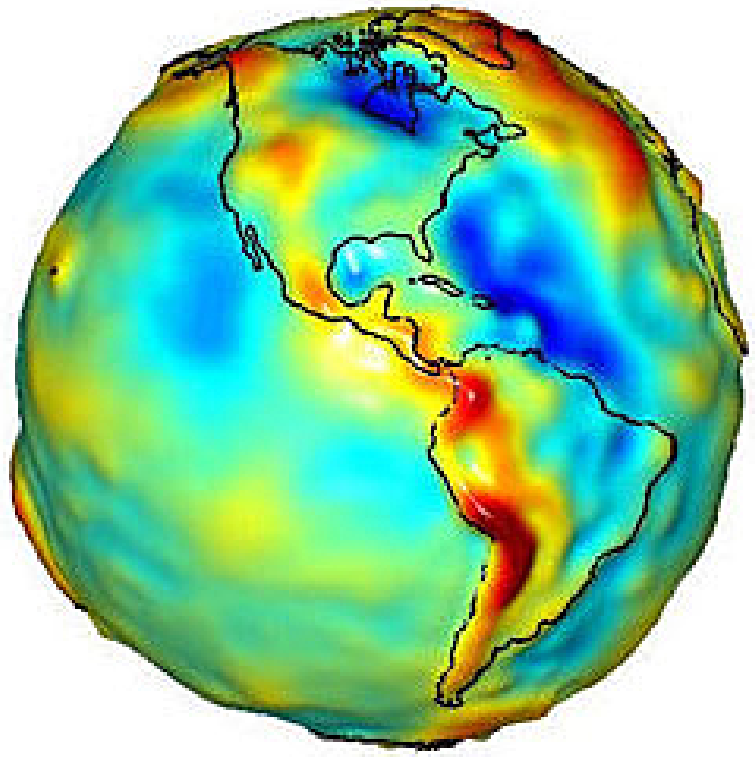
Data Science

Ingredients

1. Data Science
   – "...interdisciplinary field about scientific methods, processes and systems to extract knowledge or insights from data in various forms..."
2. Draws inspiration from e.g. statistics, machine learning, systems, databases and applied mathematics...
3. Often treats text, images, shapes as data etc.



What does great data science look like?

UCL

Broad range of topics

Methodological topics

1. Algorithms & architecture co-design
2. Data analytics platforms
3. Databases
4. Events and anomalies
5. Geometry & topology
6. Indexing, labelling etc
7. Models and learning
8. Monitoring and analytics
9. Predictive modeling
10. Privacy and security
11. Systems

Application driven topics

1. Data Science for Public Services
2. Internet of Things analytics
3. Health and Biomedicine
4. Manufacturing (Industry 4.0)
5. Legal sector
6. Physical Sciences
7. Social data science

What are the characteristics of the ideal data scientist?

What does a data science workflow look like?

Sampling paradigm?

Setting

1. Ubiquitous sensing: data is collected automatically.
2. Human subjects data.
3. No sampling paradigm: instead data=all?
4. Bias by (lack of) design: e.g. Twitter data generated by Hurricane Sandy, StreetBump smartphone app.
5. Fidelity and generalizability: see for example precision medicine.
6. Prediction vs Estimation.

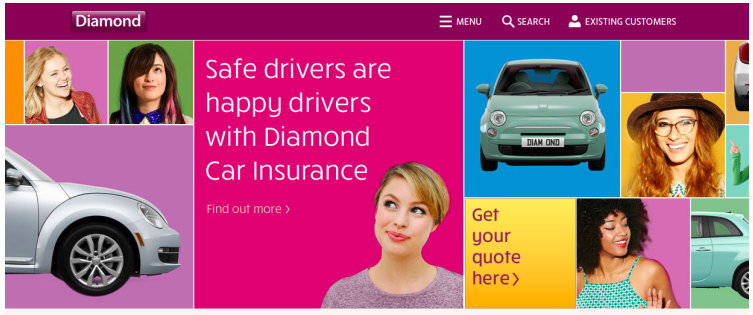
Modern forms of data.

What is a data scientist's responsibility?

Fairness & Transparency

Ingredients

1. Statistics are collected automatically from our daily lives ⇒ surveillance society?
2. Transparency: with very large amounts of complex data, and complex algorithms, how are decisions taken?
3. Fairness: What does fair mean?
4. Consent: should data be given away for perpetuity?



Article

An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures

Crime & Delinquency
2014,Vol. 60(2) 167–192

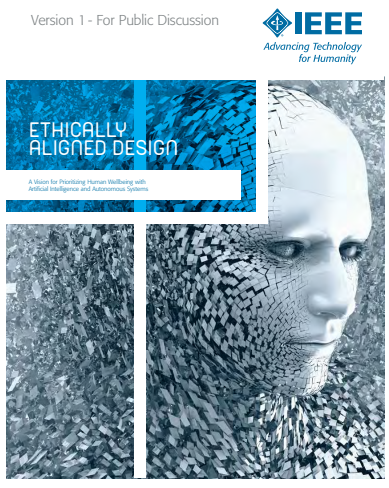
DOI: 10.1177/0011128711426544
cad.sagepub.com
SAGE

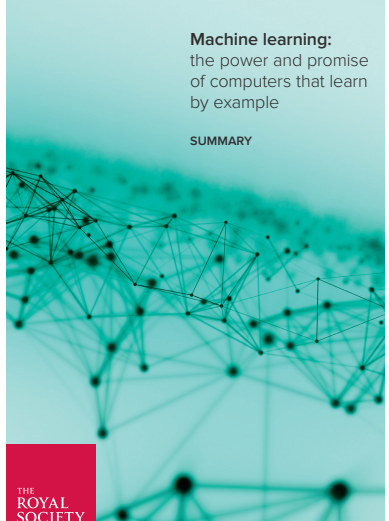
Fair? Transparent?

Impacts of data usage and algorithms

Ingredients

1. Government interest–UK GO Science & White House
2. Royal Society and British Academy project on data governance
3. IEEE: Ethical Considerations in Artificial Intelligence and Autonomous Systems
4. **Should there be an explicit data science code of ethics and behaviour?**





Theory of Big Data 3

26–28 June 2017, London, UK

1. Challenges in spatial & temporal analysis
2. High-dimensional estimation and learning
3. Privacy-preserving inference
4. Tensors and statistical modelling

Abstract submission:
14th March 2017.

- Confirmed speakers: Ming Yuan, Kostas Chatzikokolakis, Farinaz Koushanfar, Lek-Heng Lim, Christos Dimitrakakis, Martin Wegkamp, Peter Diggle, Lieven De Lathauwer, Peter Hoff, Tien Zheng, Guy Nason, Arnak Dalalyan, Arthur Gretton, Heather Battey, ...
- Mailing list: http://www.ucl.ac.uk/bigdata-theory