Collaborative Discussion 1

Initial Post:

Data Science can be defined as a 'set of fundamental principles that support and guide the principled extraction of information and knowledge from data' (Provost and Fawcett 2013). A Data Scientist uses a combination of skills from Computer Science, Statistics and Maths, to solve a variety of problems. One of the main problems that the field of Data Science looks to solve is finding patterns in massive data sets (Singh 2013). The ever-growing usage of computers and smartphones has led to more and more data being created and demanded. The internet of things (IoT) has been a catalyst to this situation. The amount of data generated in 2010 was approximately 2 zettabytes, and that figure grew to 26 zettabytes in 2017 (Durate 2023, as cited in Taylor 2022).

The drastic change in the volume of data has led to the 4th Paradigm of Science. This refers to the scientific view on how Big Data changes the very fabric of science (Hitzler and Janowicsz 2013) Hitzlet and Janowicz (2013) mention that the 'availability of data from different times, locations, perspectives, topics, cultures, resolutions, qualities, and so forth', [has led to] 'exploration' of this additional paradigm. As already mentioned earlier by Singh (2013), finding patterns in massive data sets is a benefit that Data Science provides. Kulkarni (2019) states that '95% of businesses cite the need to manage unstructured data as a problem for their business'. This has meant that a data scientist has become the fastest growing job on LinkedIn, and an estimated 11.5million data science jobs will be created by 2026 (Gour 2019). The skills of a data scientist can be used to explain findings in a story stelling format, creating new solutions and new problems simultaneously.

However, access to more data means greater responsibility for people in the field of Data Science. 'Security and privacy issues are key concerns when it comes to big data' (Hillier 2022). Exploration into the 4th paradigm and machine learning needs care and attention. The 'need to understand machine learning algorithms to build a better-automated system' will 'lead to more jobs' (Saxena 2021). Rather than less jobs, and higher unemployment.

To conclude, Data Science will have a huge influence on society and the economy of the future. As we enter the 4th Paradigm of Science and more data being generated, there is greater responsibility placed on Data Scientists when it comes to legal, ethical, and social issues. With there being many different aspects in Data Science. It is often overwhelming and referred to as the 'Dark Side' (Yildirim 2020) of Data Science.

References

Provost, F. and Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), pp.51-59.

Hitzler, P. and Janowicz, K., 2013. Linked Data, Big Data, and the 4th Paradigm. Semantic Web, 4(3), pp.233-235.

Ritesh Pratap Arjun Singh., 2013. Common Data Science Challenges of 2023, Available at: **https://www.knowledgehut.com/blog/data-science/data-science-challenges** [Accessed 13th May 2023].

Kulkarni, R 2019., 'Big Data Goes Big'. Forbes. 7th February, Available from: **https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/?sh=4bd0a97220d7** [Accessed on: 13th May 2023].

Rinu Gour 2019., Why Choose Data Science for Your Career, Available at: **https://towardsdatascience.com/why-choose-data-science-for-your-career-ca38db0c28d4** [Accessed on: 13th May 2023].

Pranjal Saxena 2019., There Will Be a Shortage of Data Science Jobs in the Next 5 Years?, Available from: **https://towardsdatascience.com/there-will-be-a-shortage-of-data-science-jobs-in-the-next-5-years-9f783737ed23** [Accessed on: 13th May 2023]

Soner Yildirim 2020., The Dark Side of the Sexiest Job of the 21st Century. Available from: **https://towardsdatascience.com/the-dark-side-of-the-sexiest-job-of-the-21st-century-fd9c46bf4cae** [Accessed on 14th May 2023].

Petroc Taylor., 2022. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025, Available at: **https://www.statista.com/statistics/871513/worldwide-data-created/** [Accessed on: 13th May].

Will Hillier 2022., Is Big Data Dangerous?, Available from: **https://careerfoundry.com/en/blog/data analytics/is-big-data-dangerous/#:~:text=Big%20data%20comes%20with%20security,scams%2C%20and%20to%20spread%20disinformation** [Accessed on: 13th May 2023].

Fabio Duarte., 2023. Amount of Data Created Daily, Available at: **https://explodingtopics.com/blog/data-generated-per-day** [Accessed 13th May].

Peer Responses

Re: Initial Post

by Nima Osman - Wednesday, 17 May 2023, 10:59 AM

Hey Chris,

I enjoyed reading your point of view on this topic. I think you stated a lot of interesting arguments.

You mentioned the 4th paradigm in your post quite often which I have not mentioned myself. I found it very insightful reading about this and how it plays a role in the potential development of a data scientist's future role.

Saxena (2021) also mentioned that the role of data scientists could evolve to automate the pipeline and let the tools make decisions on algorithm selection.

Additionally, the onset of the 4th paradigm will create a massive shift in the way we handle data.

Harvard business review suggested that there might be a rise in the need for specialised data scientists due to the differentiation of roles of data scientists to handle the surge in data (Davenport & Patel, 2022).
How do you think this could potentially affect the day-to-day role of a data scientist?

References:

Davenport, T.H. and Patel, D.J. (2022) Is data scientist still the sexiest job of the 21st century?, Harvard Business Review. Available at: https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century (Accessed: 16 May 2023).

Saxena, P. (2021) There will be a shortage of data science jobs in the next 5 years?, Medium. Available at: https://towardsdatascience.com/there-will-be-a-shortage-of-data-science-jobs-in-the-next-5-years-9f783737ed23 (Accessed: 17 May 2023).
PermalinkShow parentReply

Peer Response

by Richard Charnock - Wednesday, 17 May 2023, 9:49 PM

Hi Chris,

In your final paragraphs you mention there will be a greater focus for Data Scientists on ethical data use as more data is generated. It is a point I tend to agree with but wondered if you had any further thoughts on why more data might make that the case or if there any other aspects of the role of the Data Scientist or society which will require such responsible use?

Some of the points made by Leslie (2019) are worth consideration. Data Science models and AI are built upon already have inbuilt bias and discrimination from the source systems, societies and organisations that generated. AI systems are not accountable like a human may be; and the systems may be so complex that the results are not open to scrutiny. Importantly when considering the Data Management Life Cycle, a weak DML with poor data quality and questionable technical implementation could result in low quality outputs which are detrimental to individuals, and ultimately undermine public confidence in Data Science practices and the potential benefits.

References:

Leslie, D. (2019) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. **https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf** [Accessed 17 May 2023]

Peer Response

by Srihari Prasad Bhallamudi Bhallamudi - Saturday, 20 May 2023, 4:07 PM

Hi Chris,

I found your information on the 4th Paradigm of Science quite insightful and how it clearly highlights both the opportunities and challenges that lie ahead. The acceleration of data generation indeed makes the task of extracting meaningful patterns from these massive datasets both exciting and daunting. It is interesting how this concept can be linked and referenced to data science quite far back and as pointed out by Gray (2009), the way in which data scientists and researchers collaborate with each other in this field will offer insights into its full potential.

As pointed Yıldırım (2020), I agree that there are many challenges that data scientists face both now and in future, especially in terms of their initial workload for even entry level data scientists within the field. While this does pose a significant hurdle, it also provides an avenue for growth and learning. As Yıldırım (2020) also suggests, data scientists, despite being stretched thin, gain exposure to a variety of tasks and skills, thus providing them with an accelerated learning curve. Moreover, as the field matures, we can anticipate the development of support structures and the division of roles that can alleviate this burden. Therefore, it seems that the future will hold not only challenges but also solutions and progress in the field of data science.

Yıldırım (2020), The Dark Side of the Sexiest Job of the 21st Century **https://towardsdatascience.com/the-dark-side-of-the-sexiest-job-of-the-21st-century-fd9c46bf4cae** [Accessed 19 May 2023].

Gray (2009), The Fourth Paradigm: Data-Intensive Scientific Discovery **https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/** [Accessed 20 May 2023].

In reply to Chris Final

Peer Response

by Danielle Hall - Monday, 22 May 2023, 1:58 AM

Hi Chris!

In the third paragraph of your post, you quote Saxena (2021), who states that “the need to understand machine learning algorithms to build a better-automated system…will eventually lead to more jobs”. This leads you to state that you believe that there will be more jobs and higher employment within this field in the future. However, from my understanding Saxena first mentions that there will be a shortage of jobs due to automated pipelines, and only after a long period of time will more jobs become available as companies hire data scientists to create better automated pipelines than their competitors.

Personally, I took this as meaning that data scientists are in high demand as of right now but as companies implement automated pipelines that demand will decrease, resulting in less jobs available for data scientists. Eventually, more jobs will become available, but nowhere near the same amount as before. This was my interpretation of what Saxena (2021) stated. I'm curious as to whether you agree, or if you think that the future of data scientists is more jobs available than now, rather than less.

Reference List:

Saxena, P. (2021) There Will Be a Shortage of Data Science Jobs in the Next 5 Years? Available from: https://towardsdatascience.com/there-will-be-a-shortage-of-data-science-jobs-in-the-next-5-years-9f783737ed23 [Accessed 7 May 2023].

Summary Post

To summarise, we have investigated what a Data Science entails, and how we are entering the 4th Paradigm of Science and the exploration of Big Data. With more information to be used in the Data Management Lifecycle (DML), more accurate insights and conclusion can be drawn to benefit businesses and society. However, there is also growing responsibility on Data Scientists to act ethically and legally with the data they collect. Consequently, the General Data Protection Regulation (GDPR) was introduced into EU law in 2018. GDPR has 6 key principles which are:

· Personal information to be processed lawfully and fairly in a transparent manner.

· Personal information shall be collected for specified purposes

· Personal data shall be adequate, relevant, and limited for its intended purpose

· Personal data shall be accurate and kept up to date, when necessary

· Personal information shall be retained for a long as it's necessary

· Personal information shall be processed appropriately to ensure security

One of the main methods in maintaining GDPR standards is by using a combination of encryption and pseudonymisation for personal information (Verhenneman 2017). GDPR will also help to manage the Data Management Lifecycle (DML). As my peer, Richard Charnock mentions, not managing the DML can lead to "poor data quality and questionable technical implementation" resulting in "low quality outputs which are detrimental to individuals, and ultimately undermine public confidence in Data Science practices".

Exploration into Master Data Management (MDM) has been of focus in the course. “MDM is a combination of applications and technologies that consolidates, cleans, augments corporate data and synchronizes it with all applications, business processes, and analytical tools” (Kumra and Mishra 2011). The Master Data revolves around 4 main areas which are, People, Things, Places and Concept. As stated by Kumar & Mishra (2011) master data needs to be “clean, consistent and accurate”, which is vital for avoiding inconsistencies.

For the future of Data Science, I have already mentioned in my initial post that Data Science will generate more jobs, especially with the rapid expansion of AI and Machine Learning in recent times. However, in the short term, automation will replace some job roles within the economy as stated by Saxena (2021) and my peer, Danielle Hall. The labour supply of Data Scientists will remain restricted, but over time will expand. I believe the expansion in supply will grow exponentially as the demand for labour grows. More job losses will occur in the long term for sectors which have AI replace their roles. However, the supply of labour will grow again as the employment workforce look to transfer skills related to data science topics. My hypothesis is based off how supply and demand act as compliments in the labour force environment. Government interaction may also be needed to ensure that training/investments opportunities in the data science environment exist, creating expansion in the labour supply of Data Scientists.

References

Kumar, D & Mishra, M. (2011) A study on challenges and opportunities in master data management. International Journal of Database Management Systems, 3(2), pp.129-139.

Verhenneman, G. (2017) Encryption in GDPR. In HEAT final workshop, Date: 2017/11/27-2017/11/28, Location: Leuven.

Saxena, P. (2021) There Will Be a Shortage of Data Science Jobs in the Next 5 Years? Available at: **https://towardsdatascience.com/there-will-be-a-shortage-of-data-science-jobs-in-the-next-5-years-9f783737ed23** [Accessed 22 May 2023].