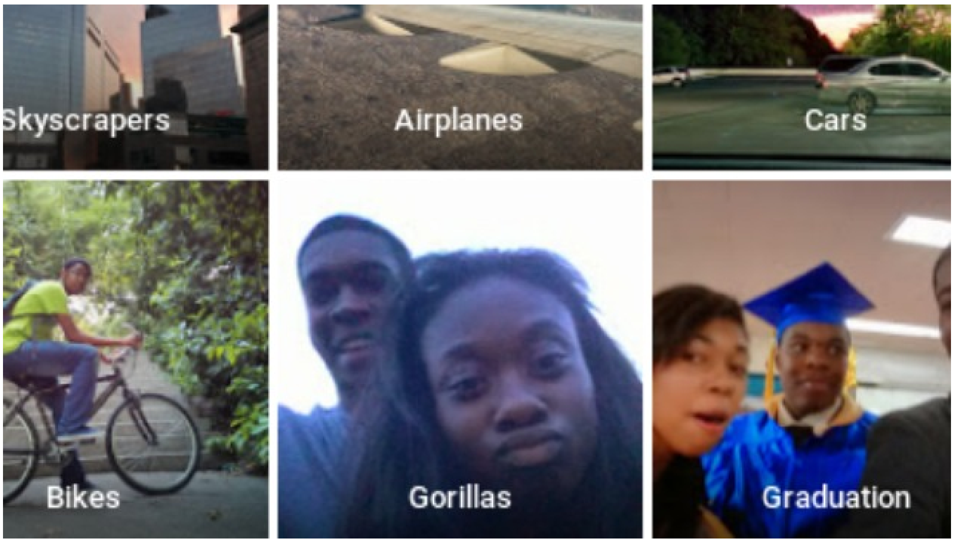

Figure 2.1: Google Photos tags two African-Americans as gorillas through facial recognition software [30].

In all these, those whose truths are being denied and whose credit worthiness scored by algorithms are often the most vulnerable members of society - the poor and racial minorities [36, 37]. These individuals suffer the consequences of decisions delivered from algorithms but often lack the financial, legal or epistemic support to dispute the decision and are often unable to tell their truths. Most of the time, most casualties of these algorithms are not even aware of the fact that algorithms score and decide their truths in the first place. Even when they do, they are simply disenfranchised individuals that have no legal or financial bodies to organize, represent and support them. As we side with the view of data and algorithms as neutral, scientific, and reflective of objective truth, we leave behind the truths of the most vulnerable individuals.

As Gitelman in *"Raw Data" is an Oxymoron* put it succinctly; "Objectivity is situated and historically specific; it comes from somewhere and is the result of ongoing changes to the conditions of inquiry, conditions that at once material, social, and ethical" [38] and we will all benefit from such recognition of nuances.

Allowing algorithms to pervade our social sphere unchecked, disguised under the cover of neutral mathematical formula, is only going to exacerbate the current lack of trans-

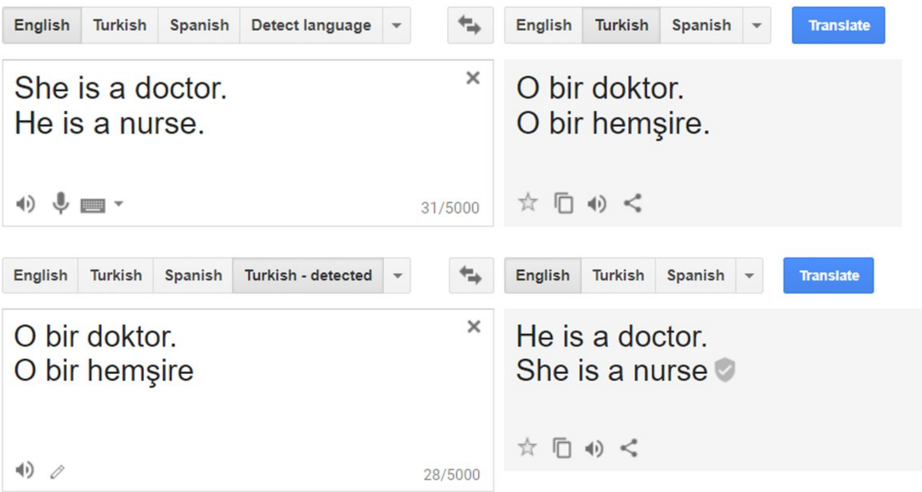

Figure 2.2: Google Translate converts Turkish sentences with gender-neutral pronouns: "O bir doktor. O bir hemsire." to the English sentences: "He is a doctor. She is a nurse." [32].

parency as to whose truth is normalised and accepted and what counts as a fact. The reality is, most algorithms, used within the social sphere, especially in the United States, remain black-boxes that their engineers themselves are unable to explain the decision-making process [39]. Furthermore, these algorithms, privately developed often with the aim of maximizing profits (and not protecting the most vulnerable members of society), remain closed to the public due to corporate intellectual property rights. These corporate companies even go further to hide not only their algorithms but also their existence altogether. This can further obfuscate the experimental pipeline and the potential for replication of results [40]. As deciphering which data sources are not the product of previous experimentation and manipulation becomes a next to impossible task [41].

Data science, like other sciences is a human endeavour and prone to human biases. Truth in data science is far from fixed, objective and independent of the values and interests of the data scientist herself. What is considered as truth, and by whom, is never a straightforward matter given the many different stakeholders. Acknowledging these complexities brings us one step closer to accepting the fact that, like other disciplines that deal with human and societal affairs, the idea of neutral and objective truth and facts in data science are illusions. Additionally, as the works of data scientists increasingly intersect with the social, political, cultural, medical and legal sphere, the need for interdisciplinary effort and ethical considerations becomes obvious and clear. Computer scientists need to be taught not just to code, but the possible consequences and impact that writing algorithms may have on people's lives. What might seem as simply digital and abstract often has impact on "real" offline lives. We would do well to recognize that there are actual people behind what we simply see as data points. The recent global phenomena involving Cambridge Analytica and Facebook data manipulation/breach to advance right-wing political agenda is a notable example [42].

[3] Data Science - Moving Forward

Our understanding can always be improved upon. Even if it is wrong, it doesn't make a preceding insight bad, it is often the necessary intermediary step to get our insight to where it is today.

Julia Shaw - Psychological Scientist - I'm a Scientist, and I Don't Believe in Facts

Since a lot of our problems seem to stem from our misconception that science is a single (as opposed to plural), neutral and value-free endeavour, we are wise to pause and scrutinize what our assumptions rest on.

Within computer science, the assumption that data tell truth and that algorithms are neutral has led to the question of ethics to be cast aside as irrelevant - or ethics belonging to a completely different category to that of computer science. However, as we find algorithms increasingly pervade the social sphere, it is becoming inevitable that ethics needs to be an integral part of data science. Integrating ethics as part of data in a sense creates responsible and aware data scientists and individuals in general.

This does not of course mean that the problem lies within individual data scientists and that training them in ethics will solve all our problems. The issue is complex and requires solutions at individual, societal, organizational, institutional and structural level. Nonetheless, ethical training in computer science is part of the solution in cultivating a culture that is somewhat aware of the intricacies of science, truth and facts. In fact, this is what we, as computer scientists involved in data ethics at University College Dublin in the school of computer science, do.

Ethical training in data science makes visible nuanced questions regarding truths and fact. What is ethically acceptable? Acceptable for whom? Whose interests, integrity, welfare and humanity are being prioritized? And whose are ignored in return? Whose voices are being amplified and in effect whose are being suppressed? Who are the stakeholders in shining light on certain truths and not others? Who is responsible and accountable when things go wrong?

It is too simplistic to assign the responsibility to a single person, team or company since all are involved; from developing algorithms, to commercial institutions, to legislators and social media giants, to individual citizens, all have parts to play. It is, for example, important for the individual citizen to be aware that whatever they expose on social media has possible consequences on their privacy. Similarly, knowing your rights to privacy and demanding it when necessary are important steps that require awareness of what's happening. Tech giants like Google and Facebook, on a similar note need to be transparent about what they do with our data, how they are playing parts in bringing about societal, political and cultural changes and whose truth they are allowing to flourish and whose they are distorting. They need to be accountable and responsible. Legislative bodies also need to play their part in making sure that there is legislation to begin with and reinforcing them further.

Furthermore, given algorithms are ubiquitous and pervade our social, political, medical and legal systems impacting the process of scientific facts and perception of truths, researching and communicating these issues in no longer a job for a single expertise. In fact, some of the misconception of "algorithms as neutral" and "data telling the truth" can be attributed to lack of communication between and collaborative work between various disciplines that such issues affect. Interdisciplinary and collaborative work, in this regard is invaluable. In fact, as authors of this piece, frustration with lack of interdisciplinary work is what led us to come together and bring our different backgrounds to think about how truths are constructed and reconstructed in computer science.

The importance of interdisciplinary effort cannot be emphasized enough when it comes to interrogating whose truths are being told and whose are left behind. Exposing the unfair and racially biased scoring algorithms of COMPAS [43], a software used across the United States to predict recidivism, for example, took a joint effort from reporters to data scientists to lawyers as well as public awareness and outcry.

Figure 3.1: *Machine Bias - There's software used across the country to predict future criminals. And it's biased against blacks.* - COMPAS [43]

Alongside interdisciplinary collaborations, diverse views (in race and gender, for example) are the last but most important components for approaching the complex explosion of truths and facts in the algorithmic age. Diversity both in terms of those who develop our algorithms and diversity in inputs and data. Whose truth we tell then has less chance of conforming to the status quo.

References

[1] xkcd. *Purity*. xkcd: A webcomic of romance, sarcasm, math, and language. 2018. URL: https://xkcd.com/435/ (visited on 03/30/2018).

[2] University of California Museum of Paleontology. *Misconceptions about science*. Understanding Science. 2018. URL: https://undsci.berkeley.edu/teaching/misconceptions.php (visited on 03/30/2018).

[3] University of Waikato. *Myths of the nature of science*. Science Learning Hub. 2018. URL: https://www.sciencelearn.org.nz/resources/415-myths-of-the-nature-of-science (visited on 03/30/2018).

[4] Anne E. Egger. *Misconceptions and missing conceptions about the process of science*. Process of Science. 2009. URL: https://serc.carleton.edu/sp/process_of_science/misconceptions.html (visited on 03/30/2018).

[5] George F Kneller. "Science as a human endeavor". In: (1978).

[6] Thomas S Kuhn. *The structure of scientific revolutions*. Vol. 2. University of Chicago press Chicago, 1963.

[7] David P Barash. *Paradigms Lost*. Aeon. 2015. URL: https://aeon.co/essays/science-needs-the-freedom-to-constantly-change-its-mind (visited on 03/30/2018).

[8] Cynthia Crossen. *Tainted truth: The manipulation of fact in America*. Simon and Schuster, 1996.

[9] Peter Vanderwicken. "Why the news is not the truth." In: *Harvard Business Review* 73.3 (1995), pp. 144–150.

[10] Hannu-Pekka Ikäheimo. *The algorithmic democracy how did the technological transformation lead to the post-truth era?* Sitra. 2017. URL: https://www.sitra.fi/en/articles/algorithmic-democracy-technological-transformation-lead-post-truth-era/ (visited on 03/30/2018).

[11] Alessandro Bessi and Emilio Ferrara. "Social bots distort the 2016 US Presidential election online discussion". In: *First Monday* 21.11 (2016).

[12] Nicole Kobie. *Who do you blame when an algorithm gets you fired?* Wired. 2016. URL: http://www.wired.co.uk/article/make-algorithms-accountable (visited on 03/30/2018).

[13] Ian Sample. *Computer says no: why making AIs fair, accountable and transparent is crucial*. The Guardian. 2017. URL: https://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial (visited on 03/30/2018).

[14] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. "Unsupervised image-to-image translation networks". In: *Advances in Neural Information Processing Systems*. 2017, pp. 700–708.

[15] J. Thies et al. "Face2Face: Real-time Face Capture and Reenactment of RGB Videos". In: *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*. 2016.

[16] Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman. "Synthesizing obama: learning lip sync from audio". In: *ACM Transactions on Graphics (TOG)* 36.4 (2017), p. 95.

[17] Aäron van den Oord et al. "WaveNet: A Generative Model for Raw Audio". In: *9th ISCA Speech Synthesis Workshop*. 2016, pp. 125–125.

[18] Soroush Mehri et al. "SampleRNN: An unconditional end-to-end neural audio generation model". In: *International Conference on Learning Representations (ICLR2016)*. 2016.

[19] Justus Thies et al. "Face2face: Real-time face capture and reenactment of rgb videos". In: *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*. IEEE. 2016, pp. 2387–2395.

[20] Will Knight. *Real or Fake? AI Is Making It Very Hard to Know*. MIT Technology Review. 2017. URL: `https://www.technologyreview.com/s/604270/real-or-fake-ai-is-making-it-very-hard-to-know/` (visited on 03/30/2018).

[21] James Vincent. *Artificial intelligence is going to make it easier than ever to fake images and video*. The Verge. 2016. URL: `https://www.theverge.com/2016/12/20/14022958/ai-image-manipulation-creation-fakes-audio-video` (visited on 03/30/2018).

[22] Amandha Rohr Lopes. "The impact of social media on social movements: The new opportunity and mobilizing structure". In: *Journal of Political Science Research* (2014).

[23] Deen Freelon, Charlton McIlwain, and Meredith Clark. "Quantifying the power and consequences of social media protest". In: *New Media & Society* 20.3 (2018), pp. 990–1011. DOI: `10.1177/1461444816676646`. eprint: `https://doi.org/10.1177/1461444816676646`. URL: `https://doi.org/10.1177/1461444816676646`.

[24] Andreas Matthias. "The responsibility gap: Ascribing responsibility for the actions of learning automata". In: *Ethics and information technology* 6.3 (2004), pp. 175–183.

[25] Jenna Burrell. "How the machine thinks: Understanding opacity in machine learning algorithms". In: *Big Data & Society* 3.1 (2016), p. 2053951715622512.

[26] Engin Bozdag. "Bias in algorithmic filtering and personalization". In: *Ethics and information technology* 15.3 (2013), pp. 209–227.

[27] Boaz Miller and Isaac Record. "Justified belief in a digital age: On the epistemic implications of secret Internet technologies". In: *Episteme* 10.2 (2013), pp. 117–134.

[28] Brent Daniel Mittelstadt et al. "The ethics of algorithms: Mapping the debate". In: *Big Data & Society* 3.2 (2016), p. 2053951716679679.

[29] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, 2016.

[30] Jacob Brogan. *Google Scrambles After Software IDs Photo of Two Black People as "Gorillas"*. Slate. 2015. URL: `http://www.slate.com/technology/2018/05/smart-clothing-may-be-the-future-of-wearables-but-will-it-really-be-useful.html` (visited on 03/30/2018).

[31] Kaveh Waddell. *How Algorithms Can Bring Down Minorities' Credit Scores*. The Atlantic. 2016. URL: `https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/` (visited on 03/30/2018).

[32] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases". In: *Science* 356.6334 (2017), pp. 183–186. ISSN: 0036-8075. DOI: `10.1126/science.aal4230`. eprint: `http://science.sciencemag.org/content/356/6334/183.full.pdf`. URL: `http://science.sciencemag.org/content/356/6334/183`.

[33] Solon Barocas and Andrew D Selbst. "Big data's disparate impact". In: *Cal. L. Rev.* 104 (2016), p. 671.

[34] Timothy Revell. *Concerns as face recognition tech used to 'identify' criminals*. New Scientist. 2016. URL: `https://www.newscientist.com/article/2114900-concerns-as-face-recognition-tech-used-to-identify-criminals/` (visited on 03/30/2018).

[35] Xiaolin Wu and Xi Zhang. "Automated Inference on Criminality using Face Images". In: ().

[36] Matthias Leese. "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union". In: *Security Dialogue* 45.5 (2014), pp. 494–511.

[37] Solon Barocas. "Data mining and the discourse on discrimination". In: *Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining*. 2014.

[38] Lisa Gitelman. *Raw data is an oxymoron*. MIT Press, 2013.

[39] Rob Kitchin. "Thinking critically about and researching algorithms". In: *Information, Communication & Society* 20.1 (2017), pp. 14–29.

[40] Luciano Floridi. *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford, 2014.

[41] Christian Sandvig et al. "Auditing algorithms: Research methods for detecting discrimination on internet platforms". In: *Data and discrimination: converting critical concerns into productive inquiry* (2014), pp. 1–23.

[42] Carole Cadwalladr and Emma Graham-Harrison. *Cambridge Analytica The Cambridge Analytica Files Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. The Guardian. 2018. URL: `https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election` (visited on 03/30/2018).

[43] Julia Angwin et al. *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.* ProPublica. 2016. URL: `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing` (visited on 03/30/2018).