

Q. How can Data Science and Humanities inform each other?

Data is like a jigsaw puzzle piece. A puzzle piece alone is meaningless, but it becomes valuable if assembled as part of a complete puzzle set. Similarly, data about someone travelling out of town will not sound interesting, even tend to be meaningless. However, the data can be valuable and interesting on the condition that it is processed into a piece of important information. Specifically, for example, information highlighting whether the increase in the number of people travelling out of town during a pandemic is proportional to the increased number of people infected in the destination city. Nevertheless, data science methods are required to process the raw data materials into this information. By definition, data science itself is a science that aims to extract knowledge or information from a set of data. The processed data results can be interpreted as a setup for artificial intelligence algorithms that can predict what we want to watch, buy, and read on social media platforms or uncover human problems. However, while current data science tools can sift through piles of data for patterns at a level of scale and speed that humans cannot achieve on their own, limitations remain evident in our existing models for understanding and determining human problems. One notable example is when Jigsaw, a technology incubator of Google that deals with online misinformation, stumbled upon a significant problem that even their powerful computer algorithms cannot solve: the complexities and contradictions of human thought. After having difficulty finding ways to convince anti-vaxxers about their misinformation, they turned to anthropologists to help them understand their behaviour. Prior, the engineers' assumed that a "debunking" site should look professional to demonstrate its credibility. However, anthropologists later discovered that professionally designed sites do not appeal to

conspiracy theories, precisely because they look like the global elites created them¹. Jigsaw, in particular, also used a combination of human oversight and data algorithms to prevent teens from being recruited by ISIS.²

Based on the previous paragraph, we then can conclude that mathematical algorithms alone will not always be successful in understanding human behaviour and solving human problems. It could only analyse and make assumptions based on specific regions, characteristics, or past occurrences based on human behaviour patterns. For example, a study based on data revealed that individuals who grew up in regions with relatively clement temperatures had higher sociability and personal growth levels than individuals who grew up in regions with warmer temperatures.³ However, we cannot necessarily conclude that people living in the hotter region are necessarily more temperamental than those living in the colder region. Therefore, this is where humanities insight is required to provide new perspectives, to help people reach conclusions beyond just the processed data result because data science exists to lead people to answers, not to provide them directly.

A good data science analysis actually requires not only technical skills, but also humanist skills. To explain this reason, we must first go back to the main purpose of data processing. Since the beginning, data has been stored and processed from and for humans, so the results of data generated and presented by engineers should be beneficial and perceptible by humans as the end-users themselves. Therefore, this process must involve human understanding

¹ Tett, G. 2021. *The human factor — why data is not enough to understand the world*. Financial Times.

² Dishman, L., & editor, s. 2019. *How Jigsaw is using AI, human connections, and AdWords to fight ISIS*. Fast Company.

³ Chen, H., Lai, K., He, L., & Yu, R. 2020. *Where You Are Is Who You Are? The Geographical Account of Psychological Phenomena*. Frontiers in Psychology.

qualitatively, not quantitatively, which is more of a humanities field of study than a scientific one. Even autonomous car, for example, needs to understand the complexities of human behaviour on the highway –such as assessing indications from a driver at a four-way stop to get either person to go first– and it requires the role of anthropologies to understand and socially assess the possibilities and meanings of patterns used by human to anticipate actions in the highway⁴. According to Carson Foster, a humanities graduate turned data scientist, humanities scholars actually have the advantage of skills that are more valuable than knowledge of any quantitative methodology in particular, which includes being able to learn deeply on any topic self-taught, stating a research question with evidences to support their answer, and be able to communicate the limitations and assumptions of their approach⁵.

Although the combination of data science and humanities may be odd for some people, they have a quite considerable correlation. Since the life of our ancestors, data has shaped and recorded the development of humankind, although not in the medium we know today. The term “data” was first used in English in the 1640s, a term derived from Latin, meant “something given” – often used as the basis for calculations. And in 1962, John Graunt studied death records data kept by London parishes to make observations about the varying death rates between genders, and even predict human life expectancy⁶. However, technically humans have been storing and processing data into information for tens of thousands of years since prehistoric times. At that time, humans carved data about their lives on rocks and walls and ceilings of caves, behaving as an ancient qualitative database. Then humans began to understand how to process the numerical data at their disposal as the Palaeolithic tribes did

⁴ *The human side of autonomous cars — with Nissan Research's Melissa Cefkin*. Medium.

⁵ Forter, C. *Humanities Graduates Should Consider Data Science*. Towards Data Science.

⁶ *The history of data*. ThinkAutomation.

around 19,000 BC, where they mark notches into sticks or bones to keep track of trading activity or supplies, and then compare them to make basic calculations about how long their food supplies would last⁷. Tally sticks were later developed in medieval England to keep loans or collect tax records by the local sheriffs⁸. Then this manual data storage method developed into digital cloud storage, while various analytical tools emerged that became the forerunner to the birth of data science. In short, human life has always been related to data and how to process it.

Whilst the correlation between data science and humanities, it's inevitable that data science needs humanities. Algorithms are simply mindless machines that can be biased based on who builds, develops, and uses them. If an error in the algorithm occurs, either intentionally or unintentionally, it can cause harm to someone. Notable examples include Amazon hiring algorithm that resulted in discrimination against women applicants because the data set is male-inclined⁹ and criminal sentencing algorithms that impose harsher penalties on defendants based on their race and ethnic background¹⁰.

Whilst the correlation between data science and humanities, it's inevitable that data science needs humanities. Algorithms are simply mindless machines that can be biased based on who builds, develops, and uses them. If an error in the algorithm occurs, either intentionally or unintentionally, it can cause harm to someone. Notable examples include the Amazon hiring algorithm that used data from a pool of primarily male applicants, which resulted in

⁷ Read <https://www.thinkautomation.com/histories/the-history-of-data/> and <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/>

⁸ Medieval London.

⁹ Adams, A., Hefter, K., Stephen, K., & Khodor, O. 2021. *Algorithmic Decisions and Their Human Consequences*. The Regulatory Review.

¹⁰ Angwin, J., Larson, J., Mattu, S., & Kirchner, L. 2016. *Machine Bias — ProPublica*. ProPublica.

discrimination against women applicants that results in discrimination against women applicants because the data set is male-inclined¹¹ and criminal sentencing algorithms that impose harsher penalties on defendants based on their race and ethnic background¹². In addition, Sam Dragga and Dan Voss have argued that data visualisation requires humanism as an essential element. They used a visualisation of Napoleon's 1812 invasion of Russia, which only highlighted his victories but did not emphasise the human misery of 412,000 casualties that resulted from the military campaign, and called this omission of data a “distortion of the reality”, and also criticising the insensitive graphics to the human condition it depicts¹³. Humanities are needed to correct these rhetorical, ethical, and communication errors in statistical graphics and produce more humane technical illustrations, since data visualisation is essential in data science to help properly curate data into a form more accessible to understand¹⁴, and these are humanities advantages over scientific majors.

In short, this is a challenge for humanities scholars to train themselves more in the core subjects of modern data, since technology has become ingrained in human life and could make an enormous contribution to humanities majors in particular. For example, digitising data collections with data science makes it easier for humanities scholars to construct their research. Data science also helped the Bridgman team in the 2019 Canadian election to find the characteristics of the public in political views. They found that people are more likely to share information from sites that align with their political views, or people are more likely to read articles about the success of the party they support, and that what politicians deem

¹¹ Adams, A., Hefter, K., Stephen, K., & Khodor, O. 2021. *Algorithmic Decisions and Their Human Consequences*. The Regulatory Review.

¹² Angwin, J., Larson, J., Mattu, S., & Kirchner, L. 2016. *Machine Bias — ProPublica*. ProPublica.

¹³ Dragga, Sam & Voss, Dan. 2001. *Cruel Pies: The Inhumanity of Technical Illustrations*. Technical Communication.

¹⁴ *What Is Data Visualization? Definition & Examples | Tableau*. (n.d.). Tableau Software

important is not always in line with the views of the Canadian public¹⁵. Another contribution of data science to humanities also includes natural language processing such as Google Translate, detecting counterfeit artwork¹⁶, learning what public policies have worked best to solve economic inequality with data visualisation, and helping archaeologists collect historical information and satellite imagery documenting the forced mass displacement of the Spanish Empire from the Inca Empire in the 16th century in online databases¹⁷.

Subsequently, the table below outlined the previous discussion's key points, explaining succinctly about how data science and humanities could inform each other with real-life examples being provided.

	Data science and humanities relevance	Real-life examples
1	Anthropology and data science	Analysing anti-vaxxers behaviour in cyberspace and creating algorithm to prevent teens being recruited by ISIS (pg. 1-2), teaching autonomous cars human behaviours (pg. 3)
2	Archaeology and data science	Documenting the forced mass displacement of the Spanish Empire from the Inca Empire in the 16th century in online databases (pg. 6)
3	Arts and data science	Creating more humane data visualisation (pg. 5) & detecting counterfeit artwork (pg. 6)
4	Languages and data science	Natural language processing (pg. 6)
5	Politic and data science	Finding characteristics of the public in political views (pg. 5-6)
6	Law and data science	Researching best public policies (pg. 6)

¹⁵ Gorbounov, S. 2020. *Making room for data science in the humanities*. The McGill Tribune.

¹⁶ Huang, C. 2018. *Art connoisseurship in the age of machine learning*. Towards Data Science.

¹⁷ Kimball, J. 2020. *How archaeologists can dig deeper with big data*. Futurity.

To conclude, when technology today allows us to analyse problems that we would not be able to do on our own, it still has its own humanities limitations that remains unchanged along the changing nature of human behaviour. This world consists of many jigsaw puzzle pieces (data), and data science is how each piece is put together into a perfect puzzle set. Meanwhile, humanities are needed to ensure that each step taken does not harm anyone in its journey, and makes data science a goal for the greater good. []

*Word Count: 1582**

**Body text only (excluding titles, references, tables, pictures, and profile)*

Sayyidah Fatimah Azzahra
Eq. to 11th grade, Homeschooling, Indonesia

Bibliography

1. Tett, G. 2021. *The human factor — why data is not enough to understand the world*. Financial Times. <https://www.ft.com/content/4f00469c-75da-4e29-baf3-b7bec470732c>
2. Dishman, L., & editor, s. 2019. *How Jigsaw is using AI, human connections, and AdWords to fight ISIS*. Fast Company. <https://www.fastcompany.com/90294876/how-jigsaw-is-using-ai-human-connections-and-adwords-to-fight-isis>
3. Chen, H., Lai, K., He, L., & Yu, R. 2020. Where You Are Is Who You Are? The Geographical Account of Psychological Phenomena. *Frontiers in Psychology*, 11. 10.3389/fpsyg.2020.00536
4. *The human side of autonomous cars — with Nissan Research's Melissa Cefkin*. Medium. <https://medium.com/@experientia/social-driving-the-human-side-of-autonomous-cars-with-nissan-reseachs-melissa-cefkin-567746c333d0>
5. Forter, C. (n.d.). *Humanities Graduates Should Consider Data Science | by Carson Forter*. Towards Data Science. <https://towardsdatascience.com/humanities-graduates-should-consider-data-science-d9fc78735b0c>
6. *The history of data*. ThinkAutomation. <https://www.thinkautomation.com/histories/the-history-of-data/>
7. The World Economic Forum. 2015. *A brief history of big data everyone should read*. <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/>
8. Fordham University. (n.d.). *Tally Stick · Medieval London Objects 3 · Medieval London*. Medieval London. <https://medievallondon.ace.fordham.edu/exhibits/show/medieval-london-objects-3/tally-stick>
9. Adams, A., Hefter, K., Stephen, K., & Khodor, O. 2021. *Algorithmic Decisions and Their Human Consequences*. The Regulatory Review. <https://www.theregreview.org/2021/11/11/adams-algorithmic-decisions-human-consequences/>
10. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. 2016. *Machine Bias — ProPublica*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
11. Dragga, Sam & Voss, Dan. 2001. Cruel Pies: The Inhumanity of Technical Illustrations. *Technical Communication*. 48. 265-274.
12. *What Is Data Visualization? Definition & Examples | Tableau*. (n.d.). Tableau Software. <https://www.tableau.com/learn/articles/data-visualization>
13. Gorbounov, S. 2020. *Making room for data science in the humanities*. The McGill Tribune. <https://www.mcgilltribune.com/sci-tech/making-room-for-data-science-in-the-humanities-280120/>
14. Huang, C. 2018. *Art connoisseurship in the age of machine learning*. Towards Data Science. <https://towardsdatascience.com/art-connoisseurship-in-the-age-of-machine-learning-d72e8a3716df>
15. Kimball, J. 2020. *How archaeologists can dig deeper with big data*. Futurity. <https://www.futurity.org/big-data-archaeology-inca-empire-2290552/>