What is the importance of ethics in data science?

The technology industry is plagued with a preoccupation with what data science *can* achieve rather than what it *should*. As a result, the data science industry faces numerous ethically critical issues, all of which have the capability to pose threats to society in the near and eventual future. The exemption of mined research data from ethical oversight now operates at a near conspiratorial level, with major US universities and journals requiring little or no Institutional Review Board approval for research conducted on farmed social media data. The environmental impact of the growth of data industries could be catastrophic, with the carbon footprint produced training Google AI Transformer alone equivalent to 288 transatlantic flights. Additional consideration must be given to the ethics of the overall direction of neural network research when deep machine learning algorithms now have the processing abilities to, for example, identify members of the LGBTQ+ community with significantly higher accuracy than humans, a technology that poses a serious threat to gay people in countries where homosexuality is illegal or socially unacceptable. I believe, however, algorithmic bias to be the most pervasive and objectionable problem currently reaching a point of no return in the data science industry which is why I have chosen it as the focus for this essay.

The number of algorithms used to allocate and assist services across the world is increasingly rapidly because of the assumption that sociologically driven data science provides an ethically neutral solution to human prejudice. The flawed logic behind this claim breaks down into three distinct problems: the inevitability of algorithmic bias; the lack of transparency perpetuated both by the algorithms and the companies that use and create them; and the reputation held by data science that serves only to exacerbate the issue further.

Emma Chandler

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), an algorithmic system used by courts to predict the probability of recidivism in several US states, represents the culmination of the three prominent issues associated with algorithmic bias I will outline. The system uses formulas that are created by data scientists but include weights derived using machine-learning of historical data.

There are several features of sociological algorithms that means that they inevitably incorporate bias and reflect the bias already existent in society, encapsulating it permanently. Algorithms that use elements of machine learning such as COMPAS are trained using a dataset manufactured from historical human decisions, which inevitably contains general bias in the data that has been collected as well as incidents of flawed or prejudicial human judgement. Instead of replacing such prejudice with fair and rational logic, the bias is integrated and learnt by the algorithm, giving it a new lease of life within the system. In non-AI algorithms, bias is incorporated through the data scientists' decisions as to which weighting to apply to each data input. Regardless of whether it is included because of the programmer's own prejudices or ones present in historical data, algorithmic bias effectively gives immortality to the status quo. It can range from men more regularly receiving targeted ads for higher-paying jobs than women, based on data informed by the gender pay gap, to loans and other services being denied in majority black areas in American cities as an everlasting effect of the Jim Crow Laws' legalisation of redlining, which led to racial inequity and economic disparity within cities that continues to the present day.

Superficially, the COMPAS recidivism prediction system appears well-positioned to provide a significant improvement upon human judgement as it removes accidental bias, such as a judge's tendency to rule more favourably following a meal as well as intentional discrimination. Nonetheless,

despite race not being inputted as one of the 137 pieces of data that defendants must provide, nor included in the original data set used to train the algorithm, traces of historical racial discrimination are impossible to remove completely. This is due to the algorithm's capability to use other related data such as racially coded names as well as formally red-lined addresses as a proxy for race data, as has been admitted by the Chief Scientist behind COMPAS. This existing bias is exemplified in a ProPublica investigative study of 10,000 offenders which found that while the COMPAS system incorrectly labelled white and black defendants at roughly the same rate, black defendants were more than twice as likely to be mistakenly considered high risk than white defendants while the opposite was true for incorrect low-risk judgements.

Algorithmic bias is difficult to spot yet even harder to correct due to the way that existing societal biases are crystalised and concealed within the complexities of the system. It is only possible to evaluate the input and output of a vastly intricate algorithm, the rest of the process is a convoluted combination of program code, statistical models, and vast quantities of data. The obscurity of biases applied to the data is further amplified by the private ownership of systems used for government decisions. The companies that create them have intellectual property rights over the systems and therefore legal grounds to unaccountability. Even if such systems were transparent, the difficulty of attaching any given output to the variables that contributed to it relinquishes huge power to the algorithms used by an increasing number of governments, services, and companies.

Northpointe, who created COMPAS, justify their algorithm's protection from public scrutiny as the need for trade secrets, suggesting that disclosing all the attributes which are used as inputs to the algorithm would afford them a competitive disadvantage. Algorithmic transparency would require Northpointe to reveal the processes that data is passed through to reach the concluding risk score as

well as the statistical models used, yet the COMPAS system contains complexities that would be difficult to explain in such a manner. This secrecy, sustained by the intricacy of machine learning used to determine weights and the reticence of the companies that use it, brings into question the way life-affecting decisions are made without public accountability and understanding, an ethical scenario that is currently given little to no contemplation by the data science industry.

Despite bearing no relevance on the accuracy or prejudice of the data produced by an algorithm, the public perceptions of data science are equally influential on its impact. Data is considered to be objective by its very nature and algorithms have only been allowed to become a powerful player in modern society due to the naivety of the public as to how they operate. The reality of algorithmic bias as a hidden form of discrimination combined with the lack of public knowledge as to how it operates, including the problems it creates and how widespread use currently is, creates a misleading reputation for the data science industry that keeps ethical queries far away from the forefront of research and innovation.

While the ProPublica investigations have revealed several ethical problems attached to the COMPAS system, it is still accompanied by a reputation for neutrality, to the extent that judges' perceptions of the efficacy and fairness of system override the caveats put in place to safeguard against algorithmic bias. The COMPAS system was introduced with the intent that the risk scores it generated would only inform the overall conclusions of judges, not replace them. The allure to judges, however, of relegating responsibility for a mistaken judgement to a system that cannot be faulted or criticised for it is tempting. The alternative is ignoring the risk score and potentially making a mistake, which causes those affected to question why the COMPAS prediction was not considered, a criticism that could not be levied against judges before the introduction of the system. As a result, the safeguard put in place

by the courts that use COMPAS and other similar systems is removed by its own inaccurate reputation.

Algorithmic bias will remain an ethical issue of the upmost importance within data science and, by consequence of the industry's far-reaching impact, society as a whole until such time as companies are held accountable for the impacts of the algorithms they create and use. Individual reforms, such as Google's AI ethics principles, are beneficial to a point but trust cannot be placed in companies to self-regulate as a solution, as evidenced by the recent dismissal of Google AI Ethicist Timnit Gebru who attempted to publish a paper criticising the company's diversity efforts within the field. Instead, industry-wide standards, ethical oversight as well as better education of the wider population as to the limits and risks of algorithmic decision-making are necessary if data science is to play a beneficial role in the future.

Bibliography

Benjamin, R., (2019). *Race After Technology*. Polity Press.

Hofer, M., 2018. *Benefits And Ethical Challenges In Data Science — COMPAS And Smart Meters*. Medium. Available at: https://towardsdatascience.com/benefits-and-ethical-challenges-in-data-science-compas-and-smart-meters-da549dacd7cd (Accessed 27 January 2021).

Kobielus, J., (2013). *The Challenges Of Transparent Accountability In Big Data Analytics | IBM Big Data & Analytics Hub*. Ibmbigdatahub.com. Available at: https://www.ibmbigdatahub.com/blog/challenges-transparent-accountability-big-data-analytics (Accessed 27 January 2021).

Kosinski, M. and Wang, Y., (2018). *Deep Neural Networks Are More Accurate Than Humans At Detecting Sexual Orientation From Facial Images*. Stanford Graduate School of Business. Available

at: https://www.gsb.stanford.edu/faculty-research/publications/deep-neural-networks-are-more-accurate-humans-detecting-sexual (Accessed 27 January 2021).

Larson, J., Mattu, S., Kirchner, L. and Angwin, J., (2016). *How We Analyzed the COMPAS Recidivism Algorithm*. ProPublica. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (Accessed 27 January 2021).

Leetaru, K., (2017). *AI 'Gaydar' And How The Future Of AI Will Be Exempt From Ethical Review*. Forbes. Available at: https://www.forbes.com/sites/kalevleetaru/2017/09/16/ai-gaydar-and-how-the-future-of-ai-will-be-exempt-from-ethical-review/?sh=33fe8d4f2c09 (Accessed 27 January 2021).

Northpointe Inc. (2015). *Practitioner's Guide to COMPAS Core*. Available at: https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf (Accessed 27 January 2021).

Singh, M., (2020). *Google Workers Demand Reinstatement And Apology For Fired Black AI Ethics Researcher*. the Guardian. Available at: https://www.theguardian.com/technology/2020/dec/16/google-timnit-gebru-fired-letter-reinstated-diversity (Accessed 27 January 2021).

Taylor, P., (2021). *Paul Taylor · Insanely Complicated, Hopelessly Inadequate · LRB 21 January 2021*. London Review of Books. Available at: https://www.lrb.co.uk/the-paper/v43/n02/paul-taylor/insanely-complicated-hopelessly-inadequate (Accessed 27 January 2021).