<u>What is the Importance of Ethics in Data Science?</u>

Analysing the plethora of data we consume daily is an instinctive part of human nature. Recognising and understanding patterns in our world allows us to make informed responses from a survival perspective down to the opinions, preferences, habits and relationships we form within our lives. These behaviours were intensified with the digital age, and with increased technologies and software, they became imprinted onto a global pile of active information where we continue to dictate our actions – or at least, we think we do. As the data pile grows exponentially, with 2013 research declaring an estimated 90% of it having only been collected within the two years prior,[1] it became apparent to those at the forefront of technological progress that sorting through it and learning from it would require a field of its own, and data science emerged [2] along with the term 'big data' which refers to the systemisation of masses of data that are large in volume and variety,[3] obtained by companies to be sifted and analysed for information. [4] By observing patterns within big data from the way that data is used and transferred, data scientists are able to analyse the very core of our technological behaviours, enabling continuous innovation and empowering much of the artificial intelligence which we have grown accustomed to within our everyday lives. With this level of intimate access to the way in which we knowingly or unknowingly share our data – which can be as directly as searching for a specific product on sites like Google or as implicit as liking a single post

---

[1] SINTEF, written by Åse Dragland. "Big Data, for better or worse: 90% of world's data generated over last two years." ScienceDaily, 22nd May 2013. https://www.sciencedaily.com/releases/2013/05/130522085217.htm
[2] Berkeley School of Information, "[Data Science] What is Data Science?" https://ischoolonline.berkeley.edu/data-science/what-is-data-science/#fn2b
[3] Analytics Software and Solutions, "Big Data | What is it and why it matters" https://www.sas.com/en_gb/insights/big-data/what-is-big-data.html accessed January 2021.
[4] Alyssa Schroer, "34 Big Data Companies Helping Us Make Sense of the World," 7th January 2019, updated 6th

August 2020. https://builtin.com/big-data/big-data-companies-roundup

Rabiah Ahmad

on a social media feed – the issues of security and privacy arise. As the range of technology that we use and the information that corporations learn simultaneously increase, with our data being harvested for access and distribution to whichever degree service providers decide upon, we might question the significance of ethics within data science.

Access to this arguably sensitive information allows data scientists to pick up and formulate impressions of us and moral opinions of right and wrong can quickly diverge, making the topic of ethics blurred to some extent.[5] Are they simply analysing anonymised patterns of usage to optimise the profitability of a certain product or expand a particular market, or are our identities less hidden than we think? Anonymised data, which could be personal but 'scrubbed' of its unique identifiers, is largely unregulated and can be sold on to third parties, including marketing companies or data brokers (whose chief service is to purchase and resell data to customers, such as businesses looking to target advertisements to potential customers or money lenders looking to detect fraud and check the credibility of loan applicants).[6] On one hand, there seems to be little ethical input required for this transactional use of data since it is generally used as a marketing tool in line with capitalist structures, but when the use of data leads to overly specific targeting of people, ethics become a consideration. Despite the protective legislations against the misuse of personal data, this industry communality means individually depersonalised identifiers can easily be cross-referenced and reidentified, [7] which seems to render the ethical challenge of misusing the capabilities of data science insignificant amongst the grander visions of market successes.

[5] Michael Smolski, "Is Data Collection an Invasion of Privacy (From the Eyes of a Privacy Company)," 26th August 2019. https://privatevpn.com/blog/business/345/is-data-collection-an-invasion-of-privacy-(from-the-eyes-of-a-privacy-company)

[6] Dan Rafter, "How data brokers find and sell your personal info," revised 18th January 2021. https://us.norton.com/internetsecurity-privacy-how-data-brokers-find-and-sell-your-personal-info.html

[7] Boris Lubarsky, 1 GEO. L. TECH. REV. 202, "Re-Identification of "Anonymized" Data," April 2017. https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/

When our data finds its way to those who understand its scientific and economic value, it can be argued that ethical conduct is its primary protector within the realms of data science. Ethics and principles could differentiate a company employee from a malicious insider - one of the most common causes of data breaches [8] - but alone will not always prevent the mishandling of data, whether it is accessed internally or stolen otherwise. It can be argued that whilst ethics are as fundamental within data science as they are in many facets of life, there are other more dependable factors involved in the area, often dictated by the companies profiting off big data. Improving data defences with knowledgeable security staff and well-maintained hardware can protect the information about the subjects unknowingly at the core of this research, but choosing to incur the costs of strengthened cyber security and better understanding of data science is a decision that sometimes has to be, to a degree, based on ethics. However, security breaches are not cheap - 2020 statistics show that the average cost of a data breach is $3.86 million.[9] With this being the case, it could be interpreted that the companies most likely to reject or avoid the costs of data protection are those which reap the biggest profit margins on their services or products, [10] and consequently also the ones likely to mine the most data whilst dismissing ethics; examples of recent significant data breaches include Twitter having 130 high-profile accounts targeted, a security breach of the Marriott-Starwood hotel chain impacting over 5.2 million hotel guests, and Uber being hacked in 2016 where information was stolen about 57 million users and drivers. [11] It took Marriot-Starwood

[8] Luke Irwin, "The most common causes of data breaches and how you can spot them," 8th April 2020. https://www.itgovernance.eu/blog/en/the-most-common-causes-of-data-breaches-and-how-you-can-spot-them

[9] Rob Sobers, "134 Cybersecurity Statistics and Trends for 2021" https://www.varonis.com/blog/cybersecurity-statistics/

[10] Department for Digital Culture, Media and Sport, Ipsos MORI Social Research Institute, University of Portsmouth, and Rishi Vaidya, "Cyber Security Breaches Survey" 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950063/Cyber_Security_Breaches_Survey_2019_-_Main_Report_-_revised_V2.pdf

[11] Josh Fruhlinger, "Marriott data breach FAQ: How did it happen and what was the impact?" 12th February 2020 https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html

Rabiah Ahmad

four years from the initial breach of over five hundred million guest records, which included sensitive information like passport numbers and financial details, to identify that the issue had occurred and subsequently release this finding publicly.[12] The Starwood brand's reservation systems were flawed and notoriously difficult to secure, [13] and after the initial attack in 2014 they were infiltrated again in 2015 in a separate attack that took eight months to be noticed.[14] Following their acquisition of Starwood in 2016, Marriott laid off many of their staff, with no exemption for those in security and information technology,[15] even though they went on to rely on the already problem-ridden Starwood systems to accommodate thousands of new customers: the payroll cuts appear to have been made in the interest of maximising their increased profitability without prioritising functionality, further evidenced by the continued neglect of the struggling Starwood systems holding their data until their hands were forced by the realisation of the breach in 2018.[16] These accumulative events could be reason to believe that the situation, risking many trusting customers, may have been avoided if ethics had more importance in their data science management. Overall, company mismanagement in the protection of the data which they rely on data science to successfully exploit seems to return to ethics: it could be argued that the decisions taken to forego extra security measures to protect consumer details - typically in preference of reducing costs or holding on to business deals to maintain and maximise profits whilst gambling the big data collected - is inherently unethical, and if ethical decision-making was given more importance in the use of data

---

[12] Josh Fruhlinger, "Marriott data breach FAQ: How did it happen and what was the impact?" 12th February 2020 https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html

[13] Patrick Nohe, "The Marriott Data Breach: Full Autopsy – 'Autopsying the Marriott Data Breach: This is why insurane matters," 22nd March 2019. https://www.thesslstore.com/blog/autopsying-the-marriott-data-breach-this-is-why-insurance-matters/

[14] Josh Fruhlinger, "Marriott data breach FAQ: How did it happen and what was the impact?" 12th February 2020 https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html

[15] Patrick Nohe, "The Marriott Data Breach: Full Autopsy – 'Autopsying the Marriott Data Breach: This is why insurane matters," 22nd March 2019. https://www.thesslstore.com/blog/autopsying-the-marriott-data-breach-this-is-why-insurance-matters/

[16] Josh Fruhlinger, "Marriott data breach FAQ: How did it happen and what was the impact?" 12th February 2020 https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html

science and implemented in the defence of consumer privacy then fewer compromises might be made, crowning ethics an influential determinant in responsible data science control.

As well as being essential in the regulation of data science, ethics are also important in the way it is applied. When sensitive data is collected, there is the inevitable risk of misuse, but when harnessed by ethically guided data scientists to identify systematic prejudices and injustices and then question the underlying causes of these, tangible progress can be made to dissolve larger societal problems that are else reflected in the algorithms that go on to be used across these same defective societies. [17] Digital algorithms used by many service providers are not sentient, but it cannot be assumed that they operate with immunity to human bias. They will process the inputs they receive in line with the way they have been programmed by the data scientists and the research that came before them, and if this disregards situational factors and enforces a systemic bias or discrimination [18] to the service provider, it can easily be accepted without ethical consideration. Ethical consideration is brought in by the human element of data science, where the faults of the models are picked up and acknowledged, and then the scientists and organisations behind the machines address the ways in which they can be targeted progressively to break down institutionalised barriers. Data scientists have shown the importance of ethics in innovative practice by responding to their ethical questioning of algorithmic bias with the discovery of a method to determine whether supposedly neutral functions are instead fulfilling the flaws of the system in which they operate [19] - an algorithmic way of contesting algorithmic issues. Ethics within data science encourage an increased

[17] Maximilian Hofer, "Benefits and Ethical Challenges in Data Science – COMPAS and Smart Meters," 10th May 2018. https://towardsdatascience.com/benefits-and-ethical-challenges-in-data-science-compas-and-smart-meters-da549dacd7cd
[18] Reuben Binns, The Conversation, "It's not big data that discriminates – it's the people that use it," 22nd March 2016. https://phys.org/news/2016-03-big-discriminates-people.html
[19] University of Utah, "Programming and prejudice: Computer scientists discover how to find bias in algorithms," 14th August 2015. https://phys.org/news/2015-08-prejudice-scientists-bias-algorithms.html

pace of progression that might be a key factor in solving some of the greatest unresolved or persisting challenges faced both technologically and socially.[20] [21]

Data science is an expansive subject, as is the importance of ethics within it. Nonetheless, even this somewhat small insight can exemplify that to optimise the potential of data science, it takes more than exploiting surface-level patterns found in large datasets and tuned algorithms; it asks for the behaviour of real people within data science, driven by ethics, to maximise the opportunities of the rapidly growing area of research, such as by resolving problems in artificial intelligence or by considerate application in widely beneficial medical research.[22] [23] Ethical implications have likely contributed to changes that mitigate the negative consequences of the growth of data science, such as the EU's GDPR regulations [24] which coerced many companies [25] [26] into considering their cyber security, but legislations are not necessarily indicative of ethics being regarded highly – these companies might not otherwise give importance to the role of ethics in their use of data science. Regardless, even if they are not yet being acted on communally, ethics appear conclusively to possess great importance in the future of data science: they are paramount to its continual successes and growth.

[20] Paul Tallon, "Why Ethics are Important in Data Science," Loyola University Maryland, 2018
https://www.loyola.edu/academics/data-science/blog/2018/why-ethics-are-important-in-data-science
[21] Jeane W. Ross, Cynthia M. Beath, Anne Quaadgras, "You May Not Need Big Data After All,"
https://hbr.org/2013/12/you-may-not-need-big-data-after-all
[22] IQVIA, "Insights | Human Data Science" https://www.iqvia.com/insights/human-data-science accessed January 2021
[23] Yilian Yuan, "A Day in the Life of a Human Data Scientist," 4th February 2019.
https://www.iqvia.com/blogs/2019/02/a-day-in-the-life-of-a-human-data-scientist-not-your-standard-statistics-and-analytics
[24] Thomas W. Dinsmore, "How GDPR affects data science," https://www.kdnuggets.com/2017/07/gdpr-affects-data-science.html
[25] Department for Digital Culture, Media and Sport, Ipsos MORI Social Research Institute, University of Portsmouth, and Rishi Vaidya, "Cyber Security Breaches Survey" 2019.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950063/Cyber_Security_Breaches_Survey_2019_-_Main_Report_-_revised_V2.pdf
[26] Izaak Crook, "How GDPR Will Affect Data Science," 13th April 2018. https://dataconomy.com/2018/04/how-gdpr-will-affect-data-science/

Rabiah Ahmad

Bibliography

https://builtin.com/data-science

https://www.sciencedaily.com/releases/2013/05/130522085217.htm

https://ischoolonline.berkeley.edu/data-science/what-is-data-science/#fn2

https://builtin.com/big-data/big-data-in-healthcare

https://builtin.com/big-data/big-data-companies-roundup

https://www.nytimes.com/2018/10/21/opinion/who-will-teach-silicon-valley-to-be-ethical.html

https://towardsdatascience.com/ai-and-the-future-of-privacy-3d5f6552a7c4

https://www.youtube.com/watch?v=bAyrObl7TYE

https://www.youtube.com/watch?v=UQeyU0YcPKY

https://www.itgovernance.eu/blog/en/the-most-common-causes-of-data-breaches-and-how-you-can-spot-them

https://searchdatamanagement.techtarget.com/definition/big-data

https://searchsecurity.techtarget.com/definition/data-breach?_ga=2.83910425.1453997096.1612001725-1328890511.1612001725

https://towardsdatascience.com/ai-and-the-future-of-privacy-3d5f6552a7c4

https://privatevpn.com/blog/business/345/is-data-collection-an-invasion-of-privacy-(from-the-eyes-of-a-privacy-company)

https://us.norton.com/internetsecurity-privacy-how-data-brokers-find-and-sell-your-personal-info.html

https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/

https://venturi-group.com/is-big-data-a-good-thing-or-an-invasion-of-privacy/ew.org/re-identification-of-anonymized-data/GLTR-04-2017/

https://venturi-group.com/is-big-data-a-good-thing-or-an-invasion-of-privacy/

https://www.theguardian.com/commentisfree/2013/mar/12/big-data-greater-good-privacy-invasion

https://www.researchgate.net/publication/337490438_Big_data_and_privacy_The_study_of_privacy_invasion_acceptance_in_the_world_of_big_data

https://www.cnet.com/news/in-the-world-of-big-data-privacy-invasion-is-the-business-model/

https://www.cnet.com/news/techs-invasion-of-our-privacy-made-us-more-paranoid-in-2018/

Rabiah Ahmad

https://theconversation.com/heres-how-tech-giants-profit-from-invading-our-privacy-and-how-we-can-start-taking-it-back-120078

https://towardsdatascience.com/algorithms-born-of-our-prejudices-8be47ca6103c

https://towardsdatascience.com/deep-biases-fixing-ais-unintentional-prejudices-234893a40a3b

https://phys.org/news/2016-03-big-discriminates-people.html

https://phys.org/news/2015-08-prejudice-scientists-bias-algorithms.html

https://towardsdatascience.com/benefits-and-ethical-challenges-in-data-science-compas-and-smart-meters-da549dacd7cd

https://www.loyola.edu/academics/data-science/blog/2018/why-ethics-are-important-in-data-science

https://towardsdatascience.com/5-principles-for-big-data-ethics-b5df1d105cd3

https://www.sas.com/en_gb/insights/big-data/what-is-big-data.html

https://www.oracle.com/uk/big-data/what-isx-big-data/

https://hbr.org/2013/12/you-may-not-need-big-data-after-all

https://dataconomy.com/2018/04/how-gdpr-will-affect-data-science/

https://blogs.thomsonreuters.com/answerson/data-science-legal-data-privacy/

https://www.kdnuggets.com/2017/07/gdpr-affects-data-science.html

https://www.nature.com/articles/d41586-019-01506-x

https://www.immuta.com/articles/data-privacy-data-science-the-next-generation-of-data-experimentation/

https://www.iqvia.com/insights/human-data-science

https://www.iqvia.com/insights/human-data-science

https://digital.nhs.uk/data-and-information/data-insights-and-statistics/innovative-uses-of-data-and-data-science

https://www.sciencedaily.com/releases/2020/07/200709141614.htm

https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/data-breach-101

https://www.itgovernance.eu/blog/en/the-most-common-causes-of-data-breaches-and-how-you-can-spot-them

https://www.varonis.com/blog/cybersecurity-statistics/

https://info.varonis.com/hubfs/docs/research_reports/2021-Financial-Data-Risk-Report.pdf?utm_content=146358482&utm_medium=social&utm_source=twitter&hss_channel=tw-21672993

Rabiah Ahmad

https://www.varonis.com/blog/cybersecurity-skills-shortage/

https://www.cnbc.com/2020/07/16/twitter-hackers-made-121000-in-bitcoin-analysis-shows.html

https://www.cnbc.com/2020/07/16/twitter-hackers-made-121000-in-bitcoin-analysis-shows.html

https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html

https://www.thesslstore.com/blog/autopsying-the-marriott-data-breach-this-is-why-insurance-matters/

https://www.csoonline.com/article/3153707/top-cybersecurity-facts-figures-and-statistics.html

https://www.helpnetsecurity.com/2019/10/01/fortune-500-ciso/

https://www.technologyreview.com/2020/02/24/276013/cybersecurity-in-2020-the-rise-of-the-ciso/

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950063/Cyber_Security_Breaches_Survey_2019_-_Main_Report_-_revised_V2.pdf

Rabiah Ahmad