


Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities?

Andrew Prescott <Andrew_dot_Prescott_at_glasgow_dot_ac_dot_uk>, University of Glasgow  <https://orcid.org/0000-0001-6474-1068>

Abstract

This article surveys the ways in which issues of race and gender bias emerge in projects involving the use of predictive analytics, big data and artificial intelligence (AI). It analyses some of the reasons biased results occur and argues for the importance of open documentation and explainability in combatting these inequities. Digital humanities can make a significant contribution in addressing these issues. This article was written in late 2020, and discussion and public debate about AI and bias has moved on enormously since the article was completed. Nevertheless, the fundamental proposition of this article has become even more important and pressing as debates around AI have progressed – namely, that as a result of the development of big data and AI, it is vital to foster critical and socially aware approaches to the construction and analysis of data. The greatest threat to humanity from AI comes not from autonomous killer robots but rather from the social dislocation and injustices caused by an overreliance on poorly designed and badly documented commercial black boxes to administer everything from health care to public order and crime.

Introduction

In 2015, I attended a workshop in Washington DC which was among the first to focus on big data in the humanities and social sciences. One of the keynote presentations was by Tom Schenk, then Chief Data Officer for the City of Chicago and the co-founder of the Civic Analytics Network at Harvard University's Ash Center for Democratic Governance and Innovation. Under Tom's leadership, Chicago was at the forefront of use of open government data to improve provision of civic services [McBride et al. 2019]. Chicago developed a pioneering open data portal which gave public access to hundreds of data sets [Chicago Data Portal, n. d.]. This data had been used to generate maps and visualisations of evident value to Chicago's citizens, such as maps showing where flu vaccinations were available or which restaurants had al fresco dining licences. 1

Particularly striking was the use in Chicago of data analytics to make predictions which either warned of potential danger or allowed the city to make better use of resources. Predictive analytics programs were developed which identified properties in the city at greatest risk of rodent infestation. This enabled rodent baiting resources to be focussed on particular areas and in 2013 resident complaints about rodents dropped by 15% [Gover 2018, 24]. Another programme used predictive analytics to improve forecasts about the risk of e-coli infection on Chicago's beaches [Lucius et al. 2019]. One of the most successful of the Chicago projects forecasted the risk of a particular restaurant failing a hygiene inspection. This enabled the city to concentrate the efforts of its small team of food hygiene inspectors on those premises where there was a greater likelihood of finding problems [Gover 2018, 23–4] [McBride et al. 2018] [McBride et al. 2019]. This model was in turn used for epidemiological investigation of food poisoning outbreaks [Sadliek et al. 2018]. 2

As Tom's description of the use of predictive analytics in Chicago and other American cities proceeded, however, I felt increasingly uneasy. In New York, predictive analytics were being used to identify which properties were more likely to have illegal flat conversions [Mayer-Schönberger and Cukier 2013, 185–9]. While this has many benefits such as 3

reducing fire risk, it was difficult to escape a feeling that data analytics were being used for greater control of poorer sections of the community. My worries became greater when I later learned about the growing use of data analytics in policing. In Chicago, the police deployed a proprietary technology called ShotSpotter which uses sound sensors across large areas of the city which register where gunshots occur. Another proprietary technology called Hunchlab then used ShotSpotter data to identify localities most likely to have gun crime, enabling police to concentrate resources in those areas. The city has claimed that these technologies reduced crime in the worst districts by about 24%, but these figures are disputed and it seems that the number of crimes detected only by use of ShotSpotter is very small [Wasney 2017]. Predictive policing packages such as ShotSpotter and Hunchlab seem in many ways to be simply a means by which police bear down even more heavily on the poorest and most deprived communities.

In the past five years, the growth of predictive analytics has expanded massively and become even more powerful as it has become linked to machine learning and artificial intelligence (AI). A number of widely publicised cases of bias in AI have confirmed the misgivings I felt as I heard Tom Schenk talk in 2015. It has become evident that AI has the potential to reinforce existing inequalities and injustices. Used carelessly, AI can be a tool to propagate racism, sexism and many other forms of prejudice [O'Neil 2016] [Eubanks 2018]. Tay, the experimental AI chatbot launched by Microsoft in 2016, was within a matter of hours taught to spout racist tweets praising Adolf Hitler [Perez 2016]. In 2015, it was pointed out that Google Photos had labelled pictures of a black man and his friends as "gorillas" [Simonite 2018]. An article in Bloomberg showed how the algorithms determining whether Amazon offers same day delivery frequently excluded postcodes with a significant black population [Ingold and Soper 2016]. An attempt by Amazon to use AI to automatically rank candidates for software development jobs was abandoned after the system systematically excluded women and produced male-only shortlists. Because of the dominance of men in computing, the system taught itself that male candidates were preferable. It downgraded graduates of all-women colleges and penalised resumes that included the word "women" in any context [Lauret 2019].

This article will survey issues of race and gender bias in AI and consider how these may affect the digital humanities. It will make preliminary suggestions as to how practitioners of the digital humanities can help address these disturbing problems. The digital humanities has begun to experiment with the use of AI. Some of these initial applications are in areas where algorithmic bias could potentially present problems, such as the automated analysis of draft legislation and identification of people in archives. As the digital humanities engage more with machine learning and AI, it is likely that use will be made of some tools and methods which caused the sort of biased results which have recently received such bad publicity. Moreover, many humanities scholars and memory institutions are heavily dependent on commercial tools such as Google Images and any suggestion that there is bias in these tools could have serious implications for wider scholarship in the humanities.

Sadly, the days when we might hope that there could be objective tools free from social or cultural bias have vanished, if indeed they ever existed. Information itself has become a site of political contention as significant as gender or race [Jordan 2015] and the political impact of large-scale machine learning tools should be an issue of central concern in the digital humanities. With its tradition of social and cultural activism, digital humanities has great potential to contribute to more ethical approaches to AI and this may be that this is an area in which digital humanities can reshape pervasive "digital modern" cultures [Smithies 2017].

The Enchantment of Big Data and AI

Much of the hype around big data in the early part of the last decade derived from claims that new analytic techniques run on more powerful machines enabled useful scientific findings to emerge spontaneously by observing co-relations in very large and messy datasets. It was suggested that if a dataset was large enough it would compensate for gaps and structural inconsistencies in the data. This stress on observing co-relations was said to be driving an epistemological shift in which there was less emphasis on exactitude and was characterised by the abandonment of a preoccupation with causality (*why*) in favour of finding co-relations (*what*). The importance of letting the data speak for itself was stressed [Mayer-Schönberger and Cukier 2013].

The manifesto for such data-driven methodologies was a notorious article in Wired by Chris Anderson [Anderson 2008]

in which he declared the “end of theory” and suggested that traditional scientific method was obsolete:

There is now a better way. Petabytes allow us to say: “Correlation is enough”. We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot [Anderson 2008].

Objections to Anderson's provocation quickly appeared. It was observed that the predictive analytics used in big data were themselves founded on statistical and mathematical theories [Mayer-Schönberger and Cukier 2013, 70–2]. Callebaut pointed out that Anderson had misrepresented the role of modelling in biological research and reminded Anderson of Darwin's dictum that “all observation must be for or against some view if it is to be of any service” [Callebaut 2012, 74]. Above all, the idea that “raw data” represents an objective factual quarry is an illusion: “raw data is both an oxymoron and a bad idea” [Bowker 2006, 184].

Despite these objections, the idea that new insights can somehow magically emerge from co-relations observed in very large amounts of data has carried over into AI. The computer scientist Stuart J. Russell has commented that:

We are just beginning now to get some theoretical understanding of when and why the deep learning hypothesis is correct, but to a large extent, it's still a kind of magic, because it really didn't have to happen that way. There seems to be a property of images in the real world, and there is some property of sound and speech signals in the real world, such that when you connect that kind of data to a deep network it will – for some reason – be relatively easy to learn a good predictor. But why this happens is still anyone's guess [Campolo and Crawford 2020, 2–3].

This emphasis on the magic of AI has led Alexander Compolo and Kate Crawford [Campolo and Crawford 2020] to compare much discussion of AI with alchemy, the magical properties of algorithms generating what they call “enchanted determinism”. This delight in “enchanted determinism” also encourages subjective responses to data.

The Importance of Explainability

These problems are compounded by the fact that so much AI development is in the hands of commercial companies, with Silicon Valley corporations dominating. Much AI implementation is commercial and the owners of proprietary algorithms are unwilling to explain their business secrets. A great deal can be achieved by reverse engineering algorithms. Nevertheless, it can be very difficult to establish the extent and nature of bias in commercial packages, so that suspicion of prejudice lingers. Silicon Valley companies will react quickly to address criticism but information about exactly how this is done is often sketchy. A great deal can be achieved by reverse engineering algorithms. Nevertheless, it can be very difficult to establish the extent and nature of bias in commercial packages, so that suspicion of prejudice lingers. The default position should perhaps be to regard all commercial AI packages that are not fully documented as biased against particular groups.

Google very quickly changed its search engine in response to the devastating criticisms of Safiya Umoja Noble who meticulously documented how searches on Google in 2011 for “black girls” and “white girls” produced shocking results reflecting racist and sexist stereotypes,[Noble 2018] but details of how Google approached these criticisms are unclear. Sometimes, the response to these issues can be very crude and makes matters worse. Google's reaction to the adverse publicity around the way images of black men were labelled as “gorillas” in Google Photos was to censor the tags, so that no images are ever labelled gorilla, chimpanzee or monkey, even if they are pictures of the primates themselves. Similarly, when a picture of a black man was labelled as an “ape” on Flickr, the term was removed from its tagging lexicon [Hern 2018].

In order to break away from the view of AI as somehow magical and resist the secretive nature of Big Tech, a greater emphasis on explainability – on documenting and discussing the assumptions behind modelling, how this feeds through into algorithms and the properties of the data used – is of vital importance. An insistence on explainability is one of the most important weapons against algorithmic bias. Rob Kitchin identifies two major epistemological approaches to big

data in the scientific community [Kitchin 2014]. On the one hand, those proclaiming the “end of theory” argue that the focus should be on observing surface patterns or anomalies in the data, a highly empirical approach which Kitchen linked to abductive reasoning, a form of logical inference starting with observation of unusual or distinctive patterns and then seeking the simplest explanation. Such an approach creates a high risk of uncritical or superficial analyses of data. On the other hand, other researchers propose that a data-driven science offers the opportunity for creating more holistic and fine grained analyses of very large data sets which can facilitate and foster more critical approaches to data. In investigating the roots of bias in AI, it is essential to adopt this second approach and explore the ways in which models, algorithms and data are constructed. We cannot understand how AI tools share and amplify human prejudices unless we look at the way the data and tools have been created.

It is oversimplistic to assume that prejudice in AI arises simply from poorly constructed algorithms. Bias can be generated by a number of factors, including the quality of data and the nature of the algorithm used. Some of the strategies used can be counterintuitive. It might be assumed that a probabilistic algorithm is more likely to embody faulty cultural assumptions and wrongly identify data concerning black and minority ethnic (BAME) populations than a deterministic algorithm requiring more precise data. However, because UK BAME data is more likely to be variable in quality, with spellings of names and locations inaccurately entered, it turns out that a probabilistic algorithm will be less biased in dealing with BAME data. This can be seen from the linking of UK National Health Service (NHS) records. In order to track the progress of individual patients in the NHS, it is necessary to link records of hospital admissions. A proprietary algorithm called HESID (Hospital Episodes ID) is used to do this. HESID information is used to help calculate commissioning of resources for NHS hospitals. HESID is a deterministic algorithm which requires precise data for such fields as NHS number, date of birth and postcode in order to match names. An analysis of HESID however found that it missed 4.1% of links and made false matches in 0.2% of cases. Moreover, it was ethnic minority patients (Black, Asian, Other) who were disproportionately affected by these missed links. The reasons for this were largely due to the way in NHS numbers were allocated [Hagger-Johnson et al. 2015].

15

In fact, a probabilistic algorithm would have been a far better choice for dealing with data of such variable quality of hospital admission records. A study investigated a probabilistic algorithm which enabled records to be linked when NHS numbers were missing by calculating the probability of a person being the same if other types of information agreed. Use of a probabilistic algorithm substantially reduced the number of missed matches, with particularly beneficial results for ethnic minorities and deprived groups. In the case of emergency hospital admissions for black patients from 1998-2003, the deterministic algorithm missed 7% of matches; the probabilistic algorithm reduced this to 2.3% missed matches. Likewise, in the case of patients from highly deprived socio-economic groups, the deterministic algorithm missed 6.8% of matches, whereas the probabilistic link missed 2.2% [Hagger-Johnson et al. 2017]. The use by the NHS of a deterministic algorithm was doubtless intended to ensure greater precision, but the probabilistic algorithm produced better results. These NHS case studies illustrate the importance of testing a range of different methods and tools and not assuming that one method is inherently superior to the other. Moreover, the results of these testing processes need to be openly available and not constrained by commercial confidentiality, as was the case with the NHS HESID system.

16

The NHS example illustrates how the most effective way of addressing racial and gender bias in AI and machine learning is by digging down into the way the data and tools function and then explaining it. Digital humanities is very well placed to play a major part in developing the explainability of AI. However, much AI implementation is commercial and the owners of proprietary algorithms are unwilling to explain their business secrets. A great deal can be achieved by reverse engineering algorithms, as the analysis above of the HESID algorithm shows. Nevertheless, without explainability, we cannot be sure if the package is biased.

17

The problems caused by the lack of explainability in a commercial AI package are further illustrated by the commercial COMPAS system used in the United States to assess the risk of prisoners reoffending. The use of predictive analytics in policing and the judicial system is particularly contentious. Many American judges, probation and parole officers make use of actuarial risk assessment instruments which automatically calculate the risk of a convict committing another offence after release. There are many of these assessment packages in use. There have been a number of studies that suggest these systems consistently give higher risk scores for black offenders, but it has never been established how the apparent bias occurs [Angwin et al. 2016].

18

In 2016, Pro Publica published a detailed analysis of COMPAS, one of the two commercial packages to assess recidivism [Angwin et al. 2016]. The study concluded that for violent recidivism:

19

Black defendants were twice as likely as white defendants to be misclassified as a higher risk of violent recidivism, and white recidivists were misclassified as low risk, 63.2 percent more often than black defendants.

This seemed to be a clear demonstration of algorithmic bias. However, a rejoinder was rapidly published which pointed to flaws in the Pro Publica analysis. In particular, the Pro Publica analysis used a data set of pre-trial defendants whereas COMPAS was designed to assess the risk of convicted defendants re-offending. Moreover, COMPAS assigned recidivism risk into three categories (low, medium and high) but the Pro Publica article lumped medium and high together as high risk. It was argued that there was no clear evidence of bias in the COMPAS algorithm [Flores et al. 2016]. A further study suggested that COMPAS was no more accurate and fair than predictions made by people with little or no criminal justice expertise which raises the question of whether it is worthwhile using this package, aside from any question of bias [Dressel and Farid 2018] [Holsinger et al. 2018]. It seems likely that the issues with these packages lie not so much in the tools themselves as in the classifications and data produced by the judicial system, particularly the classification of racial types [Benthal and Haynes 2019].

The disagreements about COMPAS illustrate why many of the problems in addressing algorithmic bias lie in the predominance of commercial packages and their lack of explainability. Although a company like Northpointe is comparatively small, it is nevertheless difficult to assess what is going on, even in a small-scale package like COMPAS. Scaling up explainability to analyse the operations of Google or Amazon is almost impossible to imagine. Yet we need to break open the black box if we are going to ensure that AI does not simply amplify and reinforce existing injustices and inequalities.

20

The performance of HESID and COMPAS is comparatively straightforward to analyse. More difficult is to assess the effect of algorithmic bias in natural language processing. A number of studies have documented how natural language processing can absorb human biases from training sets. Word embeddings trained on corpora such as newspaper articles or books exhibit the same prejudices as are evident in the training data. Word embeddings trained on Google news data complete the sentence “Man is to computer programmer as woman is to X” with the word “homemaker” [Bolukabasi et al. 2016]. Another study used association tests automatically to categorise words as having pleasant or unpleasant associations. According to the allocations generated by the algorithm, a set of African American names had more unpleasantness associations than a European American set. The same machine learning programme associated female names more with words like “parent” and “wedding” whereas male names had stronger associations with such words as “professional” and “salary” [Caliskan et al. 2017].

21

Since NLP lies at the root of many services we use every day, these gender and racial biases are imported into tools such as Google Translate. A notorious example was the way in which Google Translate initially dealt with neutral third person pronouns in languages such as Turkish, Hungarian and Finnish. Until recently, Google Translate rendered the Turkish sentences “o bir doktor” and “o bir hemşire” into English as “he is a doctor” and “she is a nurse” and the Hungarian “ő egy ápoló” as “she is a nurse”, despite the fact that the pronouns are not gender specific [Caliskan et al. 2017] [Prates et al. 2019]. This has now been corrected by Google and alternative pronouns are offered in the translation [Johnson 2020]. The Facebook translation service can also be problematic. A Palestinian was arrested by Israeli police because Facebook’s AI translation service wrongly translated the Arabic words for “good morning” as “hurt them” in English or “attack them” in Hebrew [Hern 2017]. Bias is also evident in other forms of linguistic analysis. In a test of gender and race bias in sentiment analysis systems, it was found that African American names scored higher in anger, fear and sadness, and European American names scored higher on emotions such as joy [Kiritchenko and Mohammed 2018]. The social media filter *Perspective* developed by a Google-backed incubator marks innocuous African American vernacular phrases as “rude” and categorised the statement “I am a gay black woman” as 87% toxic [Chung 2019].

22

In such cases as the problem of the gender-neutral pronoun, companies like Google are quick to try and correct blatant

23

examples of prejudice when reported by researchers. But the methods used to try and correct such problems are often crude and create as many problems as they solve. The most common method is to implement a blacklist of banned words and concepts. This was the method used to deal with the problems of Microsoft's ill-fated chatbot, *Tay*. A few months after *Tay* was taken down, Microsoft launched a replacement, *Zo*, which ran until summer 2019. *Zo* was told to shut down the conversation if words like the Middle East, Jew or Arab were mentioned. However, this was done without reference to context, so that a statement like "That song was played at my bah mitzvah" elicited the response "ugh, pass, I'd rather talk about something else". Because of the concern to ensure *Zo* was not taught to attack Jews, Microsoft ended up giving the distinct impression that *Zo* was anti-semitic [Stuart-Ulin 2018].

Many of issues of bias in AI arise from the way in which language is dealt with. The failure of *Zo* is due to its inability to deal with context. Language is of course very much the domain of the digital humanities and again digital humanities has a great deal to offer in addressing these problems. The prominent digital humanities specialists Professor Melissa Terras and David Beavan recently took part in an experiment to automatically generate a Queen's Christmas message using corpora of earlier Christmas broadcasts. The AI Queen's Christmas message contained a great deal of racist and sexist content. Terras observed that "I don't think we've really begun to train our computational systems in the philosophy of language ... And that's why these conversations between computer science folks and humanities people are so important" [Kobie 2020]. This is an urgent agenda for digital humanities in the twenty-first century.

24

Ubiquitous Dangers

As the vision of ubiquitous computing is achieved and AI penetrates every aspect of our life, the effects of gender and race bias in AI are becoming increasingly pressing. Alexa is in danger of becoming a powerful force for racism and sexism in society. As we rely increasingly on voice interaction with computers, we anthropomorphise HCI and thereby cease to notice the prejudices and biases embodied in them. Frictionless engagement with a computer is also often uncritical engagement.

25

Automated speech recognition systems are becoming an increasingly familiar part of everyday life, powering virtual assistants, facilitating automated closed captioning and enabling digital dictation platforms for health care. In a 2018 survey, 45.3% of respondents from Wales, 45.2% from Scotland and 45.1% from Yorkshire reported that they had difficulty being understood by smart home devices [Coleman 2018]. Lower accuracy in You Tube closed captioning has been found for women and speakers from Scotland [Tatman 2017].

26

A 2018 Washington Post report found significantly lower accuracy of recognition by Amazon Echo and Google Home of speakers from the Southern United States and those with Indian, Spanish or Chinese accents. The data scientist Rachel Tatman commented that: "These systems are going to work best for white, highly educated, upper-middle-class Americans, probably from the West Coast, because that's the group that's had access to the technology from the very beginning"[Harwell 2018]. It has been long recognised that natural language processing does not accommodate African American speech patterns, and this has carried over into speech recognition systems. A study recently published in the Proceedings of the National Academy of Sciences used the Corpus of Regional African American Language to analyse the performance of automated speech recognition systems and found performance was significantly poorer for African Americans [Koenecke et al. 2020]. The authors of the study commented that:

27

Our findings indicate that the racial disparities we see arise primarily from a performance gap in the acoustic models, suggesting that the systems are confused by the phonological, phonetic, or prosodic characteristics of African American Vernacular English rather than the grammatical or lexical characteristics. The likely cause of this shortcoming is insufficient audio data from black speakers when training the models.

The performance gaps we have documented suggest it is considerably harder for African Americans to benefit from the increasingly widespread use of speech recognition technology, from virtual assistants on mobile phones to hands-free computing for the physically impaired. These disparities may also actively harm African American communities when, for example, speech recognition software is used by employers to automatically evaluate candidate interviews or by

criminal justice agencies to automatically transcribe courtroom proceedings. [Koenecke et al. 2020, 7687]

A major issue with addressing these issues is the restricted availability of voice training data much of which is under the control of the larger Silicon Valley corporations. The Mozilla Foundation's *Common Voice* project was an attempt to create a more diverse and representative voice training data set [Common Voice n.d.]. The failure to create more responsive speech recognition systems reflects the lack of diversity in the Silicon Valley corporations which have developed this technology. Ruha Benjamin reports that when a member of the team which developed Siri asked why they were not considering African American English, he was told "Well, Apple products are for the premium market". This happened in 2015, one year after Dr Dre sold Beats by Dr Dre to Apple for a billion dollars. Benjamin comments on the irony of the way in which Apple could somehow devalue and value Blackness at the same time [Benjamin 2019, 28].

28

Siri, Alexa and their friends are not only racist but sexist as well. Lingel and Crawford have shown how Siri, Alexa, Cortana and other soft AI technologies

29

typically default to a feminine identity, tapping into a complex history of the secretary as a capable, supportive, ever-ready, and feminized subordinate ... These systems speak in voices that have feminine, white, and "educated" intonation, and they simultaneously harvest enormous amounts of data about the user they are meant to serve [Lingel and Crawford 2020, 2].

Although Siri, Alexa et al. offer various customisation options, including now in the case of Alexa the voice of the black American actor Samuel L. Jackson, the default is female and submissive. In choosing the voice for Alexa, Amazon had a very concrete view of the sort of person Alexa should be:

30

She comes from Colorado, a state in a region that lacks a distinctive accent. "She's the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern", [the head designer] continues. When she was a child, she won \$100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to "a very popular late-night-TV satirical pundit." And she enjoys kayaking [Lingel and Crawford 2020, 10].

While this characterisation of Alexa harks back to retrograde views of the sort of woman who makes a desirable secretary, on the other hand, as Lingel and Crawford [Lingel and Crawford 2020] emphasise, there is also a long tradition of secretaries being viewed as trusted custodians of confidential information. The friendly approachable character of Alexa makes you confident and relaxed as she absorbs and transmits to Amazon masses of personal data.

31

The more frictionless and ubiquitous technology becomes, the greater is the scope for exclusion and bias. Perhaps the most alarming from this point of view of technologies currently being rolled out is facial recognition. A seminal paper by Buolamwini and Gebu [Buolamwini and Gebu 2018] evaluated three commercially available systems by IBM, Microsoft and the Chinese company Megvii (Face++) which used facial recognition to make gender allocations. They found that darker-skinned females were the most misclassified group (with error rates of up to 34.7%), whereas the maximum error rate for lighter-skinned males was 0.8%. This bias was due to the lack of training sets with a sufficiently diverse range of images.

32

As facial recognition is increasingly used in border control, policing, store and building security, and many other purposes, these problems are becoming increasingly pressing. A further study by Raji and Buolamwini [Raji and Buolamwini 2019] investigated bias in Amazon's *Rekognition* system which had been widely marketed to police forces and judicial agencies. This showed that gender classification by the Amazon system was even more biased than in IBM, Microsoft and Megvii systems tested in the original study, with Amazon's *Rekognition* producing error rates of 31.37% for darker-skinned females and 8.66% for lighter-skinned males [Raji and Buolamwini 2019]. Amazon disputed the claims [Wood 2019], but it was emphasised by Buolamwini that Amazon had refused to submit *Rekognition* to evaluation by the National Institute of Standards and Technology (NIST) and its claims that *Rekognition* was bias free were based only on internal testing [Buolamwini 2019].

33

The tests performed by Buolamwini and her colleagues were concerned with gender classification, but inevitably raise doubts about other aspects of facial recognition packages such as identification of individuals. A 2019 NIST report found that there was indeed also bias in the use of facial recognition software to identify individuals [Grother et al. 2019]. It showed that Native American, West African, East African and East Asian people were far more likely to be wrongly identified in US domestic applications. Women were also more likely to be wrongly identified. In the case of border crossing controls, false negatives were much higher among people born in Africa and the Caribbean. In the wake of these findings and in response to the *Black Lives Matter* movement, IBM, Microsoft and Amazon all stepped back from active commercial promotion of their products [Page 2020].

34

How Should Digital Humanities Respond to This?

These are issues that should be of profound concern to practitioners of the digital humanities. Areas such as natural language processing, nominal record linkage and image recognition are of fundamental importance to the digital humanities. Thinking about how computers handle language and context is at the heart of much digital humanities research. Corpus linguists document dialect and shifting usage, and can make major contributions to more inclusive training sets for development of voice recognition software. The strong understanding of governance, regulation and transparency in both the humanities and social sciences can make a major contribution to developing governance frameworks for a more accountable and transparent AI. Digital humanities scholars such as David Berry have been at the forefront of promoting explainability in AI [Berry 2019].

35

Above all, some of these technologies are already being employed in digital humanities and there can be no doubt that, as scholars in the humanities seek to come to terms with vast quantities of born-digital data, AI tools will become of fundamental importance in humanities research. Historians and other humanities scholars will not be able to analyse the hundreds of millions of e-mails produced by governments and corporations or attempt to probe the terabytes of data produced by web archives without the aid of AI tools [Winters and Prescott 2019]. If the historical research of the future is going to be fair-minded, unbiased and just, then it will need an AI that is subject to rigorous testing, transparent in its assumptions and extensively documented.

36

AI will also be of fundamental importance to historians in the future because it will be one of the key tools used by archivists to manage born-digital data. It will be impossible for archivists manually to catalogue the petabytes of data that are already being produced by governments and corporations. Instead it is probable that finding aids will be generated by automated AI extraction of metadata [Findlay and Sheridan 2018a]. The use of AI will also be important in appraising what born-digital data should be preserved for historians and transferred to archives. AI will probably also be used in deciding which born-digital records contain sensitive information that mean they should be closed from public access [Findlay and Sheridan 2018b]. AI will without doubt be a leading force in shaping the future historical record.

37

Illustrations of some of the likely future use of AI in managing archives and libraries are given by two projects undertaken by the UK National Archives funded by the Arts and Humanities Research Council under its “Digital Transformations” strategic theme. Legal codes are now too vast to be mastered by manual reading. The UK statute book comprises 50 million words with 100,000 words changed or added every month. The *Big Data for Law* project investigated how AI methods can make it easier to understand how legislation is structured and used [UK Legislation, n.d., n.d.]. It developed tools which not only assisted in developing an overview of legislation but also suggested ways in which legislation could be improved. The second project, *Traces Through Time*, used AI to identify different mentions of a person in the archive and to build links with them [Ranade 2016].

38

Both of these pioneering projects not only give a glimpse of the likely future role of AI in the archives but also indicate some of the future ethical issues which archivists, librarians and humanities scholars will need to confront. How do we feel about machines drafting legislation which controls our behaviour? How do we know what biases and prejudices may be embedded in the tools which may be developed for legislators? Likewise, if there are clear patterns of bias in linkage in health records, how do we know that is not happening in historical archives? As humanities scholars start to make use of the possibilities provided by AI, there is a risk that humanities scholarship can become polluted by hidden gender and race bias unless the AI used is transparent, accountable and explainable.

39

Other pioneering applications of AI may on the surface seem to have a minimal risk of bias but on further examination possibilities emerge that results may be distorted by class, race or gender. For example, many studies using “distant reading” techniques make use of Google books as a base set. This may seem reasonable since Google Books purports to cover all published books. However, Google has a very top-down view of the world's knowledge and naively imagines that the great research libraries such as those at Harvard, Toronto or Oxford containing everything worth knowing. This is wrong, and Google Books omits many local or limited circulation publications are only available in local libraries whose catalogues may not even be online. Thus, if we use Google Books to analyse working-class autobiographies describing the experience of the Industrial Revolution, we find that there are significant gaps in the Google Book coverage, so that the Google sample gives disproportionate prominence to the autobiographies of successful self-made man and excludes the voices of more humble workers [Prescott 2014].

40

An important role of digital humanities in the future will be in benchmarking and documenting AI performance in areas relevant to humanities scholarship. For much of its history since the 1950s, practitioners of humanities computers and the digital humanities have had to be evangelists for the use of computers in humanities scholarship. There are still many battles to be fought over such questions as the extent to which scholars should themselves be coders or the role of quantification in humanities scholarly discourse. But increasingly as humanities scholars adopt digital methods, an important role of the digital humanities should be to promote a critical approach to the use of digital tools and methods in the humanities. Too often, scholars are happy to use n-grams or visualisations to illustrate pet theories without thinking about how the tool works or the nature of the underlying data. As AI tools and methods become increasingly available to humanities scholars, this role will be increasingly important.

41

Digital humanities is exceptionally well placed to promote an ethical AI. It is widely agreed that, in combatting algorithmic bias, an interdisciplinary approach is essential and the interdisciplinary traditions of digital humanities can make a vital contribution here. Cultural and media specialists can contribute to combatting bias in design; historians and linguists can assist in assessing the linguistic and other contexts that might generate bias. The debates around the COMPAS system to predict recidivism risk discussed above can be best understood in the context of the long and complex history of racial classification in the United States [Benthal and Haynes 2019], and such systems would perform much better if they had historians on the development team. Again, it is also agreed that in avoiding algorithmic bias, it is vital that design teams are themselves diverse in makeup. While the track record of digital humanities in ethnic and gender inclusiveness is far from perfect, there is nevertheless a strong emphasis on the importance of diversity. Digital humanities can contribute to a more inclusive and diverse approach to AI development. One area where this could be particularly important is in drawing on the experience of digital humanities with a wide range of historic, linguistic and other primary materials to create more diverse training sets for AI applications.

42

Algorithmic bias is potentially a major social and cultural crisis for humanity. It is an area where digital humanities can make a major contribution. In developing approaches to these issues, digital humanities practitioners can helpfully draw on an increasing range of recent work which outlines best practice and principles for responsible use of AI in society [Padilla 2019] [Floridi and Cowsls 2019]. Work towards an ethical AI may perhaps represent finally a coming of age for digital humanities. How might this look as a concrete plan of action? In conclusion, it might be worth setting out a short manifesto for DH AI which itemises ten key areas worth early attention. Space prevents me offering extended rationales for each action point, but it is nevertheless helpful briefly to outline them.

43

1. In many respects, digital humanities associations and organisations are often inward looking and do not pursue wider social agendas. There is room for greater dialogue with activist organisations seeking to promote the health of our digital environment. Many individual digital humanities practitioners work with Mozilla Foundation, a leading campaigner in this area [Mozilla, n. d.], and there is scope for more extended and structured engagement. Links might also be built with other campaigns, such as the Algorithmic Justice League [Algorithmic Justice League, n. d.] and Women in Voice [Women in Voice, n. d.]
2. Digital humanities offers many examples of best practice in diversity and inclusiveness for all workers in every aspect of Information Technology. As a community, we should seek to document and increase awareness of such good practice and demonstrate the benefits it brings in creating a healthier digital environment. Digital humanities has been hugely successful in encouraging reluctant and suspicious user

communities to engage with digital technology. We need to be equally forceful in encouraging humanities scholars to be highly critical and self-aware as their work becomes increasingly dependent on digital tools of all kinds.

3. The digital humanities should give priority to the articulation of international governance and benchmarking structures for the use of AI. The humanities provides an exciting venue for the exploration and articulation of such structures which can be a model for other sectors and disciplines [Cihon 2019].
4. There is a strong fit between humanities and the requirement to develop explainability. Humanities scholars have a strong awareness of the way in which the formation and processing of information is shaped by cultural and social factors. The humanities can play a major role in promoting explainability in AI, and the background and structure of digital humanities makes it an excellent vehicle to promote work in this area.
5. In this context, we need to continue to promote awareness of bias and prejudice in the history of digital humanities itself. Gender and race biases are embedded in tools such as TEI or library and museum classification systems [Lu and Pollock 2019] [Olson 2001] [Leung and López-McKnight 2021] [Macdonald 2022]. They are even evident in the Wikidata we use for linking. Root these out. The insights you gain will assist in rooting out algorithmic bias.
6. Increasingly as they deal with large image, catalogue and audio-visual data sets, libraries, museums, art galleries and archives are becoming more reliant on and expert in automated and machine learning techniques. This engagement of the heritage sector with AI will become even more important as they deal with more born-digital material. Libraries, museums, art galleries and archives can provide more diverse training data for AI in language, image and sound.
7. Language and language processing will be all important for many future developments in AI. The humanities can play a central role here. We need to continue to prioritise linguistic research in digital humanities in a way that will help tackle problems like linguistic context in AI.
8. Be more self aware. Ask ourselves what effects algorithmic bias are having within our home humanities disciplines and how we can promote awareness of this.
9. Be aware of how AI is being used in our own university environments. Amazon is promoting the use of its Rekognition facial software for proctoring of university examinations and to detect cheating. Facial recognition software is also being introduced in British schools to enable security checks and cashless payments [Winchester 2021]. Challenge such developments.
10. Develop new narratives of AI. The narratives around AI at present are too often about control, monitoring, efficiency. There are other ways we might use AI. Imagine them, suggest them and promote them.

Works Cited

Algorithmic Justice League, n. d. *Algorithmic Justice League*. <https://www.ajl.org>

Anderson 2008 Anderson, C. "The End of Theory: the Data Deluge Makes the Scientific Method Obsolete". *Wired*. 23 June 2008. <https://www.wired.com/2008/06/pb-theory/>

Angwin et al. 2016 "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks". *Pro Publica*. 23 May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Benjamin 2019 Benjamin, R. *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity Press. Cambridge (2019).

Benthal and Haynes 2019 Benthal, S. and Haynes, B. D. "Racial Categories in Machine Learning". *FAT* '19 Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 289-98. Atlanta, Georgia. January 2019. DOI: <https://doi.org/10.1145/3287560.3287575>.

Berry 2019 Berry, D. "The Explainability Turn". <https://stunlaw.blogspot.com/2020/01/the-explainability-turn.html>.

Bolukabasi et al. 2016 Bolukabasi, T., Chang, K., Zou, J., Saligrama, V. and Kalai, A. "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". *29th Conference on Neural Information Processing Systems (NIPS)*.

Bowker 2006 Bowker, Geoff. *Memory Practices in the Sciences*. MIT Press. Cambridge, Ma. (2006).

- Buolamwini 2019** Buolamwini, J. "Response: Racial and Gender bias in Amazon Rekognition - Commercial AI System for Analyzing Faces". <https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced>
- Buolamwini and Gebru 2018** Buolamwini, J. and Gebru, T. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research* 81: 77-91.
- Caliskan et al. 2017** Caliskan, A., Bryson, J. J. and Narayanan, A. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases". *Science* 356: 183-6. DOI: 10.1126/science.aal4230.
- Callebaut 2012** Callebaut, W. "Scientific perspectivism: a philosopher of science's response to the challenge of big data biology." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. 43 (2012): 69-80.
- Campolo and Crawford 2020** Campolo, A. and Crawford, K. "Enchanted Determinism: Responsibility in Artificial Intelligence." *Engaging Science, Technology, and Society* 6 (2020): 1-19. DOI: 10.17351/ests2020.277
- Chicago Data Portal, n. d.** "Chicago Data Portal". <https://data.cityofchicago.org/>.
- Chung 2019** Chung, Anna. "How Automated Tools Discriminate Against Black Language." *POCIT*. January 2019. https://peopleofcolorintech.com/author/anna_chung/
- Cihon 2019** Cihon, P. "Technical Report: Standards for AI Governance: International Standards to Enable Global Coordination in AI Research and Development." Future of Humanity Institute, University of Oxford. https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-_FHI-Technical-Report.pdf
- Coleman 2018** Coleman, D. "The Dialect of Tech". <https://spike.digital/2018/08/28/the-dialect-of-tech/>.
- Common Voice n.d.** "Mozilla Common Voice project." <https://commonvoice.mozilla.org/en>.
- Dressel and Farid 2018** Dressel, J. and Farid, H. "The accuracy, fairness, and limits of predicting recidivism." *Science Advances*. 4.1. DOI: DOI: 10.1126/sciadv.aao5580.
- Eubanks 2018** Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. Macmillan. London. (2018)
- Findlay and Sheridan 2018a** Findlay, C. and Sheridan, J. "Recordkeeping Roundcasts Episode 1: Scale and complexity, with John Sheridan". <https://rkroundtable.org/2018/08/01/recordkeeping-roundcasts-episode-1-scale-and-complexity-with-john-sheridan/>
- Findlay and Sheridan 2018b** Findlay, C. and Sheridan, J. "Recordkeeping Roundcasts Episode 3: Machine learning". <https://rkroundtable.org/2018/08/19/recordkeeping-roundcasts-episode-3-machine-learning/>
- Flores et al. 2016** Flores, A., Bechtel, K. and Lowencamp, C. "False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks'". *Federal Probation* 80.2: 38-46.
- Floridi and Cowsls 2019** Floridi, Luciano, and Josh Cowsls. "A Unified Framework of Five Principles for AI in Society". *Harvard Data Science Review* (2019). DOI: <https://doi.org/10.1162/99608f92.8cd550d1>
- Gover 2018** Gover, Jessica. "Analytics in City Government: How the Civic Analytics Network Cities are Using Data to Support Public Safety, Housing, Public Health, and Transportation". Ash Center for Democratic Governance and Innovation. Harvard Kennedy School. 2018.
- Grother et al. 2019** Grother, P., Ngan, M. and Hanaoka, K. "Face Recognition Vendor Test (FRVT) Part 2: Identification. NISTIR 8271". *National Institute of Standards and Technology*. DOI: <https://doi.org/10.6028/NIST.IR.8271>
- Hagger-Johnson et al. 2015** Hagger-Johnson, G., Harron, K., Fleming, T., Gilbert, R., Goldstein, H., Landy, R. and Parslow, R. C. "Data Linkage Errors in Hospital Administrative Data When Applying a Pseudonymisation Algorithm to Paediatric Intensive Care Records". *BMJ Open* 5: 1-8. DOI: <http://dx.doi.org/10.1136/bmjopen-2015-008118>.
- Hagger-Johnson et al. 2017** Hagger-Johnson, G., Harron, K., Goldstein, H., Aldridge, R. and Gilbert, R. "Probabilistic Linking to Enhance Deterministic Algorithms and Reduce Linkage Errors in Hospital Administrative Data". *Journal of Innovation in Health and Informatics* 24. 2: 234-46. DOI: 10.14236/jhi.v24i2.891.
- Harwell 2018** Harwell, D. "The Accent Gap". *Washington Post*. 19 July 2018. <https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/>.

- Hern 2017** Hern, A. "Facebook Translates 'Good Morning' into 'Attack Them' Leading to Arrest". *The Guardian*, 24 October. <https://www.theguardian.com/technology/2017/oct/24/facebook-palestine-israel-translates-good-morning-attack-them-arrest>.
- Hern 2018** Hern, A. "Google's solution to accidental algorithmic racism: ban gorillas". *The Guardian*. 12 January 2018. <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people>
- Holsinger et al. 2018** Holsinger, A., Lowenkamp, C., Laessa, E., Serin, R., Cohen, T., Robinson, C., Flores, A. and Vanbenschoten, S. "A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm is More Accurate than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts". *Federal Probation* 82.2: 51-6.
- Ingold and Soper 2016** Ingold, D. and Soper, S. "Amazon Doesn't Consider the Race of Its Customers. Should It?" *Bloomberg*. 21 April 2016. <https://www.bloomberg.com/graphics/2016-amazon-same-day/>
- Johnson 2020** Johnson, M. "A Scalable Approach to Reducing Gender Bias in Google Translate". *Google AI Blog*. 22 April 2020. <https://ai.googleblog.com/2020/04/a-scalable-approach-to-reducing-gender.html>.
- Jordan 2015** Jordan, Tim. *Information Politics: Liberation and Exploitation in the Digital Society*. Pluto Books. London (2015).
- Kiritchenko and Mohammed 2018** Kiritchenko, S and Mohammed, S. "Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems". *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics (*SEM)*, pp. 43-53. New Orleans. June 2018. <https://www.aclweb.org/anthology/S18-2005DOI: 10.18653/v1/S18-2005>.
- Kitchin 2014** Kitchin, Rob. "Big Data, New Epistemologies and Paradigm Shifts". *Big Data and Society*. 1.1: 1-12.
- Kobie 2020** Kobie, Nicole. "We asked an AI to write the Queen's Christmas speech". *Wired*. 24 December 2020. <https://www.wired.co.uk/article/ai-queens-speech-christmas-day>.
- Koenecke et al. 2020** Koenecke et al. 2020] Koenecke, A., Narn, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, R., Jurafsky, D., and Goel, S. "Racial Disparities in Automated Speech Recognition". *Proceedings of the National Academy of Sciences of the United States of America*. 117: 7684-89. DOI: <https://doi.org/10.1073/pnas.1915768117>.
- Lauret 2019** Lauret, J. "Amazon's Sexist AI Recruiting Tool: How Did It Go So Wrong?". *Becoming Human: Artificial Intelligence Magazine*. 16 August 2019. <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e>.
- Leung and López-McKnight 2021** Leung, S. Y., and López-McKnight, J. R. eds. *Knowledge Justice: Disrupting Library and Information Studies through Critical Race Theory*. Cambridge, MA: MIT Press (2021).
- Lingel and Crawford 2020** Lingel, Jessa and Crawford, Kate. "'Alexa, Tell Me about Your Mother:' The History of the Secretary and the End of Secrecy". *Catalyst: Feminism, Theory, Technoscience* 6.2: 1-22. DOI: <https://doi.org/10.28968/cftt.v1i1.28809>.
- Lu and Pollock 2019** Lu, Jessica and Pollock, Caitlin. "Digital Dialogue: Hacking TEI for Black Digital Humanities". *MITH in MD presentation* 5 Nov. 2019. <https://vimeo.com/372770114>.
- Lucius et al. 2019** Lucius, N., Rose, K., Osborn C., Sweeney, M. E., Chesak R., Beslow S. and Schenk T. "Predicting E. Coli Concentrations Using Limited qPCR Deployments at Chicago Beaches". *Water Research* X. 2. 100016. DOI: 10.1016/j.wroa.2018.100016
- Macdonald 2022** MacDonald, Sharon, ed. *Doing Diversity in Museums and Heritage: a Berlin Ethnography*. New York. Columbia University Press (2022).
- Mayer-Schönberger and Cukier 2013** Mayer-Schönberger, V. and Cukier, K. *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. New York. Eamon Dolan (2013).
- McBride et al. 2018** McBride K., Aavik G., Kalvet T and Krimmer R. "Co-creating an Open Government Data Driven Public Service: The Case of Chicago's Food Inspection Forecasting Model." *Proceedings of the 51st Hawaii International Conference on System Sciences* 2018. DOI: 10.24251/HICSS.2018.309
- McBride et al. 2019** McBride K., Aavik G., Toots M., Kaivet T. and Krimmer R. "How does open government data driven co-creation occur? Six factors and a 'perfect storm'; insights from Chicago's food inspection forecasting model." *Government Information Quarterly* 36, pp. 88-97. DOI: <https://doi.org/10.1016/j.giq.2018.11.006>
- Mozilla, n. d.** *Mozilla Foundation*. <https://foundation.mozilla.org>.

- Noble 2018** Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. New York. (2018).
- Olson 2001** Olson, Hope A. "The Power to Name: Representation in Library Catalogs". *Signs* 26: 639-68.
- O'Neil 2016** O'Neil, Cathy. *Weapons of Math Destruction*. Crown Books. New York. (2016).
- Padilla 2019** Padilla, Thomas. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries* Dublin, OH. OCLC Research (2019).
- Page 2020** Page, R. "Spotlight on Facial Recognition after IBM, Amazon and Microsoft Bans". *CMO*. 16 June 2020. <https://www.cmo.com.au/article/680575/spotlight-facial-recognition-after-ibm-amazon-microsoft-bans/>
- Perez 2016** Perez, S. "Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism". *TechCrunch*. 24 March 2016. <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>
- Prates et al. 2019** Prates, M., Avelar, P. and Lamb, L. "Assessing gender bias in machine translation: a case study with Google Translate". *Neural Computing and Applications* 32: 6363-81.
- Prescott 2014** Prescott, A. "I'd Rather be a Librarian". *Cultural and Social History*, 11.3, 335-41. DOI: 10.2752/147800414X13983595303192.
- Raji and Buolamwini 2019** Raji, I. D. and Buolamwini, J. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products". *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 429-35. DOI: <https://doi.org/10.1145/3306618.3314244>.
- Ranade 2016** Ranade, S. "Traces through Time: A Probabilistic Approach to Connected Archival Data". *2016 IEEE Conference on Big Data*, Washington D.C.. December 2016. DOI: 10.1109/BigData.2016.7840983
- Sadliek et al. 2018** Sadliek, A., Caty, S., DiPrete, L., Mansour, R., Schenk, T., Bergtholdt, M., Jha, A., Ramaswami, P., and Gabrilovich, E. "Machine-learned epidemiology: real-time detection of foodborne illness at scale". *npj Digital Medicine* 1, 36 (2018). DOI: <https://doi.org/10.1038/s41746-018-0045-1>
- Simonite 2018** Simonite, T. "When It Comes to Gorillas, Google Photos Remains Blind". *Wired*. 1 November 2018. <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>
- Smithies 2017** Smithies, J. *The Digital Humanities and the Digital Modern*. Palgrave Macmillan. London (2017).
- Stuart-Ulin 2018** Stuart-Ulin, Chloe Rose. "Microsoft's Politically Correct Chatbot is Even Worse than its Racist One." *Quartz*. 30 July 2018. <https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/>.
- Tatman 2017** Tatman, R. "Gender and Dialect Bias in YouTube's Automatic Captions". *Proceedings of the First Workshop on Ethics in Natural Language Processing*, pp. 53-9. Valencia, Spain. April 2017.
- UK Legislation, n.d.** "Big Data for Law." <https://www.legislation.gov.uk/projects/big-data-for-law>.
- Wasney 2017** "The Shots Heard Round the City: Are Chicago's New Shot Detection and Predictive Policing Technologies Worth It?". *South Side Weekly*. 19 December 2017. <https://southsideweekly.com/shots-heard-round-city-shotspotter-chicago-police/>
- Winchester 2021** "Facial Recognition in Schools". *House of Lords Library In Focus*. November 2021. <https://lordslibrary.parliament.uk/facial-recognition-technology-in-schools/>
- Winters and Prescott 2019** Winters, J. and Prescott, A. "Negotiating the born-digital: a problem of search". *Archives and Manuscripts*. 47.3: 391-403. DOI: <https://doi.org/10.1080/01576895.2019.1640753>.
- Women in Voice, n. d.** *Women in Voice*. <https://womeninvoice.org/>
- Wood 2019** Wood M. "Thoughts on Recent Research Paper and Associated Article on Amazon Rekognition". *AWS Machine Learning Blog*. 26 January 2019. <https://aws.amazon.com/blogs/machine-learning/thoughts-on-recent-research-paper-and-associated-article-on-amazon-rekognition/>

