Information Extraction and Ethics

Natural Language Processing Module 2019 (Dr N Aletras)

Prof Jochen L Leidner, MA MPhil PhD FRGS
⟨leidner@acm.org⟩

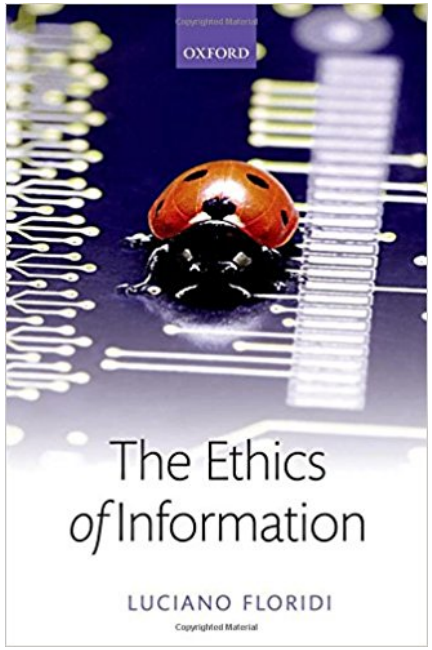
The University Of Sheffield.

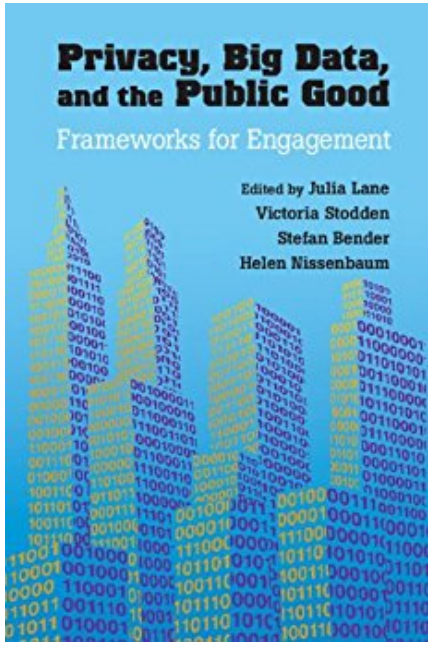
Ethics

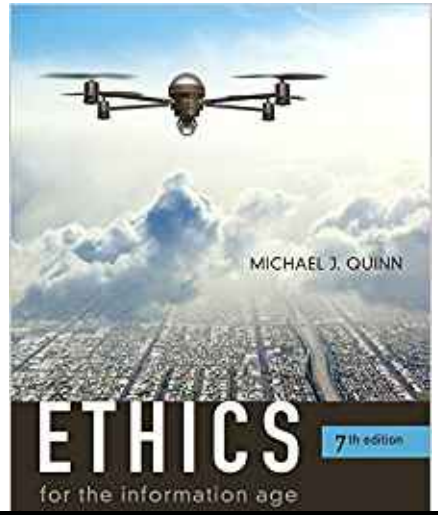
Some Book Recommendations

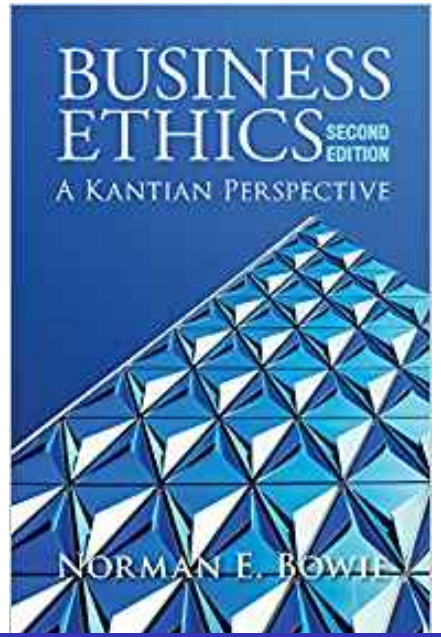
- Floridi, *The Ethics of Information*, 2013
- Lane *et al.* (ed), *Privacy, Big Data and the Public Good*, 2014
- Quinn, *Ethics for the Information Age, 7th ed.*, 2016
- Bowie, *Business Ethics: A Kantian Perspective. 2nd ed.*, 2017

Some Book Recommendations









The Ethics of Language Technology and Data Science

- Information processing does not happen in a vaccuum
- Big data risks: legal risk, compliance risk, **ethical risk**?
- What's the difference between legal and ethical issues?

Fundamental Questions of Philosophy (Kant, 1776)

- “Why am I here?” $\rightarrow$Ontology
- “What can I know?” $\rightarrow$Epistemology
- **“How should I live?” $\rightarrow$Ethics**
- “What may I hope?” $\rightarrow$Philosophy of Religion
- “What is man?” $\rightarrow$Anthropology, Politology

Ethics and Morality

- Ethics is the sub-discipline of philosophy that addresses the question “how should I live?”
- Unlike the natural sciences, which seek factual answers, ethics is looking for *normative* answers.
- Trying to live a moral life is the choice of humans as rational beings that have a free will.
- Ethical and legal considerations may sometimes overlap, but are separate realms.
  - The law codifies ethical normas in a society (at a certain time).

(Some) Ethical Issues

- Privacy: moral right to self-determination (right to be anonymous, control who knows what about self)
- Fairness: right to same treatment for everyone (avoiding bias)

(Some) Legal Issues

- Privacy: Obligation to protect personal data
  - Personal and Personally Identifiable Information (PII)
  - EU Data Protection Directive (Directive 95/46/EC)
- Copyright: protect the ownership rights of works of creative expression
- Patents: Time-limited, territorial monopoly in return for method disclosure
- Liability: obligations from holding data
- Discrimination: treating some groups differently (in a negative way)

Privacy

- Moral right of personal data holders to keep one's data private
- often enshrined in constitutions as a legal right
- How valuable is the moral right of a person to be private, in other words to keep (certain information) to themselves, away from governments and other people.
- We nowadays find it hard to live without sharing data:
  - Legal, societal, commercial pressure to disclose (ads)
  - Self-disclosure: social media, affiliate programs
  - counter-terrorism surveillance erodes this liberty
  - Risk of abuse of the data (identity theft, burglary, stalking/bullying, abduction, ...): in general, using the data against the person the data is about
- medical records: deeply personal or public research resource?
- Danger 1: on the Internet, every disclosure is permanent and globally accessible (due to Web crawlers, search engines)

Ethical Dilemmas

- Companies can provide better services (and at lower cost) if they know more about us.
- Biomedical researchers may be able to cure diseases better if they have more health data available.
- But: risk of invasion of personal space, abuse (crime, anti-social action, oppression)

Selected Ethical Schools of Thought

- Aristotelianism
- **Asimov: 3 laws of robotics**
- Consequentialism
- Deontology
- **Kantianism** (Categorial Imperative): intention→universal law
- Rawls
- **Golden Rule**
- Principalism
- **Utilitarianism**

Asimov's Laws of Robotics and Machine Ethics (Asimov 1950)

- What should a machine ethic look like?
- For example, a mobile robot or self-driving car may need to provide built-in *ethics rules* to handle conflicting situations in a well-defined way.
- Isaac Asimov (1942) stipulates the 3 Laws of Robotics in his short story"Runaround":
  1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
  2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
  3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

  — *I, Robot*

The Golden Rule

- “Do unto others as you would have them do unto you.” (already in Egypt, 2040 BC)
- Problems: subjective, assumes everybody has same moral compass

Utilitarian Ethics

- Originated in England: Mills, Hobbs, Bentham
- Take the course of action that maximizes happiness for the majority of people
- Problem: how to measure happiness? What about minorities?

Accountability and Transparency

- There should be clarity about responsibility and process in anything, by implication also in NLP, ML, big data.
- Why did an algorithm/ML model behave as it did? How can anyone adversely affected by an application seek a remedy?
- We do not know yet how to build fair systems by design
- New Workshops:
  - *Fairness, Accountability and Transparency in Machine Learning* (2014-)
  - *First Workshop on Ethics in NLP* at EACL (2017) in Valencia
  - `http://ethicsinnlp.com/`

Errors, Omission and Correction of Big Data

- What if a dataset about you contains errors?
  - What if mining data introduces errors?
- You never know who uses that data set (unanticipated use)
- User (and only right user) should have the ability to correct

Governance

- Who is responsible for holding the data, protecting it (not leaking it, deleting it when no longer needed)?
- What processes are in place to deal with the data appropriately?
- Legal/compliance term, but can serve as the “institutional home” to address ethical questions in corporate environments

Machine Learning and Bias

- ML captures properties inherent in the data
- Inherent problem: in ML, we want to find **discriminative features** (features that separate the classes well)
- This may result in **discriminating classifiers**
- Norotious (2016) example: Google's photo application (mis-)recognized a gorilla as a (black) human

Anonymity, Anonymization and De-Anonymization

- anonymity is the property of acting without disclosing one's identity
- anonymization is the transformation of a dataset into a form from which no PII is recoverable
- typically done via replacing all PII with a unidirectional hash function
- `Jochen Leidner` $\mapsto$ `7890dc541b4aea7ccbdacba55bb0d316b3f9f0c7965772c8956d1f3652`
- de-anonymization is applying the reverse transformation (in cases where the reversibility of the preceding anonymization was not actually guaranteed)

Consent and Informed Consent

- Strongest form of protection is the requirement for “express prior written consent means”
  - giving permission
  - in advance
  - in writing
- **Informed consent** (medical law): “An informed consent can be said to have been given based upon a clear appreciation and understanding of the facts, implications, and consequences of an action.” (Wikipedia)
- originated from legal/ethical obligations in medicine prior to invasive procedures, and requires volunariness, capacity to comprehend consequences and checking comprehension has happened.

Case Study: Admiral Insurance Wants Your Facebook

BBC Sign in News Sport Weather iPlayer TV Radio

NEWS

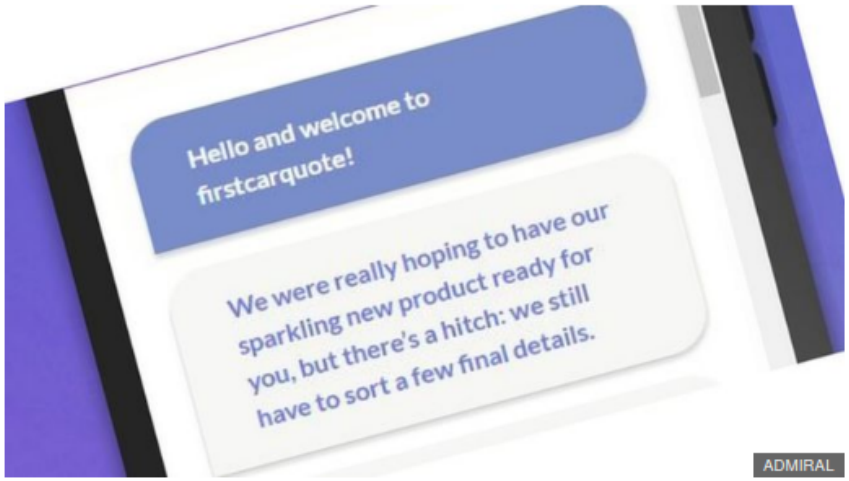
Home UK World Business Politics Tech Science Health Education Entertainm

Business Your Money Market Data Markets Companies Economy

Facebook blocks Admiral's car insurance discount plan

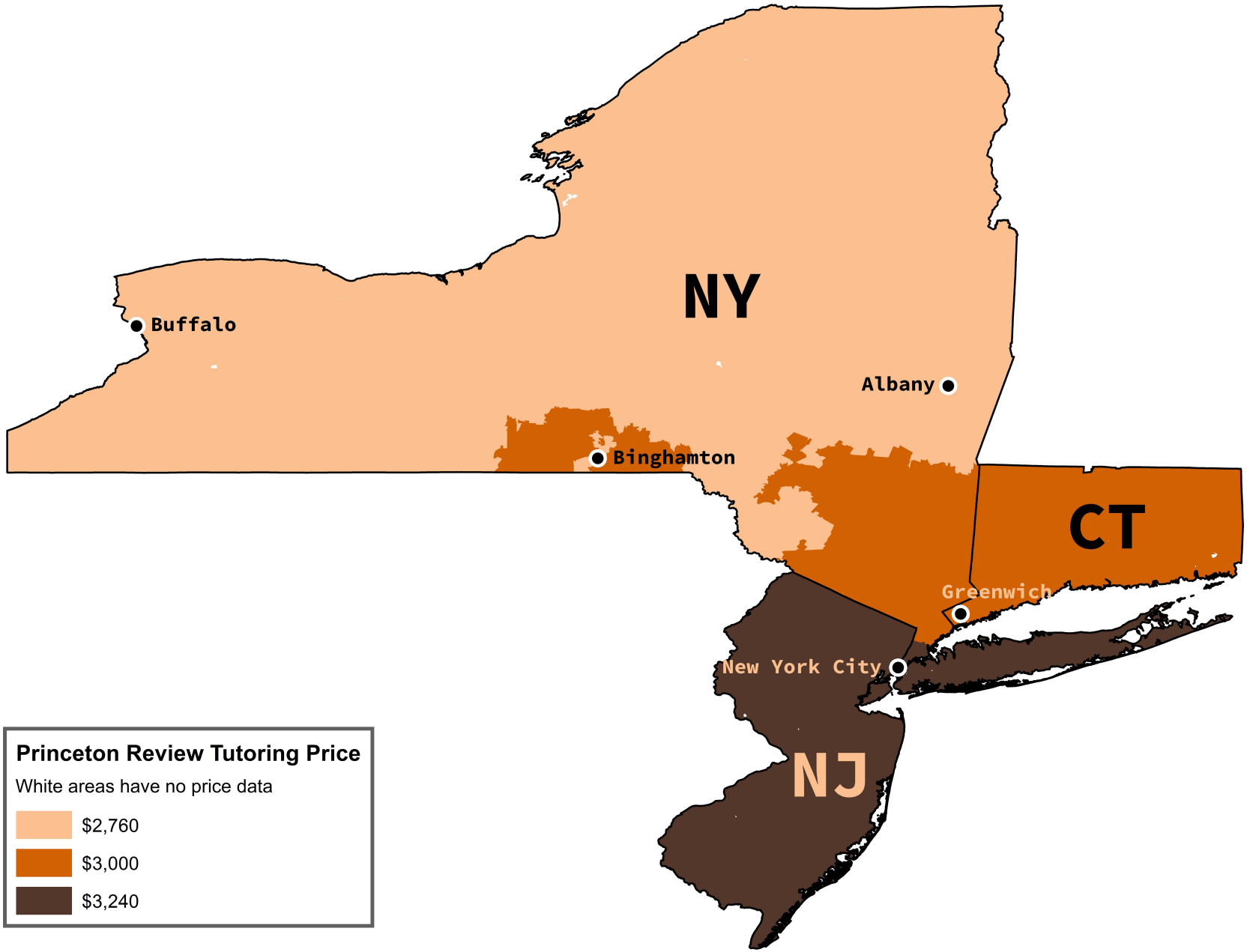
By Kevin Peachey
Personal finance reporter

2 November 2016 | Business | Share



Facebook has blocked plans by an insurer to view young drivers' profiles to help set car insurance premiums.

Case Study: Princeton Tutorial Pricing Discrimination



Case Study: Wisconsin Re-Offender Scoring

Case Study: Facebook News Bias

Facebook news selection is in hands of editors not algorithms, documents show

Exclusive: Leaked internal guidelines show human intervention at almost every stage of its news operation, akin to a traditional media organization

According to Facebook's guidelines, a team of news editors are instructed on how to 'inject' stories into the trending topics module. Photograph: Adam Berry/Getty Images

Legal and Ethical Questions for Big Data Practitioners

- **Abstraction** It is helpful to characterize a human with data, however a reduction of a human being to a mere set of data points is unethical (human dignity is also enshrined in some constitutions).
- **Algorithmic Bias** Is the big data method fair and unbiased to the whole population, intentionally or accidentally?
- **Privacy** Does a project work with PII information? Does the work respect the privacy rights of all individuals involved?
- **Copyright** Are copyright and the moral right to be recognized as author respected?
- **Competence** Does the experimenter have the statistical knowledge to conduct a big data experiment in a methodologically sound way?
- **Transparency** Can the method be inspected (in a code audit) to guarantee that what is said about what is done is actually what is done by the code? Can the user inspect what information the system holds about him or her, and correct errors in the data?

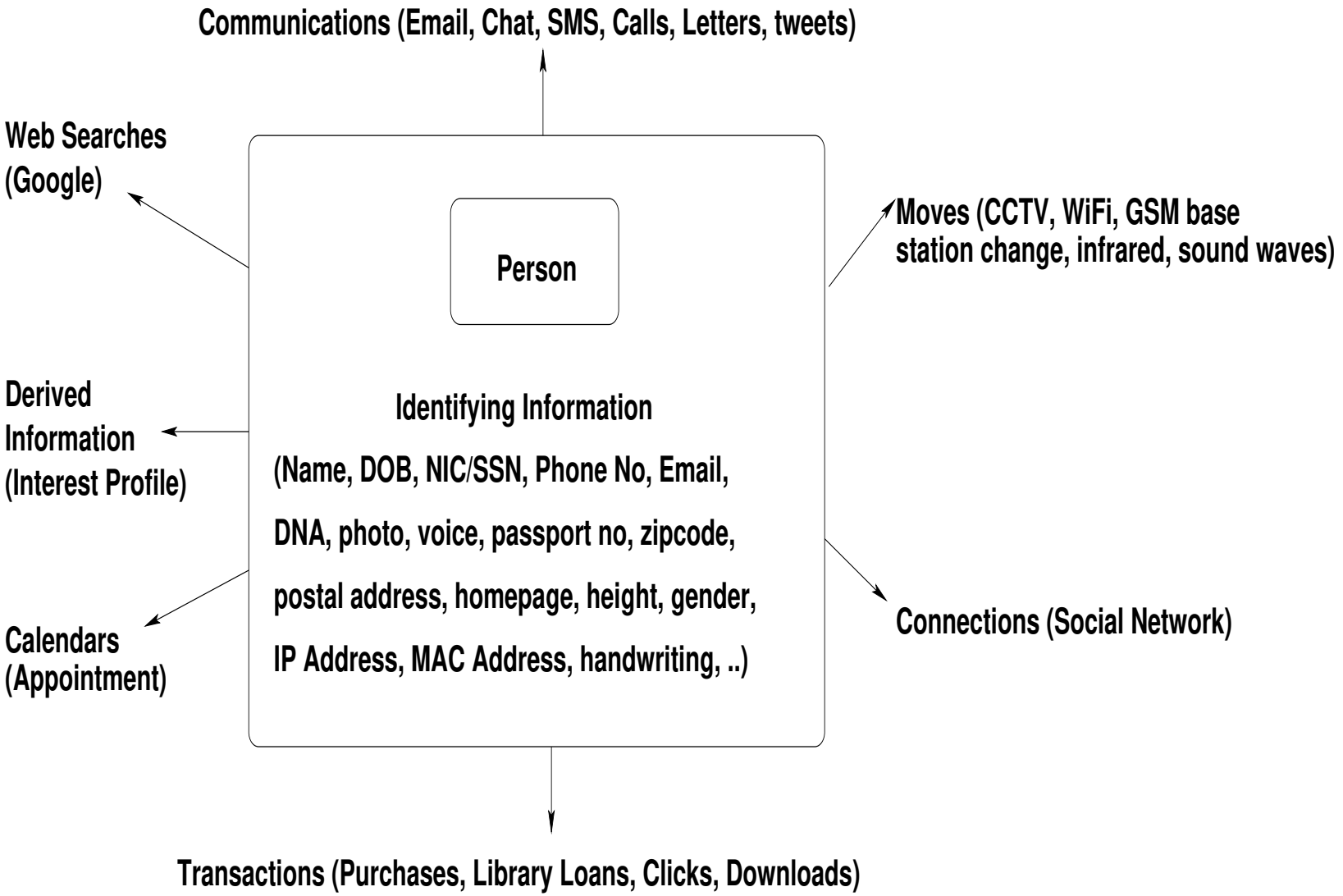
Documented Ethical Issues in NLP (Leidner and Plachouras, 2017)

- Automation: customer support chat-bots, robo-lawyers $\rightarrow$ job loss
- McIllroy's UNIX spell command: hidden e-mails to the developer
- The de-humanising API: caring about the person behind the crowdsourcing API
- Nasty chatbots: Mexican human rights activist trolled by chatbots (Thieltges, Schmidt and Hegelich, 2016)
- Privacy and consent in corpus construction: The "Pierre V." problem
- Dual use: personal profiling for marketing (customer segmentation) versus political influencing (Kosinski, Stillwell and Graepel, 2013): the *Cambridge Analytica* scandal
- Exclusion: street banks close, an elderly lady from Uddingston's accent is not recognized by ASR

Bias and Discrimination

- A health insurance may use big data to obtain statistics about diseases and fitness behavior and subsequently increase the insurance premium of some customers, or even cancel their policy due to them being not profitable.
- Car manufacturers obtain live data from modern car computers, which are already passed on to insurance companies, who analyze whether the driving style is
- Should the buyer of a car know about this? Should they be able to opt out?
- What typically happens is all market players embrace one way and nobody offers a traditional alternative. In a free market, a few individuals will be unable to create enough demand.
- A machine learning classifier learns that in a data set, people with black skin and a certain postcode are more highly correlated with crime; should the classifier be barred from being used because it is discriminating?

Dimensions of Personal Data (Modified after WEF, 2011)



Automation (1/2): The Powerless Human

- We believe in our right to self-determination.
- If more and more tasks get automated, it may no longer be possible to effect a change manually, we may lose control to automated work-flows that cannot be manually interacted with.
- The result could be a lack of freedom, and practical issues pertaining to the difficulty of correcting errors where automation gets it wrong.

Automation (2/2): The Jobless Human

- In the industrial revolution, menial mechanical jobs were automated.
- In the information revolution, more advanced, white collar jobs (e.g. business analyst) could also be automated, which raises the questions how the affected workers can earn their livelihood in a big data world.

Case Study: Data Science and the Trump Election – Cambridge Analytica

THE SPECTATOR

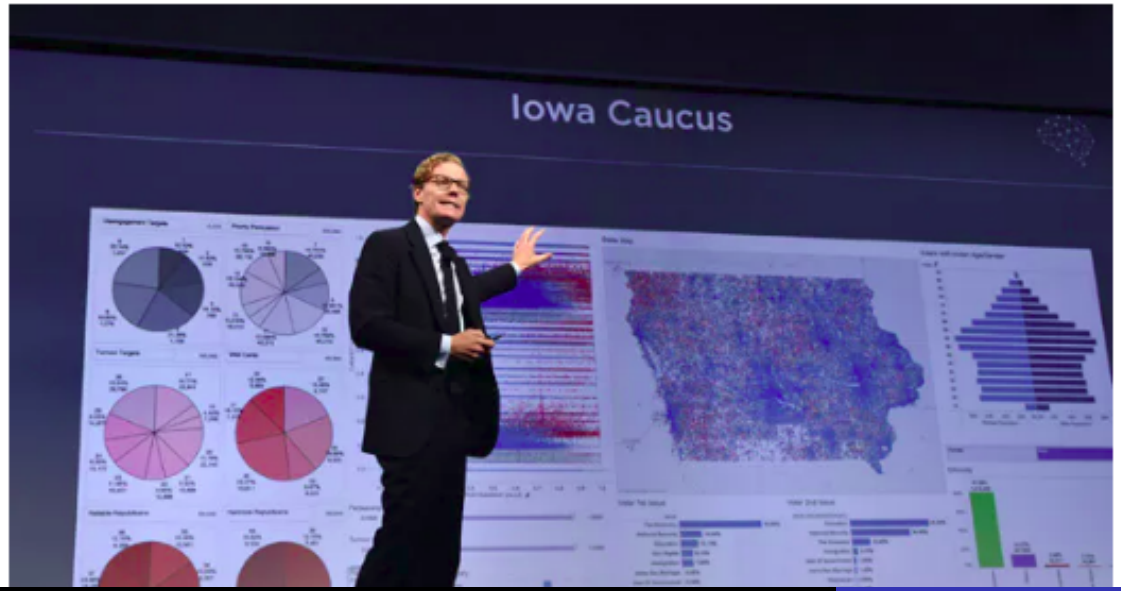
COFFEE HOUSE GENERAL ELECTION 2017 MAGAZINE WRITERS BOOKS & ARTS

FEATURES

The British data-crunchers who say they helped Donald Trump to win

Are Cambridge Analytica brilliant scientists or snake-oil salesmen?

Paul Wood



Natural Resources and Stewardship

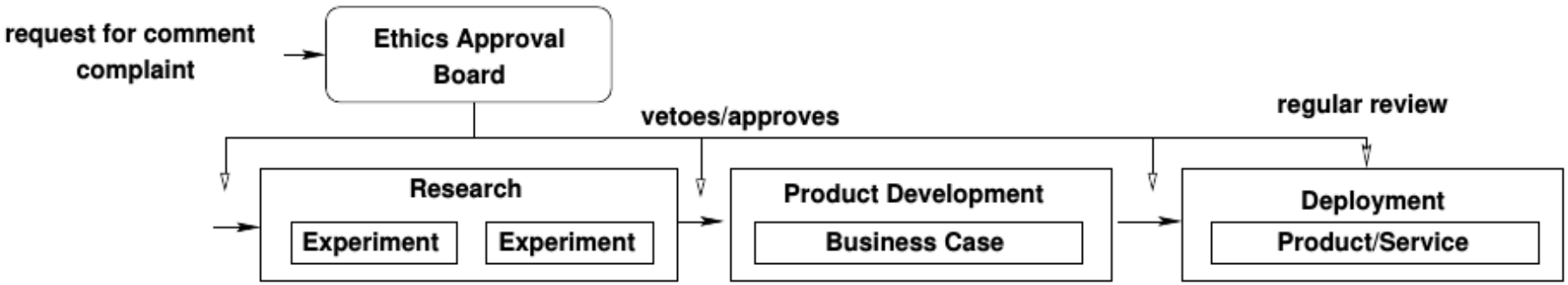
- Data centers that process enormous amounts of data also consume enormous amounts of energy. Data centers contribute to global warming.
- Does any given process justify the natural resources consumed?

Towards Solutions

So what can an individual do when confronting ethical problems?

- Seeking user's consent
- Getting a neutral third party's opinion (e.g. universities have ethics committees that approve some kinds of experiments)
- Defining a process for dealing with issues
- Protesting (logging formal documentation of disagreement): not a remedy by itself, but can start discussions leading to change
- Declining to comply, asking to be relieved of a project
- Walking away (you always have the option of not doing something): resign

Ethics & Data Science – Ethics Reviews and ERBs (Leidner & Plachouras, 2017)



Ethics & Data Science – Responding to Ethical Issues (Leidner & Plachouras, 2017)

Demonstration	to effect a change in society by public activism
Disclosure	to document/to reveal injustice to regulators, the police, investigative journalists ("Look what they do!", "Stop what they do!")
Resignation	to distance oneself III ("I should not/cannot be part of this.')
Persuasion	to influence in order to halt non-ethical activity ("Our organization should not do this.")
Rejection	to distance oneself II; to deny participation; conscientious objection ("I can't do this.")
Escalation	raise with senior management/ethics boards ("You may not know what is going on here.")
Voicing dissent	to distance oneself I ("This project is wrong.")
Documentation	ensure all the facts, plans and potential and actual issues are preserved.

IBM's Work on Fairer Face Recognition (Merkler et al., 2019)

Diversity in Faces

Michele Merler, Nalini Ratha, Rogerio Feris, John R. Smith
IBM Research AI @ IBM T. J. Watson Research Center
Yorktown Heights, NY 10598, USA
Contact: jsmith@us.ibm.com

February 22, 2019

Abstract

Face recognition is a long-standing challenge in the field of Artificial Intelligence (AI). The goal is to create systems that detect, recognize, verify and understand characteristics of human faces. There are significant technical hurdles in making these systems accurate, particularly in unconstrained settings, due to confounding factors related to pose, resolution, illumination, occlusion and viewpoint. However, with recent advances in neural networks, face recognition has achieved unprecedented accuracy, built largely on data-driven deep learning methods. While this is encouraging, a critical aspect limiting face recognition performance in practice is intrinsic facial diversity. Every face is different. Every face reflects something unique about us. Aspects of our heritage – including race, ethnicity, culture, geography – and our individual identity – age, gender and visible forms of self-expression – are reflected in our faces. Faces are personal. We expect face recognition to work accurately for each of us. Performance should not vary for different individuals or different populations. As we rely on data-driven methods to create face recognition technology, we need to answer a fundamental question: does the training data for these systems fairly represent the distribution of faces we see in the world? At the heart of this core question are deeper scientific questions about how to measure facial diversity, what features capture intrinsic facial variation and how to evaluate coverage and balance for face image data sets. Towards the goal of answering these questions, Diversity in Faces (DiF) provides a new data set of annotations of one million publicly available face images for advancing the study of facial diversity. The annotations are generated using ten facial coding schemes that provide human-interpretable quantitative measures of intrinsic facial features. We believe that making these descriptors available will encourage deeper research on this important topic and accelerate

Yesterday: Non-Consent in IBM's Face Recognition Data-Set (NBC)

Facial recognition's 'dirty little secret': Millions of online photos scraped without consent

People's faces are being used without their permission, in order to power technology that could eventually be used to surveil them, legal experts say.